Clinica Chimica Acta 448 (2015) 174–181

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/clinchim

# The application of fuzzy statistics and linear discriminant analysis as criteria for optimizing the preparation of plasma for matrix-assisted laser desorption/ionization mass spectrometry peptide profiling

Joanna Hajduk [a], Jan Matysiak [a], Piotr Kokot [b], Piotr Nowicki [a], Paweł Dereziński [a], Zenon J. Kokot [a,*]

[a] Department of Inorganic & Analytical Chemistry, Poznan University of Medical Sciences, 6 Grunwaldzka Street, 60-780 Poznań, Poland
[b] Obstetrics and Gynecology Ward, Mielec District Hospital, 22a Żeromskiego Street, 39-300 Mielec, Poland

ARTICLE INFO

*Article history:*
Received 24 January 2015
Received in revised form 25 June 2015
Accepted 25 June 2015
Available online 9 July 2015

*Keywords:*
Mass spectrometry
Matrix-assisted laser desorption/ionization
Peptide profiling
Discriminant analysis
Fuzzy inference system
Diabetes mellitus

ABSTRACT

An alternative bioinformatics approach based on fuzzy theory statistics and linear discriminant analysis is proposed for the interpretation of MALDI MS spectra in peptide profiling. When applied, the methodology enables the establishment of a reproducible plasma preparation protocol appropriate for the evaluation of small data sets. The samples were collected from pregnant women affected by gestational diabetes mellitus (GDM), n = 18 and control group, n = 13. The following pre-treatment sets were tested: pipette tips with C18 stationary phase (ZipTip, Millipore and Omix, Agilent) and magnetic bead-based weak cation exchange chromatography kit (MB WCX, Bruker Daltonics). The spectra were recorded using a MALDI TOF mass spectrometer (UltrafleXtreme, Bruker Daltonics) for a mass range of m/z from 1000 to 10,000. The significant features were selected using the wrapper selection method, and two classification systems were tested: discriminant analysis (DA) and fuzzy inference system (FIS). ClinProTools software was employed to compare the usefulness of the proposed methodology. The study showed that the optimum results for MS spectra were obtained after the use of the ZipTip as pre-treatment method in plasma preparation. Chemometric analysis allowed the differentiation of the GDM group from the control with a high degree of accuracy: 0.7333 (DA) and 0.8065 (FIS).

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

The detection of protein–peptide patterns associated with disease development has become a promising approach in proteomic studies for early and accurate diagnosis. As the proteins and peptides contained within complex biological fluids can be measured easily and precisely, the search for differences between healthy individuals and those affected by illness can contribute to a better understanding of a pathological condition. For this purpose, mass spectrometry (MS) profiling, combined with sophisticated data mining algorithms, has become more widely used, and has demonstrated clear differentiating capabilities [1,2].

Mass spectrometry has become an indispensable tool for analyzing complex biological samples as a result of the development of two soft ionization methods: MALDI (matrix-assisted laser desorption/ionization) and ESI (electrospray ionization). Currently, most profiling involves the ionization of biological molecules using MALDI and surface-enhanced laser desorption ionization (SELDI) mass spectrometry (MS) [3–5]. These high throughput methods based on soft ionization techniques are highly suitable for analyzing compounds that are thermally liable and have low volatilities. The proteins and peptides to be analyzed are introduced into the gas phase positively charged and without fragmentation, and there is a high level of sensitivity to their detection. Therefore, in protein investigations such as the analysis of MS profiling patterns, the resulting values recognize the proteins present, whilst the relative abundance of many peptides is determined on the basis of their mass/charge ratio (m/z) across the selected mass range. The employment of the SELDI platform in peptide–protein profiling has shed a new light on the possibility of further clinical applications. Petricon et al. [6] introduced this approach for the first time to identify the serum pattern in ovarian cancer. Despite the prospective application of this technology as a screening tool within a population-based assessment of proteomic patterns, however, both the variability in reproducibility of the plasma and serum samples [7,8] and the failure of the validation process indicate that this method cannot be used in its current state for routine profiling in a clinic [9]. The main drawbacks are related to the lack of sample randomization during preparation and processing, and to errors resulting from the biological sample selection [10]. Furthermore, chip-to-chip variability, the low resolution of the MS data, and the saturation effects that influence data as a result of sample concentration all impose limitations on this method [11]. It seems that MALDI has significant potential for clinical application, mainly due to the speed of analysis and ease of operation. However, issues such as reproducibility, the establishment of robust methods for sample preparation, and mathematical tools for spectra quality assessment and normalization have not been correctly addressed

* Corresponding author.
*E-mail address:* zkokot@ump.edu.pl (Z.J. Kokot).

http://dx.doi.org/10.1016/j.cca.2015.06.025
0009-8981/© 2015 Elsevier B.V. All rights reserved.

yet. Therefore, a thorough examination of this platform appears to be urgently required [9].

In clinical research, serum and plasma represent favorable diagnostic specimens since they are obtained in a minimally invasive manner and could potentially be a source of reliable physiological indicators such as metabolites, peptides or proteins. However, direct analysis of a complex biological sample using MALDI TOF MS will not result in any usable signal in the spectrum due to the presence of interfering substances such as salts, lipids, and high-abundance proteins [12]. The relative amount of these with respect to the proteins of interest hampers the ionization process and the detection of the desired components. Due to the fact that plasma concentrations of individual proteins may vary by up to 12 orders of magnitude, the dynamic range of measurement shows a major challenge in analyzing of plasma proteome [13]. Therefore, the seeking for proper pre-fractionation and cleaning strategy is necessary in sample preparation prior to MS analysis. The immunodepletion methods, using antibody affinity immunoabsorption of specific antigens, are appropriate for selective depletion of high-abundance proteins. The strategy allows the extension of proteome analysis to lesser concentration, revealing lower abundance components. However, the nature of immunodepletion may have an effect in protein–protein interactions and removal of numerous off-targeted proteins. Moreover, the variation in the efficiencies from run to run process can increase the potential for false discovery [14]. Hydrophobic or ionic solid phase extraction (SPE) sets are widely used to improve the ability to record protein and peptide signals. Commercial pipette tips such as ZipTip (Millipore) or OMIX (Agilent) pre-packed with hydrophobic phase have been reported as commonly used technologies in MS profiling [15,16]. Nevertheless, there are divergent opinions in the related literature about the effectiveness of the extraction properties of SPE pipette tips in mass spectrometric analyses [17,18]. Alternatively, in proteomics research, magnetic bead particles with hydrophobic, ionic or high metal affinity surfaces are frequently used. A human blood profiling study that used them for sample preparation was highlighted by Baumann et al. [19]. They stated that applying the standardized pre-analytical procedure in combination with magnetic bead-based fractionation decreased the variability of the proteome patterns. Although other studies have also recognized the utility of a magnetic bead-based pretreatment approach [20], there is still little knowledge about which of the methods is better in obtaining reproducible and high quality results.

In the available literature, bioinformatic analyses of MALDI MS spectra are mostly performed using data-processing software which is designed for visualization, data reduction, data mining, and the construction of predictive models from MS data [15,20,21]. The mathematical algorithms typically used when creating classification models from profiling patterns are as follows: a genetic algorithm which imitates evolution in nature; a support vector machine based on statistical learning theory; a supervised neural network (SNN); and a quick classification algorithm which takes account of p-values for classification [22]. Despite the troublesome step processing spectra that could be omitted, there are limitations in the use of these algorithms which lead us to seek a new method for processing spectra and performing statistical analysis. The main obstacles to obtaining a reliable outcome include a high rejection of spectra, and the suitability of this method only for the analysis of large data sets. Moreover, the lack of intuitive insight into the results makes some of the methods used completely incomprehensible to physicians. The fuzzy theory system, although rarely used in the field of proteomic mass spectrometry, shows great prospects in helping to retrieve various fields of information from a database [23]. The system, based on linguistic rules, can significantly enhance non-linearly separable data by classifying them with the same mean, thus facilitating interpretation. On the other hand, classical linear discriminant analysis is easy to understand and provides a good reference for comparison among data sets. The accurate interpretation of MS data is a crucial step in profiling studies and in the investigation of biomarkers; yet, it should be kept in mind that each bioinformatic analysis must be carefully assessed in depth [24].

This research uses plasma peptide patterns from matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS) as input data. For the purposes of this study, we have chosen pregnant women affected by gestational diabetes mellitus (GDM) as a research population. This metabolic disease is defined as any glucose intolerance with either its onset or its first identification occurring during pregnancy, and where the condition may or may not persist after delivery [25]. The diagnosis of GDM is made between the 24th and 28th weeks of gestation (towards the end of the second trimester) and allows for a clear differentiation of healthy women from those with carbohydrate metabolism disorders. Therefore, we decided to submit the patients to a deeper proteomic analysis and form a research group suitable for a MALDI MS profiling study. Moreover, as protein levels are highly disturbed during pregnancy, only normoglycemic pregnant women were enrolled in the control group.

Here, we report the comparative study of three different pretreatment methods based on SPE (two sets of commercially prepacked C18 SPE tips — OMIX and ZipTip respectively; and also magnetic beads — MB WCX) which are utilized for plasma preparation in GDM peptide profiling. Since we were aware of the small sample size, an alternative bioinformatics approach based on fuzzy theory statistics was proposed for the interpretation of MALDI MS spectra. Lastly, to compare the new classification and statistics performance, ClinProTools software designed for the evaluation of biomarkers was employed.

## 2. Methods

### 2.1. Materials and reagents

LC–MS grade acetonitrile (ACN), water, ethanol, acetone and isopropanol were provided by J.T. Baker (Center Valley, PA, USA). Trifluoroacetic acid (TFA) and ammonium acetate were obtained from Sigma-Aldrich (St. Louis, MO, USA). The matrix α-cyano-4-hydroxycinnamic acid, Peptide Calibration Standard, Protein Calibration Standard I, and AnchorChip 800 μm were purchased from Bruker Daltonics (Bremen, Germany).

### 2.2. Study group

The research project has been approved by the Regional Ethics Committee of the Poznan University of Medical Sciences (decision No. 200/13). The volunteers participating in the study gave written consent and completed a detailed survey before the plasma samples were taken. The GDM diagnosis was based on the Polish Gynecological Society standards of medical care in the management of women with diabetes [26]. The criteria for the diagnosis of GDM included the 75 g Oral Glucose Tolerance Test (OGTT), which was carried out between 24–28 weeks of gestation on pregnant women whose fasting glucose was below 5.5 mmol/L in early pregnancy. The test was performed under the following conditions: fasting for 8–14 h after the last meal; at least three days without any limit on the consumption of carbohydrates (normal diet); 75 g of glucose dissolved in 250–300 mL water and drunk over a period of 5 min in the sitting position, without any intake of food or inhalation of smoke. GDM was recognized when at least one value (out of three) was above normal: fasting $> 5.5$ mmol/L; after 1 h $> 10$ mmol/L; after 2 h $> 7.8$ mmol/L. The pregnant patients classified as positive cases with confirmed GDM were named as the GDM group ($n = 18$). Only women with GDM whose treatment was via their diet were examined in this study. The control group ($n = 13$) included healthy pregnant women with normoglycemia and with equivalent ethnicities, ages, body mass indexes (BMIs) and sampling times. Patients with any other chronic condition were excluded. Plasma samples were collected in the Mielec District Hospital from both healthy pregnant women (control group) and those with confirmed GDM, in the period between 24 and 28 weeks of gestation. Blood was drawn into ETDA tubes (S-Monovettes) and then centrifuged at 700 g

Download English Version:

https://daneshyari.com/en/article/8310602

Download Persian Version:

https://daneshyari.com/article/8310602

Daneshyari.com