Detection of biomarkers for lysosomal storage disorders using novel technologies
van Breemen, M.J.

Citation for published version (APA):

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

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter Twelve

Limited value of SELDI-TOF serum protein profiling for discrimination of patients suffering from Fabry disease

Mariëlle J. van Breemena, Suzanne Smitb, Anouk C. Vedderc, Huub C.J. Hoeslooth, Age K. Smildeb, Chris G. de Kosterd, Carla E.M. Hollakc, Johannes M.F.G. Aertsd

a Department of Medical Biochemistry, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
b Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands
c Department of Internal Medicine, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
d Mass Spectrometry of Biomacromolecules, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands
Abstract

Fabry disease is an X-linked lysosomal storage disorder, due to a deficiency in α-galactosidase A. Accumulation of globotriaosylceramide in the endothelium is thought to cause onset and manifestations of Fabry disease. Currently no blood biomarker is available that reflects the clinical manifestation. Directed searches in plasma or serum of Fabry patients for markers of endothelial cell activation have given negative results. To find a biomarker, we compared serum of controls and Fabry patients using SELDI-TOF MS, an approach that earlier allowed classification of patients suffering from Gaucher disease, another lysosomal storage disorder. SELDI-TOF MS serum profiles of symptomatic Fabry patients and control subjects were classified using principal component discriminant analysis (PCDA) and support vector machines (SVM). Distinction between Fabry patients and controls using PCDA showed high error rates, also after variable selection. With SVM, the prediction error was lower. The permutation test showed that the classification result is significant, but the misclassification rate is still 16%. Of note, healthy family members from Fabry patients were misclassified, suggesting that not a true disease-specific classification is made. In conclusion, our study failed to detect useful discriminatory differences between Fabry and control SELDI-TOF MS serum profiles.
Introduction

Fabry disease (McKusick 301500) is an X-linked lysosomal storage disorder. Deficient activity of α-Galactosidase A leads to accumulation of glycosphingolipids (mainly globotriaosylceramide, Gb3) in lysosomes [1,2]. Extensive storage occurs in arterial walls, in particular in endothelial cells. This accumulation is believed to underlie the clinical manifestations in Fabry disease: progressive renal insufficiency, cardiac infarction or hypertrophy, arrhythmias and cerebral infarctions [3].

Clinical observations reveal a high incidence of thrombosis in Fabry disease patients [4] and in mouse models [5,6]. In addition, based upon case histories [7-9] and a study in mice [10], an association between α-Galactosidase A deficiency and the early development of atherosclerosis has been suggested, though a more recent study revealed an increased carotid intima-media thickness in the absence of atherosclerosis in Fabry disease patients [11]. Laboratory investigations that have been performed to assess determinants of coagulation or activation of the endothelium are not always in accordance. Elevated levels of soluble sICAM-1, sVCAM-1, P-selectin, plasminogen activator inhibitor (PAI) and decreased thrombomodulin [12] suggest a prothrombotic profile in patients with Fabry disease, although only an elevated level of sVCAM-1 could be confirmed by Demuth et al. [13]. In a very recent study conducted with a large cohort of Fabry patients in the Academic Medical Center in Amsterdam, only minimal abnormalities in indicators of coagulation, fibrinolysis and platelet activation as well as endothelial activation were detected (Vedder, A.C., Biró, É., Aerts J.M.F.G., Nieuwland, R., Sturk, A. and Hollak, C.E.M., unpublished data). Very severely affected patients with renal impairment formed an exception in this respect. The noted plasma abnormalities in these individuals might be ascribed to their renal insufficiency rather than the underlying disorder itself. Unfortunately, it has to be concluded that at present no single plasma protein biomarker is available that reflects unambiguously and reliably the clinical manifestation of Fabry disease.

Gaucher disease (McKusick 230800), another lysosomal storage disorder caused by deficiency of glucocerebrosidase, can be effectively treated by enzyme replacement therapy. This therapeutic approach has been copied for Fabry disease. The recent availability of therapy based on chronic intravenous administration of recombinant α-Galactosidase A preparations [14,15] has stimulated the search for surrogate markers of disease in serum of Fabry patients. It is envisioned that such markers can be exploited to monitor disease manifestation and the response to therapeutic intervention. Given the present lack of a single serum protein biomarker for Fabry disease, attention has been paid to the discovery of discriminative serum protein profiles. Profiling of serum proteins by means of SELDI-TOF MS (surface-enhanced laser desorption/ionization time-of-flight mass spectrometry) has become a popular approach to obtain a disease-specific protein profile. Indeed, we recently have demonstrated that principal component discriminant analysis of SELDI-TOF MS data obtained from serum specimens allowed classification of Gaucher disease patients [16]. Cross-validation showed that the sensitivity of the
discriminatory model was 89% and the specificity 90%. We have next studied in a similar fashion the value of SELDI-TOF MS serum profiling for discrimination of symptomatic Fabry disease. The outcome of this investigation is here reported and discussed.

Materials & methods

Subjects

Patients. All patients with Fabry disease studied (n=20, 14 males and 6 females; mean age 41 years, range 18–57) were known by referral to the Academic Medical Center. Table 1 shows the clinical characteristics of the investigated Fabry patients.

Control subjects. The control group consisted of male (6) and female (11) healthy volunteers (mean age 36 years, range 23–54).

Relatives of Fabry patients. None of the healthy relatives (3) carried the α-Galactosidase A mutation.

Overall severity of disease was assessed using the Mainz Severity Score Index (MSSI) [17]. In brief, the MSSI is composed of four sections that cover the general, neurological, cardiovascular, and renal signs and symptoms of the disease. The total scores are reported to represent mild (<20), moderate (20–40), or severe (>40) Fabry disease.

<table>
<thead>
<tr>
<th>Table 1. Patient characteristics of 20 Fabry patients (Median en range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
</tr>
<tr>
<td>No. of patients</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>MSSI score</td>
</tr>
</tbody>
</table>

Note: NS, not significant.

Serum samples

Blood samples were collected from patients (before therapeutic intervention) and other subjects in 7 mL, BD Vacutainer, ‘red-top’ tubes (BD # 367625), and sera were prepared. Collection protocols were the same for both groups. Blood samples were allowed to clot at room temperature for 30 min. Subsequently, blood samples were centrifuged at 1300 RCF for 10 min at room temperature. All serum samples were stored at -20°C until required. Serum samples of Fabry patients were obtained before initiation of therapy. Approval was obtained from the Ethic Committee. Informed consent was provided according to the Declaration of Helsinki.

SELDI-TOF MS

Serum samples were surveyed for basic proteins with SELDI-TOF MS making use of the anionic surface of CM10 ProteinChip® Arrays (Ciphergen Biosystems Inc., Fremont, CA, USA). Serum samples (10 μL) were first mixed with 90 μL of denaturation solution (9M urea [Sigma Chemical Company, St. Louis, MO, USA], 2% CHAPS [Fluka Biochemika,
Limited value of serum protein profiling for Fabry disease

Buchs, Switzerland, and 1% DTT [Sigma Chemical Company, St. Louis, MO, USA]) and incubated at room temperature for 1 hour. An aliquot (10 μL) of this solution was mixed with 90 μL binding buffer (50 mM Tris [Sigma Chemical Company, St. Louis, MO, USA] + 0.1% Triton X-100 [BHD Laboratory Supplies, Poole, Dorset, UK], adjusted to pH 7 with hydrochloric acid [Merck, Darmstadt, Germany]). Before application of the sample to a CM10 ProteinChip® Array, all spots were equilibrated. To equilibrate the CM10 ProteinChip® Array, spots were washed with 200 μL of binding buffer (two times, 5 min on a platform shaker) by using a Ciphergen Biosystems 96-well bioprocessor. After equilibration, buffer was removed and samples were added. Fabry and control samples were applied in random order. The samples were allowed to bind to the anionic surface for 40 min at room temperature on a platform shaker. Subsequently the ProteinChip® Arrays were washed with 200 μL binding buffer (two times, 5 min on a platform shaker). Next the ProteinChip® Arrays were washed with 200 μL binding buffer without Triton X-100 (two times, 5 min on a platform shaker). After a brief wash with deionized water (to remove salts) ProteinChip® Arrays were dried on air. Prior to SELDI-TOF MS analysis, matrix was added to each spot (two times 0.5 μL of sinapinic acid [Fluka Biochemika, Buchs, Switzerland] (10 mg/mL) in 50% aqueous acetonitrile [Merck, Darmstadt, Germany] containing 1% TFA [Fluka Biochemika, Buchs, Switzerland]). After co-crystallization of the (bound) proteins with the matrix molecules, a pulsed nitrogen laser was used for ionization of the samples. ProteinChip® Arrays were analyzed using a PBSIIc ProteinChip® Reader (Ciphergen Biosystems Inc., Fremont, CA, USA), a linear laser desorption/ionization time-of-flight mass spectrometer equipped with time-lag focussing. The result is a mass spectrum composed of the mass to charge ratios (m/z values) and peak intensities originating from the desorbed (poly)peptide ions. All spectra were acquired in the positive-ion mode.

Preprocessing of SELDI-TOF MS data for further analysis

External calibration
Spectra were externally calibrated against a mixture of known peptides (All-in-1 Peptide Standard, Ciphergen Biosystems Inc., Fremont, CA, USA). The pre-mixed peptide standard includes arg8-vasopressin (1084 Da), somatostatin (1637 Da), porcine dynorphin (2147 Da), human adrenocorticotropic hormone (1–24) (2933 Da), bovine insulin β-chain (3495 Da), human insulin (5807 Da), and hirudin BHVK (7033 Da).

Spot-to-spot calibration
Spot-to-spot calibration is a feature of the ProteinChip® Software that accounts for the spot-to-spot variation that can occur on an individual array. To determine the correction factors for the different positions on an array we used a set of peaks that is always present in our spectra. The correction factors for the different positions on an array are applied to the corresponding mass spectra and used in the recalculation of the masses.
Baseline subtraction

The ProteinChip® baseline algorithm removes offsets in the spectra that are the result of how the signal is collected electronically and of chemical noise contributed from the energy absorbing molecules in the matrix. The algorithm is a modified piecewise convex-hull that attempts to find the bottom of the spectra and correct the peak height and area. Baseline subtraction is applied to all spectra.

Peak detection

Peaks were detected with Biomarker Wizard™. Biomarker Wizard™ is a feature of the ProteinChip® Software that is used.

Data analysis

Classification

To find differences between the SELDI-TOF MS serum protein profiles of controls and Fabry patients we used two classification methods: principal component discriminant analysis (PCDA) and support vector machines (SVM). These two methods construct classification rules in different ways, thus we have the opportunity to draw classifier independent conclusions.

PCDA is a combination of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). First, PCA is applied to the data to reduce the dimensionality. The PCA scores are then used in LDA to find a direction that discriminates between the two groups, by maximizing the ratio of the variance between the groups to the variance within the groups [18]. PCDA was used exactly as described previously [16].

The rank products variable selection method [19] can be conveniently combined with PCDA [16]. As a by-product, the PCDA training procedure generates several discrimination models, all of which describe the difference between cases and controls, albeit with slightly different discriminant vectors. The loading of a variable in a discriminant vector can be regarded as a measure of its importance. In each of the models obtained with cross validation, the variables can be ranked by their absolute loading in the discriminant vector. Then, if p-fold cross validation is used, where the data is divided in p parts and in every fold a different part forms the test set and the remaining p-1 parts form the training set, each variable is ranked p-times. The p ranks of a variable are multiplied to obtain the variable’s rank product, which is a measure of its overall importance.

A support vector machine (SVM) [20,21] with linear kernel was used to find a hyperplane that separates the Fabry profiles from the controls. When the classes are linearly separable, the optimal hyperplane maximizes the distance from the closest objects to the hyperplane. The class assignment of new samples depends on which side of the hyperplane they are.

All data analyses were performed in Matlab (Mathworks). The SVM algorithm is a routine in the Bioinformatics Toolbox (Mathworks). The PCDA and Rank Products codes are available at www.bdagroup.nl
Normalization and scaling
The data were normalised by dividing each spectrum by its median intensity, making the intensities of the peaks comparable. Thereafter, the data are auto-scaled: all variables have zero mean and unit variance. In auto-scaled data, the contribution of a variable to the classification model is not dependent on the intensity of the signal, but on the relative difference in signal intensity between the classes.

Statistical validation

Prediction error
The prediction error is used as a measure of the performance of the PCDA and SVM classification rules. We calculate the prediction error as the misclassification rate in a 10-fold cross-validation scheme. In this scheme, a model is constructed on a training set after which an independent set of samples is used to test the model.

Permutation test
The significance of the prediction error is determined using permutation tests. In a permutation test, the class labels are repeatedly removed and randomly reassigned to samples to create an uninformative data set of the same size as the data under study. Building and testing a classifier on many permutations of the data gives a distribution of the performance found by chance, to which the performance of the classifier on the original data can be compared. The same classifier building protocol that is applied to the data is applied to the permutations, including any filtering or other selection of variables and parameter tuning [22].

Results

Data
Serum samples of controls and Fabry patients were measured with SELDI-TOF MS. Preprocessing of the spectra was performed according to the descriptions given above. The resulting data set contained 20 control and 20 Fabry spectra, each consisting of 590 m/z values between 1000 en 10,000 Da. The protein profiles were normalized by dividing each profile by its median to arrive at comparable spectra. To prevent the largest peaks in the protein profiles from dominating the PCA part of the model, the data were auto-scaled. For cross-validation, auto-scaling was always performed on the training data before modelling and then the test data was scaled prior to prediction with the scaling parameters of the training set. By doing this, it is ensured that the prediction of the test data is truly independent.

Fabry patients vs Controls
Amongst the 20 controls, 3 relatives of Fabry patients were present. These mass spectra were removed from the control group. The remaining 20 Fabry and 17 control samples
were used to construct classification models with PCDA and SVM. The models were tested with 100 cross-validations, repeatedly leaving a small set of samples completely out of the model training phase. The class labels for the test samples are then predicted. Table 2 shows how often each sample is misclassified with both methods. On average, PCDA misclassifies 9.3 samples, or 25%. Although the misclassification rate is high, the P-value obtained from 10,000 permutations is 0.004, suggesting that the differences found are significant.

Table 2. Percentage misclassified in 100 predictions with PCDA and with SVM

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>PCDA</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>59</td>
<td>2</td>
</tr>
<tr>
<td>F2</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>F3</td>
<td>75</td>
<td>2</td>
</tr>
<tr>
<td>F4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F6</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>F7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F9</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>F10</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>F11</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F13</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>F14</td>
<td>100</td>
<td>98</td>
</tr>
<tr>
<td>F15</td>
<td>23</td>
<td>5</td>
</tr>
<tr>
<td>F16</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>F17</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F18</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>F19</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F20</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig. 1 shows how the misclassification rate of the PCDA classifier depends on the number of variables selected with rank products. The variables included in the models are best discriminating m/z values. The performance of PCDA can be improved; using a selection of 100 m/z values, the misclassification rate is decreased to 21% (7.6 misclassifications). The five m/z values (>1500 Da) that rate highest in the rank products selection are given in Table 3. Analysis of the individual m/z values did not reveal a clear relationship with Fabry status of individuals.
The SVM classifier performs somewhat better than PCDA; on average 5.9 samples are misclassified (16%). The P-value for the SVM result is 0.0001.

**Fabry relatives**

SELDI-TOF MS protein profiles of three relatives of Fabry patients were predicted with a PCDA and with a SVM model, which are both constructed using all 20 Fabry patients and 17 controls. Interestingly, in all cases the subjects were misclassified as being most likely Fabry patients.

### Table 3. The best discriminating variables were selected with rank products

<table>
<thead>
<tr>
<th>m/z (Da)</th>
<th>Rank product</th>
</tr>
</thead>
<tbody>
<tr>
<td>2057.02</td>
<td>222208</td>
</tr>
<tr>
<td>1785.08</td>
<td>4.06E+09</td>
</tr>
<tr>
<td>1755.77</td>
<td>2.00E+10</td>
</tr>
<tr>
<td>1828.05</td>
<td>1.02E+12</td>
</tr>
<tr>
<td>3439.20</td>
<td>1.25E+12</td>
</tr>
</tbody>
</table>

Note: The m/z values with the lowest rank products are shown. The lower the rank product, the better discriminating the m/z value is.

**Discussion**

In sharp contrast to the earlier positive findings with serum specimens of Gaucher disease patients, comparable SELDI-TOF MS profiling and PCDA analysis rendered no reliable discrimination between symptomatic Fabry patients and normal subjects. Six out of 17 control subjects were misclassified as patients. Four out of 20 Fabry patients were misclassified as normal. It should be noted that the three of the four misclassified patients were mildly to moderately affected (MSSI: F1(16), F4(24), and F6(23)), However, one misclassified patient, F14 (MSSI: 46), showed characteristic severe Fabry disease manifestations.

SVM analysis of the profiles rendered slightly better results; four control subjects and two patients, F6 and F14, being misclassified. PCDA and SVM analysis were both used to exclude the possibility that the obtained results are the consequence of the used classification method. It seems thus unlikely that the poor discrimination that was obtained both with the PCDA and SVM analysis can be contributed to a particular classification method.

It might be argued that the procedure used for protein profiling is not sensitive enough to detect early manifestations of Fabry disease. However, concomitant with misclassification of Fabry patients as being normal, some control subjects are classified as diseased Fabry patients. Strikingly, all three unaffected relatives of Fabry patients (R1, R2 and R3) that were tested were classified as being patient, either using SVM or PCDA. This suggests that the discrimination may not be primarily based on the underlying disorder but rather on other characteristics shared by families. This illustrates the importance to use matched patient and control subjects, possibly even from the same family, in these types of studies.

In conclusion, the outcome of our investigation is negative. SELDI-TOF MS protein
profiling rendered no reliable discrimination between diseased Fabry patients and healthy control subjects. In hindsight, the result of our investigation is not so surprising since no single serum biomarker for Fabry disease has been detected so far [23]. It appears that in contrast to the earlier common believe [3], lipid-laden endothelial cells of Fabry patients are not grossly abnormal in behavior and function, and are not releasing specific proteins into the circulation that are detectable by serum protein profiling with the currently available SELDI-TOF MS methodology.

Supporting information available
SELDI-TOF MS data in mat-files and software for PCDA, cross-validation and rank products in m-files are available free of charge at www.bdagroup.nl

Acknowledgements
The authors thank TNO Quality of Life (Zeist, NL) for providing PCDA m-files for use with Matlab®.

References


