Available online at www.sciencedirect.com
ScienceDirect
Journal of Chromatography B, 847 (2007) 12–23
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Review

SELDI-TOF mass spectra: A view on sources of variation ☆

Martijn Dijkstra [a,*], Roel J. Vonk [b], Ritsert C. Jansen [a]

[a] *Groningen Bioinformatics Centre, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, P.O. Box 800, NL-9700 AV Groningen, The Netherlands*
[b] *Centre for Medical Biomics, University Medical Centre Groningen, A. Deusinglaan 1, 9713 AV Groningen, The Netherlands*

Received 30 June 2006; accepted 5 November 2006
Available online 21 November 2006

Abstract

Adequate interpretation of mass spectrometry data can yield valuable biomarkers. However, spectrum interpretation is a complicated task. This paper reviews the various factors that determine a sample's spectrum and demonstrates the role of these factors in the interpretation process. We derive a simulation model that adequately predicts the expected spectrum based on known sample content and, in the reverse mode, obtain an analysis model that adequately fits an observed spectrum based on the hypothesized sources of variation.

Keywords: Mass spectrometry; Quantitative analysis; SELDI; Mixture models; Spectrum analysis

Contents

1. Introduction 13
2. SELDI-TOF technology and sources of variation 13
   2.1. From sample to chip 14
   2.2. From chip to detector 14
      2.2.1. The desorption/ionization-process 14
      2.2.2. Intermolecular complexes 15
      2.2.3. Separating molecules 15
      2.2.4. Delayed extraction 15
      2.2.5. Chemical noise 15
   2.3. From detector to spectrum 16
   2.4. The spectrum 16
      2.4.1. The TOF-spectrum and the *m*/*z*-spectrum 16
   2.5. Other sources of variation 16
      2.5.1. Post-translational modifications 16
      2.5.2. Averaging spectra 16
3. A practical example 16
   3.1. Sample preparation 16
   3.2. Peak positions 17
   3.3. Peak areas 17
   3.4. Peak shape 17
   3.5. Baseline 18
   3.6. The machine parameters 18

☆ This paper was presented at Biomarker Discovery by Mass Spectrometry, Amsterdam, The Netherlands, 18–19 May 2006.

* Corresponding author. Tel.: +31 503638091.
E-mail address: m.dijkstra@rug.nl (M. Dijkstra).

1570-0232/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2006.11.004

4. Analyzing our practical example 20
4.1. Simulation and analysis model 20
4.2. Mixture components 20
4.3. Parameter estimation 21
4.4. Quantifying sources of variation 21
5. Discussion 22
Acknowledgement 23
References 23

1. Introduction

Mass spectrometry is an analytical technique that analyses a physical sample (e.g. a peptide/protein mixture) and generates a "mass spectrum" with peaks that represent the masses and the amounts of the sample components. However, retrieving these masses and amounts from a mass spectrum is a complicated task and requires an adequate interpretation of the spectrum.

A mass spectrum usually contains many more peaks than the number of different molecule species present in the sample, because molecules form complexes and/or carry multiple charges and therefore appear at several locations as peaks in the spectrum. Fig. 1, for example, shows a mass spectrum of a single protein, which nevertheless contains a large number of peaks. Even apparent singleton peaks are often still a mixture of several smaller peaks, leading to a right-skewed peak shape. This is primarily due to matrix adducts, post-translational and on-chip modifications, and isotopic variants. The broadness of such peaks depends on the heterogeneity of the corresponding molecules. Peak areas on the time of flight (TOF) scale are assumed to be proportional to the measured concentrations of the corresponding molecules [1], but only if overlapping peaks are properly resolved [2].

Taking all the different sources of variation in mass spectra into account should enable the exact determination of the variation contributed by individual sample proteins. Correlating these sample protein variations to sample phenotypes may facilitate the discovery of biomarker proteins (i.e. those proteins which are expressed differently between different phenotypes).

Section 2 presents a theoretical view on sources of variation in surface enhanced laser desorption/ionization time of flight (SELDI-TOF) analysis. Section 3 illustrates this theoretical view with real example data from designed experiments. Section 4 develops and applies a statistical analysis approach to quantify the discussed sources of variation in the example data. The final section summarizes and discusses our findings on sources of variation in SELDI-TOF and related techniques.

2. SELDI-TOF technology and sources of variation

For SELDI a biological sample (usually a protein solution) is put on a pre-coated stainless steel slide. The coating 'enhances' the surface to bind preferentially a particular class of proteins based on their chemical properties. Different coatings give different 'chip types' which bind to different classes of proteins [3]. Washing the chip removes weakly bound proteins, which have low affinity for the specific chip. Such fractionation reduces the complexity of a mixture and prevents high-abundance proteins suppressing the low-abundance proteins. The protein fraction, which is retained by the chip is then desorbed and ionized as in the matrix assisted laser desorption/ionization (MALDI) method. The subsections below describe the SELDI process in more detail and highlight the most important sources of variation, numbered (1–20), along the proteins' route from sample to spectrum (summarized in Table 1).

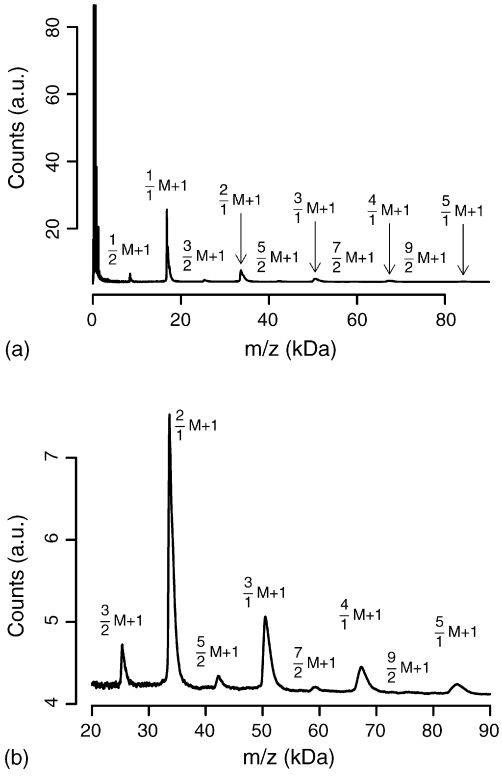

Fig. 1. A practical example. The mass spectrum of the myoglobin molecule (M = 16,951 Da) shows peaks at several locations in the spectrum because molecules form complexes and/or carry multiple charges. Hence, the peak positions, which are indicated by the peak labels, are all related to the molecular weight of myoglobin (M), by multiplying M with a fraction. The nominator is the number of myoglobin molecules in the complex, which corresponds to that peak. The denominator is the number of charges, which the complex carries. For example, the singly charged myoglobin peak comprises one myoglobin molecule and one proton (unit mass and unit charge), has $m/z = (1 \times M + 1)/1 = (1/1) \times M + 1$ as label. (a) Shows an overview of the full spectrum (0–90 kDa) and (b) shows a close-up of the mass range 20–90 kDa.

Table 1
Sources of variation

	Short description	Impact on the spectrum
(1)	Different chip types bind different classes of proteins	Determines the peak set
(2)	Washing the chip removes weakly bound proteins	Removes peaks. Reduces peak area[a]
(3)	Competition for sites in the matrix	Peak area[a]
(4)	Molecules can get more than one charge	Many more peaks
(5)	Laser energy	Peak area (0–100%). Peak resolution
(6)	Proteins can denature. Denatured proteins have larger surface area	Peak area[a]
(7)	Proteins with larger surface area can carry more charge	Peak area[a]
(8)	Molecules form intermolecular complexes (cluster ions)	Many extra peaks (∼80): m-mers ($m = 1 \ldots 10$), 8 satellite peaks per m-mer
(9)	Complexes formed with salt ions (which have mass and non-zero charge) from the washing buffer	Peak shift
(10)	Fragmentation of molecules and complexes	Chemical noise—many minor peaks (<8 kDa)
(11)	Competition for being protonized by the matrix	Peak area[a]
(12)	Electric field potential	Peak TOF-position. Peak area (3–14%)
(13)	Delayed extraction	Peak resolution. Peak area (2–17%)
(14)	Detector sensitivity determines multiplication factor of incoming ions and of electric noise	Peak area (4–26%). Noise level
(15)	Electric noise	Noise level (constant across spectrum)
(16)	Detected air molecules because flight tube is not completely vacuum	Noise level (constant across spectrum)
(17)	Digitizer rate	Number of data points
(18)	m/z-Calibration	Peak shift[a]
(19)	Post-translational modification	Peak shift
(20)	Variation in the density of the matrix crystal	Spectral area[a]

Table 1 summarizes the discussed sources of variation (1–20) and their impact on the spectrum. We analyzed our experimental data for the sources (4) multiple charges, (5) laser energy, (8) formation of intermolecular complexes, (10) fragmentation of molecules and intermolecular complexes, (12) electric field potential, (13) delayed extraction, (14) detector sensitivity, (15) electric noise, (16) detector noise, (18) m/z-calibration, (19) post-translational modifications.

[a] We did not exactly quantify all the sources of variation.

2.1. From sample to chip

Different SELDI chip types have surfaces, which range from chromatographic chemistries that bind many different molecules, to surfaces with a specific biomolecular affinity (e.g. antibodies, receptors, enzymes and ligands) that bind one specific molecule or molecular class. After putting the sample on a chosen chip type (1), and washing (2) weakly bound proteins away, the sample is mixed with small photosensitive molecules, which causes the entire mixture to crystallize and form a so called matrix on the chip as it dries. These photosensitive (matrix) molecules facilitate desorption and ionization of proteins (below). Proteins compete for sites in the matrix crystals (3) if the protein concentration is so high that not all proteins can be incorporated in the matrix. Proteins that are more easily embedded will then have a higher concentration in the matrix. The sample is then put into an almost vacuum chamber, the so called flight tube.

2.2. From chip to detector

SELDI shares several characteristics with MALDI. We refer to Chapter 3 in ref. [4] for an overview of the MALDI method. Zenobi and Knochenmuss [5] give an extensive review of ion formation in MALDI mass spectrometry.

2.2.1. The desorption/ionization-process

A short laser pulse hits the crystal structure and excites the matrix molecules. The energy of the excited matrix molecules leads to excitation of other matrix molecules, and is converted to thermal energy which heats up the crystal locally to around 1000 K within a fraction of 1 ns [5]. Excited matrix molecules can protonate another matrix or protein molecule. The overheated part of the crystal explodes together with the embedded proteins into a plume. The physics of this process is not fully understood. There are two types of ionization processes: primary ion formation that occurs during and immediately after the laser pulse, and secondary ion forming reactions that take place in the plume later on [5]. The plume expands but keeps a rather high density and a temperature of 500 K for 100 ns [5], causing many collisions between proteins and matrix molecules or matrix clusters. Some matrix molecules are protonated during the primary ion formation, i.e. received a proton. These collisions induce reactions, which include matrix-protein reactions where the protonated matrix transfers its proton to the uncharged protein and creates a charged protein. The process can repeat, such that proteins get multiple charges (4). In the gas phase, proteins have higher proton affinities than the matrix molecules [5]. Different kinds of matrix molecules can be used in the SELDI analysis. The two most common are alpha-cyano-4-hydroxy cinnamic acid (enables efficient laser desorption and ionization of small proteins, e.g. <15 kDa) and sinapinic acid (enables efficient laser desorption and ionization of large proteins, e.g. >10 kDa) [6]. In our spectra (e.g. Fig. 1), the singly charged molecules were generally more abundant than doubly charged ones. Remarkably, the length of the laser pulse does not have much influence on mass spectra as long as the total energy per pulse stays constant [5]. Using more laser energy (5) on the

other hand generates more ions and influences the mass spectra (see below).

Proteins often denature (i.e. unfold) in an acidic environment [7,8], such as the washing buffer used for SELDI. Denatured proteins have a larger surface area (6), and molecules with a larger surface area can carry more charge (7) [9].

2.2.2. Intermolecular complexes

Before, and maybe even during the laser excitation, some of the molecules form intermolecular complexes, also known as cluster ions (8), which are non-specific, non-covalent adducts [10]. The formation of these complexes increases the number of peaks in the spectrum enormously. Self-association or *m*-merization is the process where several ($m=2$ or more) protein molecules link to each other and form complexes [10,11]. Some molecules already have a net charge before laser excitation due to complexes formed with salt ions (9) (e.g. K^+ and Na^+), which are present in a washing buffer [4]. According to the authors of ref. [5], there is generally little prompt fragmentation in the plume and most fragmentation (10) happens by the decay of metastable ions in the flight tube, partially induced by collisions with background gas.

As discussed above, during crystallisation there is a competition for lattice sites (3) and the presence of one analyte may prevent another from being included into the MALDI crystal [4]. During ionization in the plume, analytes compete (11) for protons that are transferred by matrix molecules and if a protonated analyte collides with an unprotonated one which has the higher gas phase basicity, it may pass its proton to the collision partner [5]. Therefore, the presence of one analyte may diminish the signal intensity of another. This phenomenon is called 'suppression effect' [5].

2.2.3. Separating molecules

The physical principle of the TOF analyzer in SELDI is that sublimated molecules, which have a different mass (m) over charge (z) ratio (m/z) are accelerated differently and enter the flight tube with different velocities. Therefore, the time for an ion to pass the flight tube depends on its m/z. The relation between TOF and m/z can be calculated by using the law of energy conservation. The electron volt (symbol eV) is a unit of energy. It is the amount of kinetic energy gained by a free particle, which has a charge that is equal to the elementary charge when it passes through an electrostatic potential difference of 1 V, in vacuum. Both an electron (−1) and a proton (+1) have a charge which equals the elementary charge, but with opposite sign, such that they will be accelerated in opposite directions. A protein which passes through a potential difference of U V, acquires an energy of $V=U$ eV for each charge (12). Hence, a protein with a net charge z acquires energy zV. The energy expresses as kinetic energy, $(mv^2/2)$, where m is the protein mass and v is its velocity:

$$zV = \frac{1}{2}mv^2. \tag{1}$$

Hence, the TOF to the detector equals

$$\mathrm{TOF} = \frac{x}{v} = x\sqrt{\frac{m}{2zV}}, \tag{2}$$

where x is the length of the flight tube.

2.2.4. Delayed extraction

The laser excitation initializes the desorption/ionization-process. It creates a plume of ions, which may have different initial velocities in the direction towards the detector. The electric field accelerates the ions and increases their initial velocity towards the detector. If the electric field is switched on synchronously with the end of the laser pulse, then the ions of a given m/z all get the same increase in velocity towards the detector because they all pass through the same electrostatic potential difference. Ions of a given m/z value which have a wide variation in initial velocities will have the same wide variation in their final velocity. Such ions show a broad distribution in their TOFs to the detector and hence induce a broad peak in the spectrum.

If the electric field is not switched on synchronously with the end of the laser pulse, but after a certain delay (13), then each ion will fly a certain distance into the acceleration trajectory according to its initial velocity. Ions with a higher initial velocity fly further into the acceleration trajectory. The slower ions are accelerated over a longer distance, which partially compensates for their initial slower velocity. The optimal delay depends on the particles' m/z. More precisely, the optimal delay theoretically is proportional to the square root of the particles' m/z [12], and raises from 1 Da for low masses (∼100 Da) up to more than 50 Da for 100 kDa molecules [13].

2.2.5. Chemical noise

If molecules and intermolecular complexes collide with each other or with the background gas, a small molecule group can split off. This phenomenon is called fragmentation. We now discuss how the position in the flight tube, where the fragmentation of the intact molecule occurs, affects the peak position. If the fragmentation occurs before the acceleration, then each fragment appears in the spectrum at the position, which corresponds to its own m/z. If the fragmentation occurs after the acceleration, then all the fragments appear at the same position in the spectrum. This position corresponds to the m/z of the intact molecule. If the fragmentation occurs during the acceleration, then each fragment appears in the spectrum between the position corresponding to the m/z of the intact molecule and the position corresponding to its own m/z. The fragmentation most often occurs in the free flight tube, i.e. after the acceleration took place [5].

We consider a detected molecule, which was not present as one molecule in the original sample, as "chemical noise" (10). Chemical reactions between different molecules and/or fragmentation cause chemical noise. The chemical reactions can take place between sample protein molecules, matrix molecules, molecules from the washing buffer and molecules from sample impurities. Some of the chemical reactions use energy originating from the laser.

2.3. From detector to spectrum

Basically, the detector measures the period between the moment the electric field switches on and the moment a particle hits the detector. In this paper, we focus on the mass spectra produced by the PBS-II instrument [6] implemented in the Ciphergen ProteinChip System. When molecules strike the first detector plate, the detector plate releases a certain multiple of electrons. The released electrons strike another detector plate, which again releases a certain multiple of electrons. At each detector plate, the multiplication process repeats. The detector gain is defined by the ratio of the eventually released electrons and the number of molecules striking the first detector plate. The detector gain strongly depends on the kinetic energy of the detected molecules [14]. The user can scale the detector gain by a machine setting called 'detector sensitivity' (14) which corresponds to the potential difference between the detector plates. Apart from electrons that are released due to particles striking the detector, there is coincidental electron emission from each plate inside the detector due to thermal energy. This is called 'dark current' or 'electric noise' (15). The molecules fly in a flight tube, which is close to vacuum. Because the flight tube is not completely evacuated, coincidentally air molecules are detected too (16). Dark current and detected air molecules induce a reasonably flat basic signal level, which is unrelated to sample content. The detector counts all the released electrons during a time interval and sends these numbers via a detector signal to the computer. A spectrum comprises many (say 15,000) consecutive time intervals. The user configurable detector frequency, the so-called digitizer rate, (17) determines the length of the time intervals.

2.4. The spectrum

The computer displays the counted totals per time interval in the spectrum. Because the time intervals are small, the spectrum can be interpreted as an almost continuous, 'smoothed' histogram (time versus numbers of electrons). Within the spectrum, we distinguish one baseline and zero or more peaks (definitions follow). Dark current and detected air molecules together form the sample-independent part of the baseline. The chemical noise forms the sample-specific part of the baseline. The detection of molecules, which all have the same molecular formula and occur in the same charge state, induces a signal, which we define as a *singleton peak*. A singleton peak may show overlap with other singleton peaks. We define a *peak* in the spectrum as the signal, which is induced by one or more neighboring singleton peaks together. A (singleton) peak area is assumed to be proportional to the corresponding detected numbers of molecules [1].

2.4.1. The TOF-spectrum and the m/z-spectrum

Theoretically, Eq. (1) converts the TOF-axis to an m/z-axis by $(m/z) = (2\,\mathrm{eV}/v^2) = (2\,\mathrm{eV})/(x/\mathrm{TOF})^2 = (2\,\mathrm{eV})/(x^2)\mathrm{TOF}^2$. However, in practice, linear deviations from the expected relation are observed (18). A calibration equation, which deals with the linear deviation by inserting the extraction delay (t_0) and the calibration parameters ($\tilde{\alpha}$ which we use as a temporary dummy variable, and β), is $(m/z) = \tilde{\alpha}(2\,\mathrm{eV}/x^2)(\mathrm{TOF} - t_0)^2 + \beta$. Substitution of $\tilde{\alpha}(2e/x^2)$ by α (i.e. a new calibration parameter which replaces $\tilde{\alpha}$) gives the calibration equation: $(m/z) = \alpha V(\mathrm{TOF} - t_0)^2 + \beta$. Measuring the TOFs of molecules with known masses yield the calibration parameters α and β. Different spectra may have different α and β [15]. Ciphergen allows the user to consider t_0 as an extra calibration parameter, which can be estimated together with α and β from the data. A more detailed discussion about the calibration applied by Ciphergen, and also algorithms for the alignment of mass spectra, can be found in ref. [15]. Conversion of the horizontal axis from TOF to m/z according to the calibration equation changes the areas under the curves. On the one hand, at a given peak position on m/z-scale, the areas are still proportional to the numbers of molecules when comparing different spectra. On the other hand, at different peak positions within one spectrum, say one in the lower and another in the higher m/z-range, detecting the same numbers of molecules results in different peak areas [2]. Therefore, for quantifying (detected amounts of) molecules, the data are preferably analyzed on the TOF-scale, and not on the m/z-scale. The next section identifies baseline and peaks in a practical example spectrum. The following subsection discusses some remaining sources of variation.

2.5. Other sources of variation

2.5.1. Post-translational modifications

Molecules form intermolecular complexes by non-specific binding. Binding can also be biology specific. Post-translational modifications (19) play an important role in biological pathways. Phosphorylation, for example, is a chemical modification of certain amino acids that can switch the activity of a protein on and off. Therefore, the amount of phosphorylation for a certain protein can provide important information about the state of a cell. This modification adds 80 Da to the mass of the protein, and thus leads to a shift in peak position [16].

2.5.2. Averaging spectra

A final spectrum is an average of spectra acquired by several individual laser shots. Each single laser shot is fired on a different position on the dried droplet. The crystal density and size varies with the position within the dried droplet (20). The Ciphergen software automatically removes shots, which generate a signal that is too high or too low [6].

3. A practical example

3.1. Sample preparation

Myoglobin (16,951 Da), a single-chain protein of 153 amino acids, is the primary oxygen-carrying pigment of muscle tissues. Myoglobin can contain a non-covalently linked heme group (616 Da) in its center [7]. Denatured myoglobin releases the heme group. We applied non-denatured as well as denatured myoglobin to different spots on an NP20-chip. The NP20-

chip mimics normal phase chromatography [6]. After a short incubation period, we washed the chip twice with buffer and once with distilled water and applied the 'matrix molecules', a saturated sinapinic acid (SPA, 224 Da) solution, to the chip. After drying, we put the chip in the vacuum chamber, hit the dried matrix with a laser (220 μJ) and applied an electric field (10 kV) in positive ion mode after a short delay of ∼1000 ns. This delay was chosen based on myoglobin's molecular weight, according to the Ciphergen software [6]. Thirty-five shots were averaged in the final spectrum, which is shown in Fig. 1. The final spectrum does not include the two warming shots with an increased laser energy of 5 μJ. We used a detector sensitivity of 9 and a digitizer rate of 250 MHz.

3.2. Peak positions

The m/z-locations of the peaks in the spectrum correspond to m/z-values of the intermolecular complexes formed. Let the non-negative variables m_0, m_1 and m_2, respectively, denote the numbers of myoglobin, SPA and heme molecules that comprise the complex. The set $\{m_0 \times 16{,}951 + m_1 \times 224 + m_2 \times 616 + p \times 1\}$ Da describes the masses of possible complexes, where the positive variable p denotes the number of extra protons, obtained from the matrix. The set $\{(m_0 \times 16{,}951 + m_1 \times 224 + m_2 \times 616 + p \times 1)/p\}$ Da describes the m/z-values of these complexes. However, fragments of these complexes, salty adducts and various atomic isotopes can generate peaks as well. In our data, we see the satellite peaks at about +206 Da to the right of the myoglobin peak, instead of at the described +224 Da, which is the molecular mass of the SPA molecules. Perhaps the acid forms a covalent ester bond with the protein alcoholic side chains, or perhaps the acid forms a covalent peptide bond with the amide of the protein via a condensation reaction. In either case, a water molecule (18 Da) is one of the products. Hence, 224 is better substituted by 206 in the formula above: $\{(m_0 \times 16{,}951 + m_1 \times 206 + m_2 \times 616 + p \times 1)/p\}$ Da. We now use Fig. 1 to elucidate on these peaks and we show that some complexes are more likely to occur than others.

The singly charged myoglobin monomers (i.e. $m_0 = p = 1$) cause the peak at $m/z = 16{,}952$ Da. The singly charged m_0-mers ($m_0 = 2, \ldots, 5$ and $p = 1$) cause peaks at $m_0 \times 16{,}951 + 1$ Da. Peaks for $m_0 > 5$ exceed the mass axis. The doubly charged myoglobin molecules ($m_0 = 2, \ldots, 10$ and $p = 2$) cause peaks at $((m_0 \times 16{,}951 + 2)/2)$ Da. Doubly charged m_0-mer peaks have almost the same m/z value as singly charged $(m_0/2)$-mer peaks, for even values of m_0, respectively. Such peaks will show large overlap.

Fig. 2a shows that the singly charged myoglobin proteins (i.e. $m_0 = p = 1$) occur in complex form, linked to $m_1 = 1, 2, 3, \ldots$ (reactive forms of the) SPA molecules. Myoglobin which is non-covalently linked with a heme group, occurs as a peak at the position corresponding to $m_0 = m_2 = 1$, see, e.g. Fig. 2b. The non-covalent linking might be both biologically specific as it is in vivo, or non-specific like the matrix adducts we observe.

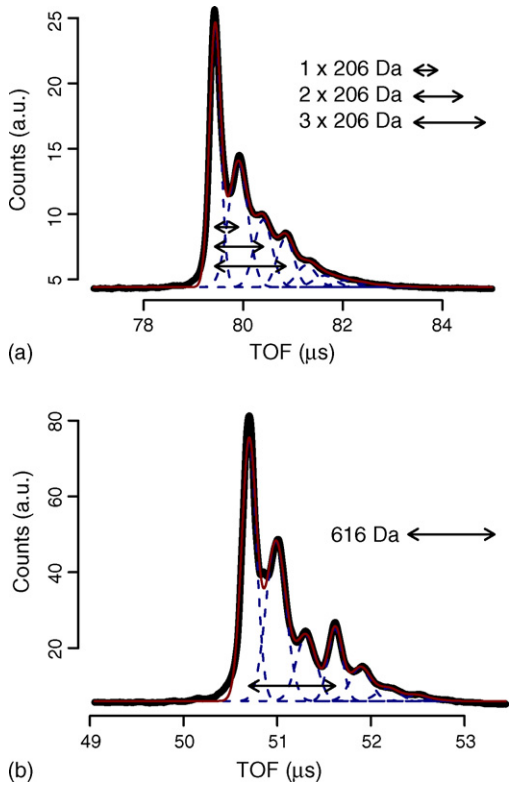

Fig. 2. Mixture model analysis of myoglobin and its satellite peaks. We analyzed the same myoglobin sample in two different runs and show the corresponding singly charged peak in (a) and (b). Myoglobin (16,951 Da) can (i) form complexes with m reactive matrix molecules ($m \times 206$ Da) (a), generating peaks at 16,951 Da + $m \times 206$ Da, and/or can (ii) contain a non-covalently bound heme group (616 Da) in its center [7] (b). The black line is the observed data. The red/brown curve is the fitted mixture model and the blue dashed curves are the fitted individual components.

3.3. Peak areas

The area of a peak is proportional to the corresponding number of detected ions within a given spectrum [1]. The mutual repulsion between ions with charge of equal sign causes the ion cloud to expand during its flight. We expect that the numbers of detected ions approximate the numbers of ions, which were formed during the desorption/ionization-process, with a certain trade-off for ions with a larger m/z (because these have a larger TOF and their ion cloud expands more, so more ions miss the detector and are not counted). Overlap between adjacent peaks complicates the estimation of individual peak areas. Section 4 presents a method to resolve overlapping peaks by using normal distributions. In that section, we use the areas of our spectral components to quantify described sources of intraspectral and of interspectral variation. Section 4.4 shows how the peak areas are related.

3.4. Peak shape

Singleton peaks can be slightly skewed to the right due to the expansion of the ion cloud and due to the isotopic distribution. A (singleton) peak is induced by an ion cloud. If all molecules within an ion cloud travel at constant speed towards the detector

and the molecule density is point symmetric around the center of the ion cloud, then the generated peak would be symmetric around its mean. However, molecules within one ion cloud travel with different velocities towards the detector and the ion cloud expands. An ion cloud, which expands more during detection induces more right skewness in the peak shape. The detection of ion clouds takes longer if the ion clouds have a longer TOF, such that peak skewness increases with TOF. The authors of ref. [17] present the peak shape as a function of the expanding ion cloud. However, Fig. 2 shows that the singleton peaks, which we analyzed did not show a large deviation from the symmetric normal distribution.

The isotopic distribution is another minor cause of singleton peak right-skewness. Except for the lower TOF-range (shown in Fig. 3), SELDI resolution is too low to observe the individual isotopic peaks. In other words, the individual isotopic peaks occur too close to each other (i.e. at mass differences of 1 Da at *m*/*z*-scale) to be individually observed. After longer TOFs the individual "isotopic" ion clouds, which travel close to each other due to their small mass differences, expand and overlap. These isotopic ion clouds then cause overlapping isotopic peaks, which together generate one 'singleton peak'. The isotopic distributions are skewed to the right for low molecular weight peaks, but tend quickly to a normal distribution for the higher masses. Hence, the isotopic cluster of a given molecule species generates one apparent singleton peak in the spectrum, which can be modeled with one single normal distribution.

A lower voltage setting leads to longer TOFs (cf. Eq. (2)) which leads to a larger separation between the centers of ion clouds and thus to better separation between different ions. On the other hand, a "singleton" ion cloud expands more if it has a longer TOF. Both larger ion clouds and a larger separation between those clouds lead to broader peaks. Fig. 4 shows that the overlap between the singleton satellite peaks caused by singly charged complexes comprising exactly one myoglobin molecule increases for lower voltage settings (9–25 kV shown). The overlapping singleton peaks together induce one single right-skewed peak in the spectrum.

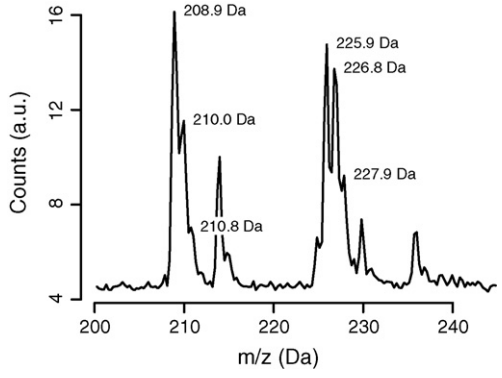

Fig. 3. Isotopic resolution in the lower *m*/*z*-range. We analyzed a new sample comprising the two commonly used matrix molecule species, alpha-cyano-4-hydroxy cinnamic acid and sinapinic acid. The close-up of the low molecular mass range illustrates the most elementary singleton peaks which can occur in mass spectra, i.e. the isotopes of the molecule, having different numbers of neutrons (1 unit mass, no charge). The theoretical mass difference between two adjacent isotopes is 1 Da. However, the overlap between adjacent (isotopic) peaks causes slight mass shifts in the corresponding local modes. We labeled some local modes with the corresponding mass position in the spectrum.

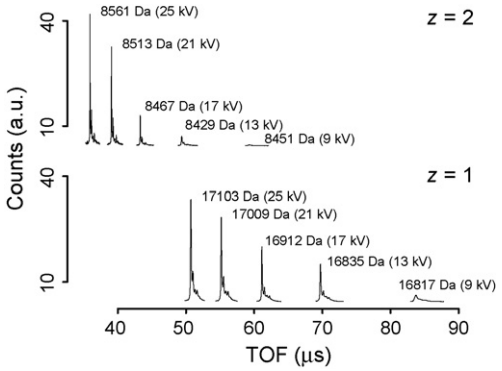

Fig. 4. The applied electric field potential. We measured the chip with the myoglobin sample several times and applied a different field potential in each run. We cut the doubly charged myoglobin peak from each spectrum and show them together in one "spectrum", above ($z = 2$), and we show the singly charged myoglobin peaks below ($z = 1$). The peak labels indicate the corresponding *m*/*z*-position (Da), and the applied field potential (kV). A higher voltage leads to faster ions, which have a shorter TOF to the detector and induce more electrons when striking the detector. This leads to larger peak areas. Ion clouds, which have a longer TOF expand more and generate broader and lower peaks; hence, the singleton peaks show less overlap if the applied voltage is higher. Surprisingly, the size of the doubly charged peak increases more rapidly than the size of the singly charged peak, when applying a higher voltage. The peak labels show that a higher applied voltage leads to a larger mass estimate. The machine should be recalibrated when changing the voltage settings.

3.5. Baseline

Dark current and detected air molecules contribute relatively little extra variation to the signal. We assume that this variation does not depend on TOF. It can be observed in a region, which does not contain peaks caused by detected ions. Detected matrix molecules and detected molecule fragments cause chemical noise. Fig. 1 does not show much chemical noise. Fig. 5 displays myoglobin spectra, which are measured with different laser energies applied. It shows that the chemical noise is more abundant when higher laser energies are applied. Probably there is more fragmentation. The chemical noise is more abundant in more complex samples comprising many different protein species due to fragmentation and metastable decay due to phenomena like collisions with the background gas [4,5]. The chemical noise shows exponential decay with TOF [2].

3.6. The machine parameters

Fig. 4 shows that a higher applied voltage (which creates an electric field with a larger potential difference) generates faster ions, which have a shorter TOF to the detector and induce more electrons when striking the detector, leading to higher peaks. Ion clouds, which have a longer TOF expand more and generate broader and lower peaks. Surprisingly, the height of the doubly charged peak increases more rapidly than the size of the singly charged peak.

The detector with higher sensitivity settings generates more electrons when detecting the same amounts of ions, leading to

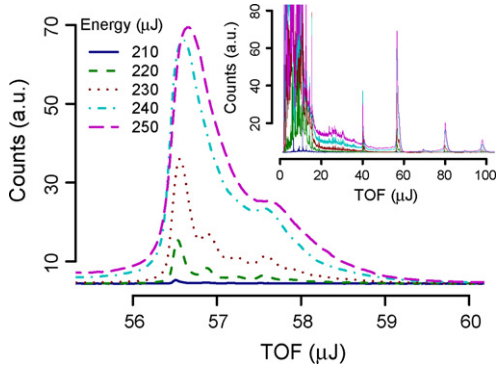

Fig. 5. The applied laser energy. We measured the same myoglobin sample five times with different laser energies (210, 220, . . ., 250 µJ) in each run. Spectra which are acquired with higher laser energies have a larger area under the curve, because applying higher laser energies generates more ions during the desorption/ionization-process; a 210 µJ laser pulse generates almost no ions which leads to small peaks, while a 250 µJ laser pulse generates many ions which leads to large peaks with low resolution. Higher laser energy increases the thermal energy of the ions, resulting in more and harder collisions between the ions and increasing fragmentation and chemical noise. The increased thermal energy expresses as an increased kinetic energy, resulting in higher initial velocities. The tail at the left side of the peak at 80 µs, which is best observed in the 250 µJ spectrum, might be explained by the short extraction delay (see text for details).

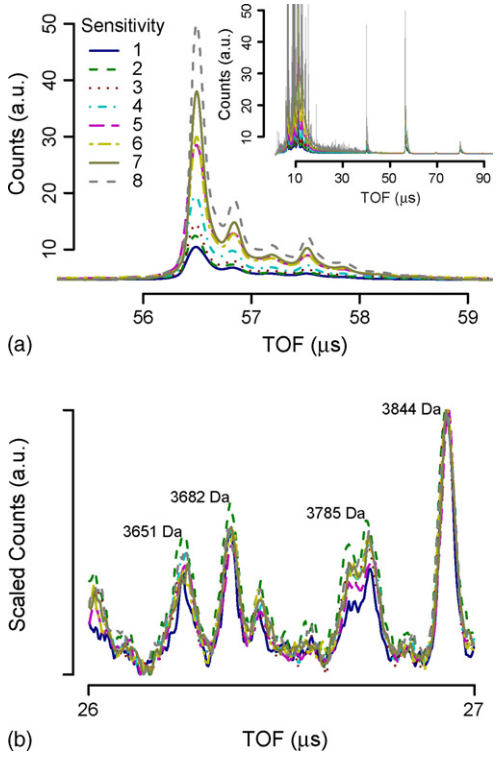

Fig. 6. The detector sensitivity settings and chemical noise. We measured the same myoglobin sample eight times, with different sensitivity settings (1, 2, . . ., 8) in each run. (a) Shows an overview of the full spectrum and a close-up of the singly charged myoglobin peak. Higher sensitivity settings generate larger peaks and does not affect the resolution considerably. The overview shows many reproducible peaks between 20 and 40 µs. (b) Shows a close-up of the spectra between 26 and 27 µs, acquired with the above settings. The *y*-axes were scaled to enable a direct comparison. Some of the reproducible peaks have a mass label. Although these peaks probably originate from our sample molecules, we cannot explain their positions easily by fragmentation, complex formation and multiple charges. We consider these peaks as chemical noise.

higher peaks, without considerably affecting peak shape and peak resolution (Fig. 6a).

The energy of the laser is the only machine parameter which enormously affects the desorption/ionization-process and the generated number of ions; a 210 µJ laser pulse generates almost no ions which leads to small peaks, while a 250 µJ laser pulse generates many ions which leads to large peaks with low resolution (Fig. 5). Higher laser energy increases the thermal energy of the ions, resulting in more and harder collisions between the ions and increasing fragmentation and chemical noise.

The time lag between the end of the laser pulse and the moment the electric field switches on is called the extraction delay. Only one position in the spectrum has optimal resolution. A longer delay shifts this position to the right (ratio explained above), and leads moreover to smaller peaks, probably because some of the ions are accelerated over a shorter distance and get less kinetic energy (Fig. 7).

Fig. 8 plots the area of the singly charged myoglobin peak with default machine settings, and varying one setting at a time from low to high. The largest peak was scaled to 100% and the other peaks were scaled accordingly. The observed impact of the machine settings on the peak area are: (i) the laser energy (0–100%), (ii) the sensitivity setting (4–26%), (iii) the extraction delay (2–17%) and (iv) the applied voltage (3–14%).

Moreover, Figs. 4, 5 and 7 show that changing the machine settings: (i) the applied voltage, (ii) the applied laser energy and (iii) the extraction delay, assign other *m*/*z*-values to the "same peak"; hence, the machine should be recalibrated when changing these machine settings. We give two examples. The singly charged peak in Fig. 4, measured with 25 kV gets the label 17,103 Da, while the peak measured with 9 kV gets the label 16,817 Da, which is a difference of 286 Da. The same peak in Fig. 7 gets the label 16,920 Da when the delay is 241 ns, and 17,096 Da when the delay is 2007 ns, which is a difference of 176 Da.

The repeated measurements in Fig. 6b show a close-up (26–27 µs) of the reproducible peaks, which we observe in the

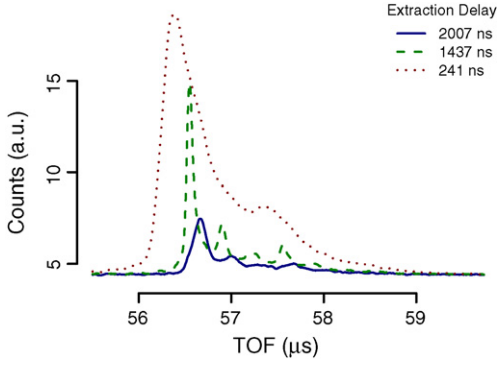

Fig. 7. The extraction delay. The time lag between the end of the laser pulse and the moment the electric field switches on is called the extraction delay. Only one position in the spectrum has optimal resolution. A longer delay shifts this position to the right (ratio explained in the text), and leads moreover to smaller peaks, probably because some of the ions are accelerated over a shorter distance and get less kinetic energy. The machine should be recalibrated when changing the sensitivity settings.

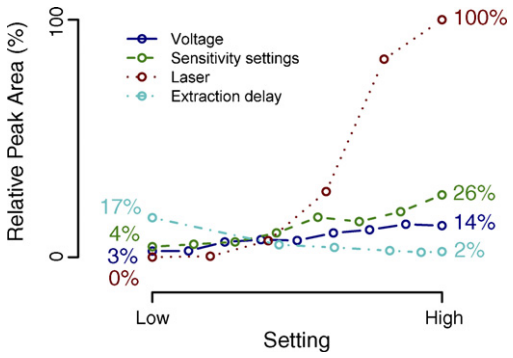

Fig. 8. Machine parameters and peak area. The plot shows the area of the singly charged myoglobin peak with default machine settings, and varying one setting at a time from low to high. The largest peak was scaled to 100% and the other peaks were scaled accordingly. The observed impact of the machine settings on the peak area are: (i) the laser energy (0–100%), (ii) the sensitivity setting (4–26%), (iii) the extraction delay (2–17%) and (iv) the applied voltage (3–14%).

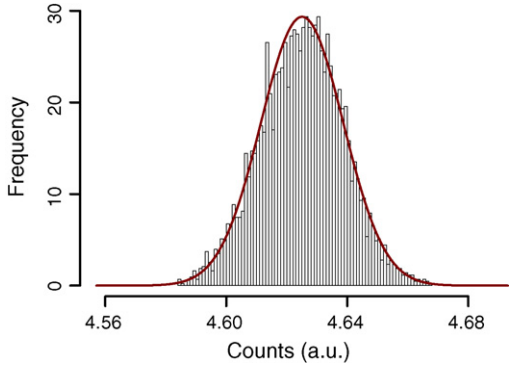

Fig. 9. Normal distributed detector noise. Electric noise and coincidentally detected air molecules from the free flight tube contribute little to the overall variation in the signal. The histogram of the detector signal in a region where we expect no protein peaks, is approximated with a normal distribution with mean 4.6 and indeed a very small variance <0.0002.

region (20–40 μs). These peaks are probably caused by fragments of sample molecules, which we call chemical noise.

Electric noise and coincidentally detected air molecules from the free flight tube contribute little to the overall variation in the signal. Fig. 9 approximates the histogram of the detector signal in a region where we expect no protein peaks, with a normal distribution with mean 4.6 and indeed a very small variance <0.0002.

4. Analyzing our practical example

This section uses statistical mixture models to quantify the major sources of the variation of the spectral components (described in Section 2.4) in our practical example data (introduced in Section 3). The first subsection below introduces the statistical mixture model analysis of the spectrum. The second subsection defines our mixture model components and the third subsection shows to fit our model to the spectrum. The fit of the mixture model produces estimates of the positions and the proportions of the spectral components, including peak positions and peak areas. The fourth subsection quantifies the variation in the positions and the proportions of the spectral components and connects these variations to the associated sources.

4.1. Simulation and analysis model

We study spectra with the x-axis on TOF-scale. Such a spectrum represents the observed TOFs of all the detected molecules, or more precisely, it represents the observed TOFs that correspond to the counted electrons. Each of the counted electrons originates either from a detected molecule species, or from dark current. We assume that the TOF that corresponds to a counted electron follows a probability density distribution. The origination of the electron determines the exact distribution it follows. We assume that the TOF of a detected molecule species (e.g. a myoglobin monomer) or complex (e.g. a myoglobin dimer) is derived from a unique normal distribution. Additional observations (i.e. due to dark current, detector noise) are present with more-or-less equal probability of occurrence along the spectrum, which we hence model with a uniform distribution. We choose and justify these distributions based on empirical considerations, i.e. the proposed distributions fit well to the observed ones, as illustrated in the practical data (e.g. Fig. 2). The origination of a counted electron cannot be observed with absolute certainty. So, the observed TOF is a finite mixture of multiple distributions. The next section shows how we describe a spectrum with a mixture model. Our model adequately simulates the expected spectrum based on known sample content, and in the reverse mode, adequately fits the observed spectrum.

4.2. Mixture components

We now define the individual component distributions and their mixture distribution. The spectrum can be interpreted as a histogram with times of flight, say $t_1, t_2, \ldots, t_I$, ordered from short (left) to long (right) on the x-axis. Let $n_1, n_2, \ldots, n_I$, denote the counted numbers of electrons which arrived after the times of flight $t_1, t_2, \ldots, t_I$, respectively. The uniform distribution is defined by

$$f_0(t) = \frac{1}{t_1 - t_I}, \quad \text{for} \quad t_1 \le t \le t_I.$$

Suppose that the spectrum contains M peaks. The normal distributions j ($j = 1, 2, \ldots, M$) are defined by

$$f_j(t) = \frac{1}{\sigma_j\sqrt{2\pi}} \exp\left(-\frac{(t-\mu_j)^2}{2\sigma_j^2}\right),$$

where μ_j (mean) and σ_j (standard deviation) are indexed by peak number. The finite mixture distribution is described by

$$f(t) = \sum_{j=0}^{M} p_j f_j(t)$$

where $f_j(\cdot)$ is the probability density function of the jth component. A distribution f_j ($j = 0, 1, 2, \ldots, M$) occurs with a proportion $p_j \in [0, 1]$, and $\sum_{j=0}^{M} p_j = 1$, such that the area under our finite mixture distribution equals one. Superimposing the

individual distributions and their mixture distribution on top of the spectrum after multiplication with the area under the spectrum, shows the fit of the model to the spectrum, see p. 450 in ref. [18].

4.3. Parameter estimation

Section 3.2 predicts the expected peak positions on m/z-scale. We use the calibration equation to calculate the corresponding peak positions on TOF-scale and use these to initialize the parameters μ_j ($j = 1, 2, \ldots, M$) as described in ref. [2]. We initialize $\sigma_j = 0.1$ ($j = 1, 2, \ldots, M$), and $p_j = (1/(M+1))$ ($j = 0, 1, 2, \ldots, M$). We now briefly describe how we iteratively apply the EM-algorithm [19] to estimate the maximum likelihood values for the parameters in our finite mixture model. Each iteration consists of an E- and an M-step.

The E-step calculates the conditional component membership probabilities, $p_{j|i} = p_j\, f_j(t)/f(t)$ for $j = 0, \ldots, M$, given the current parameter estimates. The M-step calculates the updated means by

$$\hat{\mu}_j = \frac{\sum_{i=1}^{I} n_i p_{j|i} t_i}{\sum_{i=1}^{I} n_i p_{j|i}},$$

and uses the newly obtained means to calculate the updated variances by

$$\hat{\sigma}_j^2 = \frac{\sum_{i=1}^{I} n_i p_{j|i} (t_i - \hat{\mu}_j)^2}{\sum_{i=1}^{I} n_i p_{j|i}}, \quad \text{for} \quad j = 1, \ldots, M.$$

Finally, the M-step calculates the updated proportions by

$$\hat{p}_j = \frac{\sum_{i=1}^{I} n_i p_{j|i}}{\sum_{i=1}^{I} n_i} \quad \text{for} \quad j = 0, 1, \ldots, M-1, \quad \text{and}$$

$$\hat{p}_M = 1 - \sum_{j=0}^{M-1} \hat{p}_j.$$

Fig. 2 shows a fit of our model to the singly charged myoglobin peak and its satellite peaks. We ran the EM-iteration process until convergence of the parameter estimates.

4.4. Quantifying sources of variation

This section uses our mixture model analysis to quantify the sources of variation, which we identified in Section 3. The expected peak positions have already been discussed in Section 3.2. We first study the relation between the areas of the peaks in the spectrum. We next elucidate on the quantification of biological modifications. Then, we discuss the effect of peak broadening due to the isotopic distribution. This section ends by a quantification of the detector noise and some notes about the effect of different machine settings.

Section 3.3 explained that we expect the peak area within a given spectrum to be approximately proportional to the corresponding detected number of ions, i.e. the charged complexes formed during the desorption/ionization-process. We analyzed the areas of two series of peaks in the spectrum displayed by

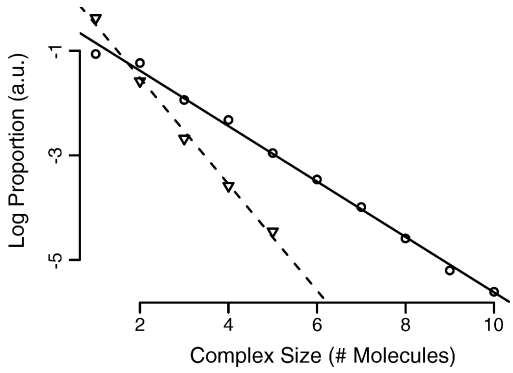

Fig. 10. Log-linear peak areas. The five triangles represent the peaks which correspond to the m-mers of myoglobin for $m = 1, 2, 3, 4$ and 5. The 10 circles represent the 10 peaks, which correspond to the monomer (i.e. 1-mer) of myoglobin and to the nine satellite peaks that we analyzed (cf. Fig. 2). The considered peak area decays log-linear as a function of the considered number of molecules that form the complexes corresponding to the peak. The areas of the series of m-mer peaks were rescaled and the areas of the series of the monomer and its satellite peaks were rescaled such that both series have the same sum of the peak areas.

Fig. 1. The m-mers ($m = 1, 2, \ldots, 5$) form the first series of peaks. The five triangles in Fig. 10 represent these five m-mers. Fig. 2a shows the singleton peaks, which together form the singly charged myoglobin peak. The circles in Fig. 10 represent these singleton peaks, which form the second series we analyzed. Fig. 10 shows that the proportions of the m-mers show a log-linear decay with m. And, that the proportions of the second series of peaks show log-linear decay with the total numbers of molecules, which comprise the complex. The lines show different slopes. Firstly, our sample contains many more SPA molecules than myoglobin molecules. Secondly, it might be more likely that a myoglobin molecule, which is relatively large, forms a complex with one of the many SPA molecule which are relatively small, than that it forms a complex with one of the other less abundant and large myoglobin molecules. These two issues might explain why the second series shows a steeper slope in Fig. 10.

We analyzed the same myoglobin sample twice, in different runs. Myoglobin (16,951 Da) formed complexes with m reactive matrix molecules (206 Da), generating peaks at 16,951 Da + $m \times 206$, for $m = 0, 1, 2, \ldots$ (Fig. 2a). Fig. 10 shows that the singleton peak areas decay log-linearly as function of m. Fig. 2b shows the spectrum obtained in the second run, with the fourth singleton peak being larger than expected on the basis of Fig. 2a. This peak corresponds to myoglobin linked to a 616 Da compound, which probably is a heme-group. The myoglobin-heme complex has a mass which is almost equal to the myoglobin-m-matrix ($m = 3$) complex, because 3×206 Da = 618 Da. We found indeed a peak at 616 Da, which was larger in sample 2a, where it exceeded the detection limit, than in sample 2b. Our mixture model analysis should enable the quantification of the biological modification, e.g. by analyzing the deviation of the expected proportion of the considered singleton peak.

Isotopic peaks are distinguishable in the lower mass region (Fig. 3). In the higher mass region, peaks become negligibly

broader due to the isotopic variation in comparison with the effect of the expanding ion cloud.

Fig. 9 shows a histogram of the signal intensities, which were measured in a region without protein peaks. The normal distribution (red line) fits well to the histogram. The protein signal increases with the machine parameter, detector sensitivity. Fig. 4 shows the effect of different applied voltages. Fig. 6 shows the effect of the sensitivity settings on the singly charged myoglobin peak. Fig. 5 illustrates the effect of laser energy. Each of the settings induces a larger area under the curve, when higher settings are applied. However, only the applied laser energy affects the numbers of ions. Our paper does not address variation due to chips with different surface types, different positions on the surface, and reproducibility issues.

5. Discussion

Table 1 summarizes the discussed sources of variation (1)–(20) and their impact on the spectrum. We analyzed our experimental data for the sources (5) laser energy, (8) formation of intermolecular complexes, (10) fragmentation of molecules and intermolecular complexes, (12) electric field potential, (13) delayed extraction, (14) detector sensitivity, (15) electric noise, (16) detector noise, (18) *m/z*-calibration and (19) post-translational modifications.

The many peaks in complex SELDI mass spectra might be described in a parsimonious way by only a few major proteins. We show, as is commonly known, that one protein species can generate about 10 major peaks and many minor satellite peaks. Adding another protein species to our sample might add another 10 major peaks and many minor peaks to our spectrum, which are directly attributable to the new protein. However, if the two proteins can bind to each other, many more peaks will show up. In theory, the number of possible peaks increases exponentially with the number of protein species in the sample. Affinity between different proteins can be biologically specific. Analyzing peaks, which are generated by complexes comprising two proteins of two different species might indicate the affinity between the two species. However, their affinity under SELDI experimental conditions might not represent their affinity in vivo.

Proteins consist of H, C, N, S and O atoms which all have a known isotopic mass distribution. Singleton peaks can be further decomposed in isotopic peaks by using our approach, which should enable the accurate prediction of the atomic composition of the proteins. However, it remains to be seen whether the resolution in SELDI spectra is high enough to do so.

Sample and matrix molecules which form complexes and/or carry multiple charges do not well explain the many reproducible peaks which we observe; e.g. the peaks between 20 and 40 μs in Fig. 6. MALDI is generally characterized by little prompt fragmentation [5]. Hence, we might expect that these reproducible peaks are mainly caused by sample impurities and by molecules which are formed by chemical reactions which frequently take place on the chip. The chemical reactions can take place between sample protein molecules, matrix molecules, molecules from the washing buffer and molecules from sample impurities. We now give a hypothesis for the left tail of the 80 μs peak. Higher laser energy increases the thermal energy and gives ions a higher initial velocity. We applied a very short extraction delay. Hence, even the fast myoglobin ions passed through a large part of the field, ending up with a higher initial velocity and arriving earlier at the detector. A longer extraction delay would probably slow down the faster ions, diminishing the left tail.

Our model uses three parameters per (normal) component (i.e. μ, σ and p). Reducing the number of parameters improves the robustness of the estimation procedure but can give a worse fit. Combining parameters might be considered when analyzing complex spectra. The position parameters, μ, could be connected by making use of the known relations between the peak positions in the spectrum as described (cf. Section 3.2, e.g. if μ is the position of the monomer, then 2μ is the position of the dimer). The satellite peaks of a given protein peak can, for example, be combined in a simple relation because the satellite peaks are known to occur equidistant on mass scale (at +206 Da in our examples). We observed log-linear relationships between the proportion parameters, p. The parameter σ can be considered as a monotonously increasing function of TOF because it is related to the size of the ion clouds, which depends mainly on TOF. Peaks, which get a larger σ than expected based on the relation of σ with TOF, might indicate that the complexes, which correspond to the peak occur in multiple biological variants with slight mass differences. Myoglobin, for example, can be oxidized (an O atom has a mass of 16 Da) multiple times (m) which results in extra peaks at $+m \times 16$ Da with respect to the main myoglobin peak [8]. The oxygen adduct peaks would show strong overlap in SELDI spectra.

In ref. [2], we analyze multiple SELDI spectra (8 from adipose tissue and 64 from serum) and use log-normal distributions to fit peaks. Our current paper shows that the peak can be further deconvoluted with mixtures of normal distributions (at the price of many more components). The next step is to build models with interrelationships between μ's and σ's and p's explicably incorporated. It remains to be seen whether complex spectra can then be "corrected" for the main "families" of peaks arising from single molecules/complexes.

We believe that our study of sources of variation and our computational methods are of great value to biomarker discovery for the following reasons: (i) proteins with approximately the same mass will show-up with overlapping peaks in SELDI-TOF spectra. Our methods developed in ref. [2] and in the current manuscript make it possible to correctly deconvolute the spectrum, which will result in more accurate and reliable estimates of protein abundance. (ii) A given protein can show up at multiple locations in a SELDI-TOF spectrum as shown in Fig. 1. Our methods developed in the current manuscript make it possible to deconvolute a spectrum in a more meaningful way by appropriately linking the peaks that correspond to the same protein. We have shown how to do this for relatively simple samples, and anticipate that the same strategy will also work in more complex mixtures. (iii) SELDI-TOF spectra obtained with different machine settings can look quite differently as demonstrated in Figs. 4–7. Our preliminary experiments suggest that a laser pulse of 230 μJ (Fig. 5) and an extraction delay of 1437 ns (Fig. 7)

give the best peak resolution. Further research, preferably using a multifactorial statistical design for experimentation, should confirm or improve these settings.

We strongly believe that better pre-processing of data from SELDI-TOF spectra generates more accurate and reliable estimates of protein abundance, which in turn will lead to more reliable and powerful biomarker discovery.

Acknowledgement

We thank Bart Charbon for the laboratory work.

References

[1] M. Merchant, S.R. Weinberger, Electrophoresis 21 (2000).
[2] M. Dijkstra, H. Roelofsen, R.J. Vonk, R.C. Jansen, Proteomics 6 (2006) 19.
[3] N. Tang, P. Tornatore, S.R. Weinberger, Mass Spect. Rev. 23 (2004).
[4] M. Müller, Molecular scanner data analysis, Doctoral dissertation, University of Geneva, 2003.
[5] R. Zenobi, R. Knochenmuss, Mass Spect. Rev. 17 (1998).
[6] ProteinChip Software 3.1 Operation Manual, Fremont, CA Ciphergen Biosystems, Inc., 2002 (www.chiphergen.com).
[7] V.W.S. Lee, Y.L. Chen, L. Konermann, Anal. Chem. 71 (1999) 19.
[8] I.A. Kaltashov, S.J. Eyles, Mass Spect. Rev. 21 (2002).
[9] I.A. Kaltashov, A. Mohimen, Anal. Chem. 77 (2005) 16.
[10] V. Livadaris, J.C. Blais, J.C. Tabet, Eur. J. Mass Spectrom. 6 (2000).
[11] R.K. Chitta, M.L. Gross, Biophys. J. 86 (2004).
[12] M. Vestal, P. Juhasz, J. Am. Soc. Mass. Spectrom. 9 (1998).
[13] M. Hilario, A. Kalousis, C. Pellegrini, M. Müller, Mass Spect. Rev. 25 (2006).
[14] M. Hellsing, L. Karlsson, H.-O. Andren, H. Norden, J. Phys. E: Sci. Instrum. 18 (1985).
[15] N. Jeffries, Bioinformatics 21 (2005) 14.
[16] S. Vorderwülbecke, S. Cleverley, S.R. Weinberger, A. Wiesner, Nat. Methods 2 (2005) 5.
[17] I.I. Stewart, J.W. Olesik, Am. Soc. Mass Spectrom. 10 (1999).
[18] R.C. Jansen, in: D.J. Balding, M. Bishop, C. Cannings (Eds.), Handbook of Statistical Genetics, second ed., Wiley, West Sussex, 2003, p. 445.
[19] A.P. Dempster, N.M. Laird, D.B. Rubin, J. R. Stat. Soc. Ser. B—Methodol. 39 (1977).