Plasma Phospholipid Metabolic Profiling and Biomarkers of Type 2 Diabetes Mellitus Based on High-Performance Liquid Chromatography/Electrospray Mass Spectrometry and Multivariate S...

ARTICLE *in* ANALYTICAL CHEMISTRY · AUGUST 2005

Impact Factor: 5.64 · DOI: 10.1021/ac0481001 · Source: PubMed

7 AUTHORS, INCLUDING:

Chang Wang
Suzhou University
20 PUBLICATIONS **657** CITATIONS
SEE PROFILE

Jun Yang
University of California, Davis
94 PUBLICATIONS **1,873** CITATIONS
SEE PROFILE

Guowang Xu
Dalian Institute of Chemical Physics, Chinese...
378 PUBLICATIONS **6,333** CITATIONS
SEE PROFILE

Anal. Chem. **2005,** *77,* 4108−4116

Plasma Phospholipid Metabolic Profiling and Biomarkers of Type 2 Diabetes Mellitus Based on High-Performance Liquid Chromatography/ Electrospray Mass Spectrometry and Multivariate Statistical Analysis

Chang Wang,[†] Hongwei Kong,[†] Yufeng Guan,[‡] Jun Yang,[†] Jianren Gu,[§] Shengli Yang,[†,§,||] and Guowang Xu*,[†]

National Chromatographic R&A Center, Dalian Institute of Chemical Physics, the Chinese Academy of Sciences, Dalian 116023, China, The Second Affiliated Hospital of Dalian Medical University, Dalian 116023, China, National Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Shanghai Jiao Tong University, Shanghai 200032, China, and Shanghai Research Center of Biotechnology, Chinese Academy of Sciences, Shanghai, China

Liquid chromatography/mass spectrometry (LC/MS) followed by multivariate statistical analysis has been successfully applied to the plasma phospholipids metabolic profiling in type 2 diabetes mellitus (DM-2). Principal components analysis and partial least-squares discriminant analysis (PLS-DA) models were tested and compared in class separation between the DM2 and control. The application of an orthogonal signal correction filtered model highly improved the class distinction and predictive power of PLS-DA models. Additionally, unit variance scaling was also tested. With this methodology, it was possible not only to differentiate the DM2 from the control but also to discover and identify the potential biomarkers with LC/MS/MS. The proposed method shows that LC/MS combining with multivariate statistical analysis is a complement or an alternative to NMR for metabonomics applications.

Metabonomics is the identification and measurement of metabolic profile dynamics of host changes as the result of exposure to a toxin or drug, to environmental effects, or to the onset of disease.[1] Based upon the multivariate analysis of complex biological profiles, metabonomics has been used for toxicological screening or disease diagnosis. These techniques involved NMR,[1−4] mass spectrometry,[1,5−8] and optical spectroscopic techniques.[9] Although NMR has the advantages of being nondestructive, applicable to intact biomaterials, and accessing intrinsically more information rich in complex-mixture analyses, owing to its low sensitivity, it is difficult to satisfy the multiparallel analysis of hundreds of metabolites for generation of large metabolic networks, especially for a low-abundance analyte. Gas chromatography/mass spectrometry (GC/MS) is a relatively low cost alternative that provides high separation efficiency to resolve the complex biological mixtures; however, it requires that samples should be volatile. This requirement can be readily accomplished by chemical derivatization, but it could be at the cost of additional time, processing, and variance.[10] HPLC coupled to MS (LC/MS) is a powerful alternative that offers high selectivity and good sensitivity. The extra resolution provided by the chromatography step further facilitates a reduction in chemical interference, particularly for low-abundance analytes. The technique coupled to pattern recognition has been proved to be a good choice for metabolite detection and metabolite profiling.[11−16]

* To whom correspondence should be addressed. Tel./Fax: 0086-411-84379559. E-mail: dicp402@mail.dlptt.ln.cn.

[†] Dalian Institute of Chemical Physics.

[‡] The Second Affiliated Hospital of Dalian Medical University.

[§] Shanghai Jiao Tong University.

[||] Shanghai Research Center of Biotechnology.

(1) Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E. *Nat. Rev.* **2002**, *1*, 153−161.

(2) Brindle, J. T.; Antti, H.; Holmes, E.; Tranter, G.; Nicholson, J. K.; Bethell, H. W. L.; Clarke, S.; Schofield, P. M.; Mckilligin, E.; Mosedale, D. E.; Grainger, D. *J. Nat. Med.* **2002**, *8*, 1439−1444.

(3) Keun, H. C.; Beckonert, O.; Griffin, J. L.; Richter, C.; Moskall, D.; Lindon, J. C.; Nicholson, J. K. *Anal. Chem.* **2002**, *74*, 4588−4593.

(4) Ebbels, T. M. D.; Antti, H.; Bollard, M. E.; Beckonert, O.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Anal. Chim. Acta* **2003**, *490*, 265−276.

(5) Goodacre, R.; Rooney, P. J.; Kell, D. B. *J. Antimicrob. Chemother.* **1998**, *41*, 27−34.

(6) Freeman, R.; Goodacre, R.; Sisson, P. R.; Magee, J. G.; Ward, A. C.; Lightfoot, N. F. *J. Med. Microbiol.* **1994**, *40*, 170−173.

(7) Halket, J. M.; Przyborowska, A.; Stein, S. E.; Mallard, W. G.; Down, S.; Chalmers, R. A. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 279−284.

(8) Taylor, J.; King, R. D.; Altmann, T.; Fiehn, O. *Bioinformatics* **2002**, *18*, S241−S248.

(9) Winson, M. K.; Goodacre, R.; Timmins, E. M.; Jones, A.; Alsberg, B. K.; Woodward, A. M.; Rowland, J. J.; Kell, D. B. *Anal. Chim. Acta*. **1997**, *348*, 273−282.

(10) Sumner, L. W.; Mendes, P.; Dixon, R. A. *Phytochemistry* **2003**, *62*, 817−836.

(11) Plumb, R. S.; Stumpf, C. L.; Gorenstein, M. V.; Castro-Perez, J.; Dear, G. J.; Anthony, M.; Sweatman, B. C.; Connor, S. C.; Haselden, J. N. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1991−1996.

(12) Plumb, R. S.; Stumpf, C. L.; Granger, J. H.; Castro-Perez, J.; Haselden, J. N.; Dear, G. J. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2632−2638.

(13) Kerwin, G. M.; Watts, J. D.; Aebersold R. *Anal. Biochem.* **1997**, *24*, 347−356.

(14) Raithk, Z. S.; Lasch, J.; Neubert, R. H. H. *Anal. Chim. Acta* **2000**, *418*, 167−173.

10.1021/ac0481001 CCC: $30.25
Published on Web 05/17/2005

Among metabolic diseases, diabetes mellitus is considered one of the most prevalent throughout the world. Currently, statistics shows that over 10% of the world's aged population (60 years and older) suffers from this disease. As a consequence, it consumes a considerable proportion of world health expenditure.[17] Diabetes mellitus is a complex polygenic disorder of intermediary metabolism. Although type 1 diabetes is due to insulin deficiency and type 2 diabetes is due to insulin resistance, many studies have demonstrated that diabetes mellitus is intimately associated with metabolism disorder of lipid or fatty acid in phospholipids.[18−22] Phospholipids are a class of important constituents in the biomembranes. Both the physical and chemical properties of the membrane bilayer can be affected by the variation of phospholipid compositions. Membrane phospholipids are a complex mixture of molecular species containing a variety of fatty acyl and headgroup compositions. In addition to their structural role, some phospholipids also participate in biological processes in various pathways. Other phospholipids, such as polyphosphoinositides, are important in cellular signaling systems.[23,24] Phospholipids serve as a reservoir for arachidonic acid (20:4 *n*-6) and other polyunsaturated fatty acids that can be metabolized to biologically active eicosanoids such as prostaglandins, thromboxanes, leukotrienes, and lipoxins.[25,26] Phospholipids have been given increased attention in many fields, for example, as biomarkers in chemotaxonomical studies and in the making of liposomes for drug delivery or cosmetics/detergents. The commercial use of phospholipids is increasing in fields such as biomembranes, skin-care formulations, and drug delivery.

Analysis of these phospholipids has been carried out with chromatographic techniques such as thin-layer chromatography[27] and high-performance liquid chromatography (HPLC).[28−30] However, due to the complexity of phospholipids, the detection methods including UV absorbance and evaporative light scattering have shown their limitation.[31] MS offers an attractive alternative for the analysis of phospholipid compositions because of its high sensitivity, specificity, and (apparent) simplicity. Particularly, electrospray ionization-mass spectrometry (ESI-MS) has been shown to be a very promising technique.[32] However, it is important to have a chromatographic system that separates the different phospholipid classes due to possible mass overlap. Thus, in the analysis of samples taken from a complex biological matrix, there is a need for class separation by LC followed by subsequent species identification utilizing mass spectrometry.[33] Recently, HPLC/MS has been applied in the lipid analysis.[34−36] Previously, we coupled HPLC to MS and identified over 100 phospholipid species in human blood samples.[31]

A single LC/MS metabolite profile can yield hundreds of components. This provides a wealth of information to be interpreted and leads to significant challenges in processing the data. Thus, it is necessary to utilize a wide range of statistical data reduction and multivariate analyses,[37,38] such as principal components analysis (PCA) and partial least squares-discriminant analysis (PLS-DA). In addition, to improve the performance of multivariate pattern-recognition analysis and enhance the predictive power of the model, data scaling and data preprocessing methods, such as unit variance (Uv) and orthogonal signal correction (OSC),[39] are used to optimize the separation.

In this paper, we coupled LC/MS technology to multivariate statistical analysis in order to study phospholipid metabolic profiling in diabetes mellitus and to discover the potential biomarkers. PCA and PLS-DA models were compared in class separation of type 2 diabetes mellitus (DM2) patients and healthy persons (control). Uv scaling and OSC data preprocessing methods were also developed to improve class separation. Product ion spectra following collisionally activated dissociation (CAD) of quasi-molecular ions in a tandem mass spectrometry was used to identify individual molecular composition of phospholipid biomarkers.

EXPERIMENTAL SECTION

1,2-Dimyristoyl-*sn*-glycero-3-phosphoethanolamine (C14:0/C14:0 PE, phosphatidylethanolamine), 1,2-dimyristoyl -*sn*-glycero-3-phosphocholine (C14:0/C14:0 PC, phosphatidylcholine), 1,2-dimyristoyl-*sn*-glycero-3-phospho-L-serine (sodium salt) (C14:0/C14:0 PS, phosphatidylserine), and 1-lauroyl-2-hydroxy-*sn*-glycero-3-phosphocholine (C12:0 lysoPC, lysophosphatidylcholine) were from Avanti Polar Lipids (Alabaster, AL). Other phospholipid standards were from Avanti Polar Lipids or Sigma (St. Louis, MO). 2,6-Di-*tert*-butyl-4-methylphenol was from Aldrich-Chemie (Steinheim, Germany). Formic acid and all the solvents were HPLC grade (TEDIA); ammonia (25%) was analytical grade from Lian-Bang (Shenyang, China).

Sample Preparation. Phospholipid standards were dissolved (∼1 mg/mL) in chloroform/methanol (2:1, v/v) and further diluted in hexane/1-propanol (3:2, v/v).

(15) Huhman, D. V.; Summer, L. W. *Phytochemistry* **2002**, *59*, 347−360.
(16) Idborg-Bj_rkman, H.; Edlund, P. O.; Kvalheim, O. M.; Schuppe-Koistinen, I.; Jacobsson, S. P. *Anal. Chem.* **2003**, *75*, 4784−4792.
(17) Jones, D. R.; Varela-Nieto, I. *Mol. Med.* **1999**, *5* (8), 505−514.
(18) Han, X. L.; Abendschein, D. R.; Kelley, J. G.; Gross, R. W. *Biochem. J.* **2000**, *352*, 79−89.
(19) Tan, K. C.; Shiu, S. W.; Wong, Y. *Eur. J. Clin. Invest.* **2003**, *33* (4), 301−306.
(20) Perassolo M. S.; Almeida, J. C.; Pra, R. L.; Mello, V. D.; Moulin, C. C.; Zelmanovitz, T.; Azevedo, M. J.; Gross, J. L. *Diabetes Care.* **2003**, *26* (3), 613−618.
(21) Zendzian-Piotrowska, M.; Bucki, R.; Gorska, M.; Gorski, J. *Horm. Metab. Res.* **2000**, *32* (10), 386-389.
(22) Hsu, F. F.; Bohrer A.; Wohltmann, M.; Ramanadham, S.; Ma, Z.; Yarashesk, K.; Truk, *J. Lipids* **2000**, *35* (8), 839−854.
(23) Berridge, M. J. *Nature* **1993**, *361*, 315−325.
(24) Exton, J. H. *Biochim. Biophys. Acta* **1994**, *1212*, 26−42.
(25) Marcus, A. J.; Hajjar, D. P. *J. Lipid Res.* **1993**, *34*, 2017−2031.
(26) Piomelli, D. *Crit. Rev. Neurobiol* **1994**, *8*, 65−83.
(27) Suzuki, E.; Sao, A.; Kuriki, T.; Miki, T. *Biol. Pharm. Bull.* **1993**, *16*, 77−80.
(28) Seewald, M.; Eichinger, H. M. *J. Chromatogr., A* **1989**, *469*, 271−280.
(29) Bearerogers, J. L.; Bonekampnasner, A.; Dieffenbacher, A. *Pure Appl. Chem.* **1992**, *64*, 447−454.
(30) Caboni, M. F.; Menotta, S.; Lercker, G. *J. Chromatogr., A* **1994**, *683*, 59−65.
(31) Wang, C.; Xie, S. G.; Yang, J.; Yang, Q.; Xu, G. W. *Anal. Chim. Acta* **2004**, *525*, 1−10.
(32) Koivusalo, M.; Haimi, P.; Heikinheimo, L.; Kostiainen, R.; Somerharju P. *J. Lipid Res.* **2001**, *42*, 663−672.
(33) Karlsson, A. Å.; Michelsen, P.; Larsen, Å.; Odham, G. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 775−780.
(34) Kim, H. Y.; Wang, T. C. L.; Ma, Y. C. *Anal. Chem.* **1994**, *66*, 3977−3982.
(35) Taguchi, R.; Hayakawa, J.; Takeuchi, Y.; Ishida, M. *J. Mass Spectrom.* **2000**, *35*, 953−966.
(36) Delong, C. J.; Baker, P. R. R.; Samuel, M.; Cui, Z.; Thomas, M. J. *J. Lipid Res.* **2001**, *42*, 1959−1968.
(37) Eriksson, L.; Johansson, E.; Kettaneh-Wold, H.; Wold, S. *Introduction to Multi- and Megavariate Analysis using projection Methods (PCA & PLS-DA);* Umetrics Inc.: Box 7960, SE90719 Umeå, Sweden, 1999; pp 267−296.
(38) Jurs, P. C. *Science* **1986**, *232*, 1219−1224.

Table 1. Linear Gradient Composition[a]

time (min)	A, %	B, %
0	68	32
20	20	80
33	20	80
38	68	32
60	68	32

[a] Solvent mixture A: hexane/1-propanol/formic acid/ammonia (79/20/0.6/0.07, v/v). Solvent mixture B: 1-propanol/water/formic acid/ammonia (88/10/0.6/0.07, v/v).

Human plasma samples were collected from 35 healthy adults and 34 patients with type 2 diabetes mellitus. The age range was 30−80 years. All patients were from the Second Affiliated Hospital of Dalian Medical University of China with the fasting plasma glucose concentration above 7.0 mmol/L. Plasma samples were collected and kept at −20 °C until analysis. The lipids in the 500-μL plasma samples were extracted essentially as described earlier.[40] Appropriate amounts of internal standards (e.g., C14:0/C14:0 PE, C14:0/C14:0 PC, C14:0/C14:0 PS, C12:0 lysoPC) were added during phospholipid extraction. Prior to analysis, the extracted samples were redissolved in 200 μL of chloroform/methanol (2:1, v/v) and then were diluted 10 times with hexane/1-propanol (3:2, v/v).

High-Performance Liquid Chromatography. An HP 1100 series HPLC system (Agilent Technologies, Palo Alto, CA) was used. The HPLC separation was performed on a diol column (Nucleosil, 100-5 OH), 250 (mm) × 3.0 (mm, i.d.) × 5.0 (μm, particle size). The total flow rate was 0.4 mL/min. The column temperature was kept at 35 °C. The linear solvent gradient was shown in Table 1. Solvent mixture A was hexane/1-propanol/formic acid/ammonia (79:20:0.6:0.07, v/v); solvent mixture B was 1-propanol/water/formic acid/ammonia (88:10:0.6:0.07, v/v).

Mass Spectrometry. The mass spectrometric detection was conducted on a QTRAP LC/MS/MS system from Applied Biosystems/MDS Sciex equipped with a turbo ion spray source. The combination of highly selective triple-quadrupole MS/MS scans and high-sensitivity ion trap product scans on the same instrument platform provided rapid identification of phospholipids of extracted human plasma sample. The detection of phospholipids eluted from the chromatographic column was performed in the "enhanced MS" (EMS), single-quadrupole mode (so-called "survey scan"), where ions were accumulated and then filtered in the Q3-linear ion trap. The structure of phospholipid is elucidated by the "enhanced" product ion (EPI) scan mode where ions were trapped in the third quadrupole before filtration.

The HPLC effluent entered the MS through a steel ES ionization needle set at 5500 (in positive-ion mode) or 4500 V (in negative-ion mode) and a heated capillary set at 375 °C. The ion source and ion optic parameters were optimized with respect to the positive or negative molecular related ions of the phospholipid standards. The nitrogen drying gas and turbo gas were 45 and 40 psi, respectively. The curtain gas that was to prevent contamination of the ion optics was set at 30 psi. The declustering potential (DP) was set at 80 psi. The other parameters were as follows: EMS as survey scan (mass range m/z 414−917, scan speed 1000 Da s^{-1}, trap time 20 ms) and EPI as dependent scan (scan speed 1000 Da s^{-1}, trap time 150 ms, collision energy set at +35 eV in the positive-ion mode and −40 eV in the negative-ion mode). Q0 trapping was always on.

Data Collection and Normalization. Negative ion LC/MS chromatograms were inspected for profiling the phospholipid species in plasma. Negative ion mode ESI-MS was chosen because it gave more information-rich data than positive-ion ESI. Masses corresponding to the quasi-molecular anions $[M - H]^-$ (for PE, PS, and phosphatidylinositol (PI) species) or $[M - 15]^-$ (for PC, sphingomyelin (SM), and lyso-PC species) for each phospholipid species were plotted against elution time. From the LC/MS profile of plasma samples, 83 phospholipid species in which the ion intensities exceeded 4500 counts/s (after the background was subtracted) and where the peaks constantly occurred in each DM2 sample were collected to form a database. The matching of these peaks (83 peaks) was based on the mass-to-charge ratio and retention time of each class of phospholipid (the retention times of different molecular species with the same polar head are very similar; see LC/MS Analysis section). The retention time of each class phospholipid was aligned by the use of time window determinations.[44] After peak matching, the peak intensities in the extracted ion chromatography were normalized by that of internal standard, using the homemade software, and then were used for PCA or PLS-DA.

Fisher Weight (F). For

$$F_i = (m_{i1} - m_{i2})^2/(V_{i1} + V_{i2})$$

where m_{i1} and m_{i2} are the mean of the ratios of individual phospholipid species of DM2 and control samples to corresponding internal standard, respectively, and V_{i1}, V_{i2} are the variance of these ratios. The higher the value of F_i, the more effect the variance can have on the separation of DM2 and control samples.

Multivariate Analysis. *Principal Components Analysis.* PCA is a bilinear decomposition method used for overviewing "clusters" within multivariate data.[37] The data (X) are represented in K-dimensional space (where K is equal to the number of variables) and reduced to a few principal components (or latent variables) that describe the maximum variation within the data, independent of any knowledge of class membership (that is, "unsupervised"). The principal components are displayed as a set of "scores" (t), which highlight clustering or outliers, and a set of "loadings" (p), which highlight the influence of input variables on t.

Partial Least Squares-Discriminant Analysis and Orthogonal Signal Correction Filtering. Multivariate analysis was performed using SIMCA-P software (version 10.0, Umeå, Sweden). PLS-DA is a multivariate classification method based on PLS, the regression extension of PCA.[37] Whereas PCA works to explain the maximum variation between samples, PLS-DA explains the maximum separation between defined class samples in the data (X) using a **Y** matrix that represents an orthogonal unit vector for each class. PLS-DA is done by a PLS regression against a "dummy matrix" (**Y**), which describes variation according to class. Variation is interpreted in terms of X-scores (T) and X-weights

(39) Wold, S.; Antti, H.; Lindgren, F.; Ohman, *J. Chemom. Intell. Lab. Syst.* **1998**, *44*, 175−185.
(40) Uran, S.; Larsen, Å. Jacobsen, P. B.; Skotland, T. *J. Chromatogr., B* **2001**, *758*, 265−275.

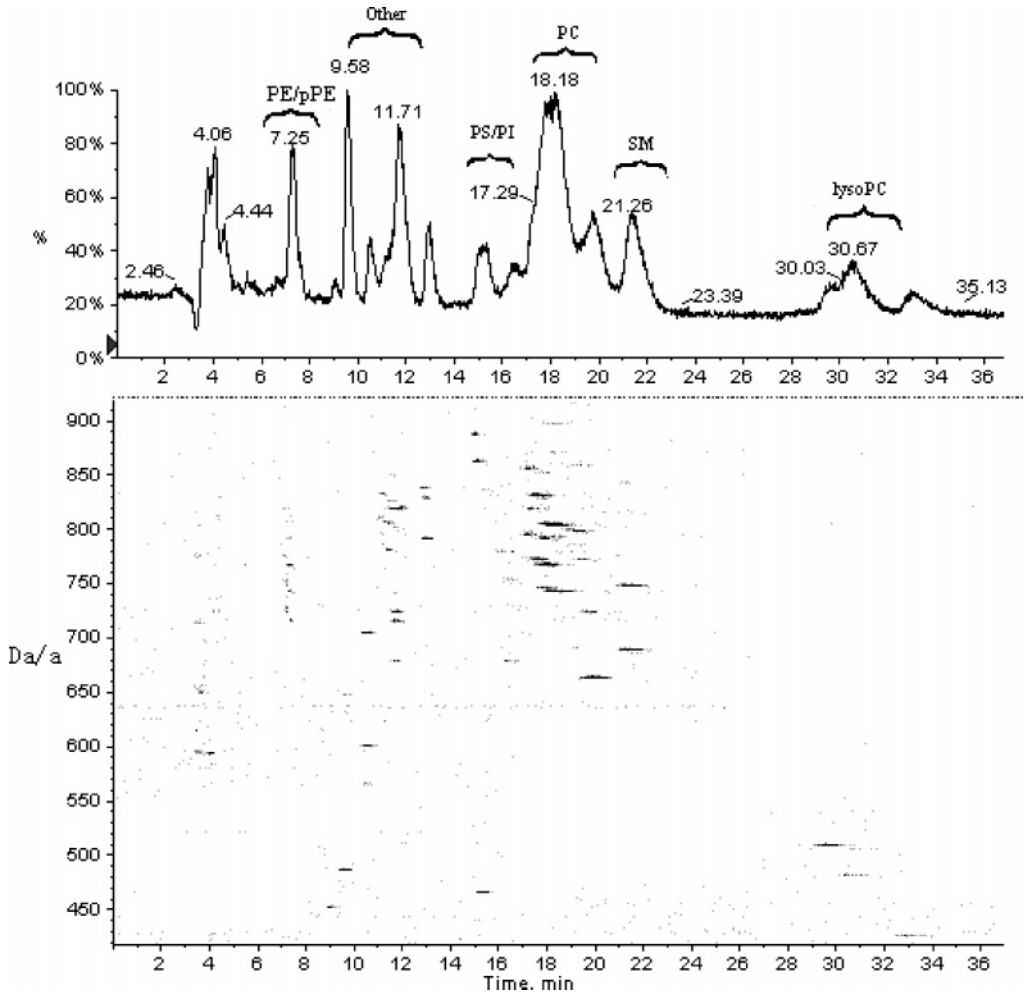

Figure 1. Gradient negative-ion LC/ESI-MS of phospholipid mixture obtained from DM2 patient plasma showing both the total ion current mass chromatogram (top) and a 2D mass chromatogram (bottom). The experimental conditions of LC/ESI-MS are described in text. PE, phosphatidylethanolamine; pPE, plasmalogen phosphatidylethanolamine; PI, phosphatidylinositol; PS, phosphatidylserine; PC, phosphatidylcholine; SM, sphingomyelin; lysoPC, lysophosphatidylcholine.

(*W*, *C*). Once a PLS-DA model is calculated and validated, it can be used for prediction of class membership for unknown samples.

OSC was implemented in the manner described by Wold et al.[39] The class (DM2 or control) identified was used as a responded vector **Y** to describe the variation between the sample classes. The OSC determines components that explain the maximum variance that is orthogonal to **Y**. The use of the residual data from this orthogonal model effectively filters obscuring variation in the data set. The resultant data set has been filtered to allow pattern recognition focused on the variation correlated to features of interest within the sample population, rather than noncorrelated, orthogonal variation.

RESULTS AND DISCUSSION

LC/MS Analysis. The full scan of phospholipid species in plasma were in the negative-ion mode because in this mode most of phospholipid species have relatively high sensitivity. For analysis of the elution patterns of the phospholipids, a two-dimensional (2D) map was constructed with one dimension set to retention time and the other set to mass-to-charge ratio. Figure 1 illustrates the complexity of phospholipid species in DM2 patient plasma. However, not all of these spots represented individual compounds because many of them came from ion source fragmentation or adduct formation. Under our separation condition, PE was first eluted, followed by PS (PI), PC, SM, and lysoPC in a successive manner for phospholipids containing a given fatty acyl composition. Because different molecular species within a class have a same polar head, their retention time in HPLC is very similar. The retention time difference of compounds within one same class is less than that in two different classes, which can be used to align the retention time of phospholipid species in the extraction ion chromatography in order to avoid the retention time fluctuation between different injections. In the negative-ion spectra, molecular species of PE, PS, and PI mainly give the $[M - H]^-$, $[M - 15]^-$, and $[M + 45]^-$ ions for PC, SM, and lysoPC species. The internal standards (C14:0/C14:0 PE, C14:0/C14:0 PS, C14:0/C14:0 PC, and C12:0 lysoPC) were selected based on their solubility and the lack of any demonstrated endogenous molecular ions in that region, which was verified by acquiring a mass spectrum without internal standards. These phospholipid species were quasi-quantified by comparison of the individual ion peak intensity with that of a corresponding internal standard. For PI and SM species, due to the lack of commercial internal standards, and their retention times being close to those of the PS and PC

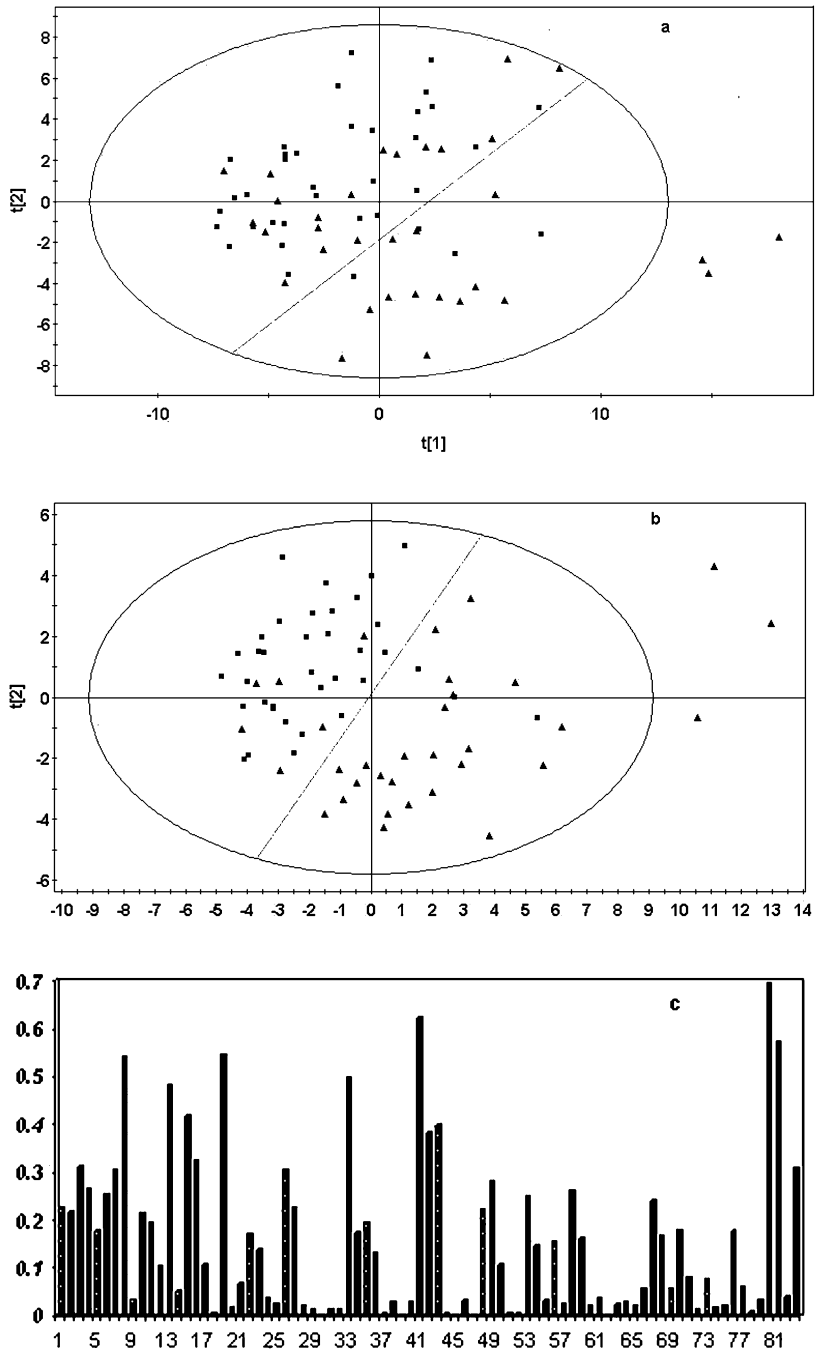

Figure 2. PCA plot and Fisher weights of phospholipid species from 34 DM2 and 35 controls (a) without screening by Fisher weight, (b) with screening by Fisher weight, snd (c) Fisher weights of entire variable. The solid ellipse denotes the 95% significance limit. (■) Control; (▲) DM2.

species, their quasi-quantification was based on PS and PC internal standards (i.e., C14:0/C14:0 PS, C14:0/C14:0 PC), respectively.

After being normalized by the corresponding internal standard, the 83 variables reduced from the LC/MS spectrum were set as columns of **X** matrix; 69 samples (34 DM2 patients and 35 control) were arranged in rows of **X** matrix. Then the **X** matrix was performed to PCA and PLS-DA analysis.

Multivariate Analysis. Our aim of applying multivariate statistical methods to the analysis of phospholipids data was to identify the profiling characteristic for DM2 as compared with control.

First, the 83 variables (phospholipid species) obtained from 34 DM2 patients and 35 control samples were performed with PCA analysis. From Figure 2a, it seemed not difficult to conclude that the pure PCA analysis poorly classified two groups (only with 69.1% correct rate). However, when the variables were screened with the Fisher weight, some improvement was observed in the separation between two groups with 88.2% correct rate (Figure 2b). The cause is that PCA is an unsupervised technique; the excessive irrelevant variables could result in error during the comparison of samples. In contrast, feature extraction of variables is beneficial to unsupervised multivariate analysis. The Fisher weights of these variables are listed in Figure 2c.

Simple, unsupervised chemometrics methods obviously worked well for a data set with a limited number of well-defined classes. Biological systems are seldom simple and many of the biofluid

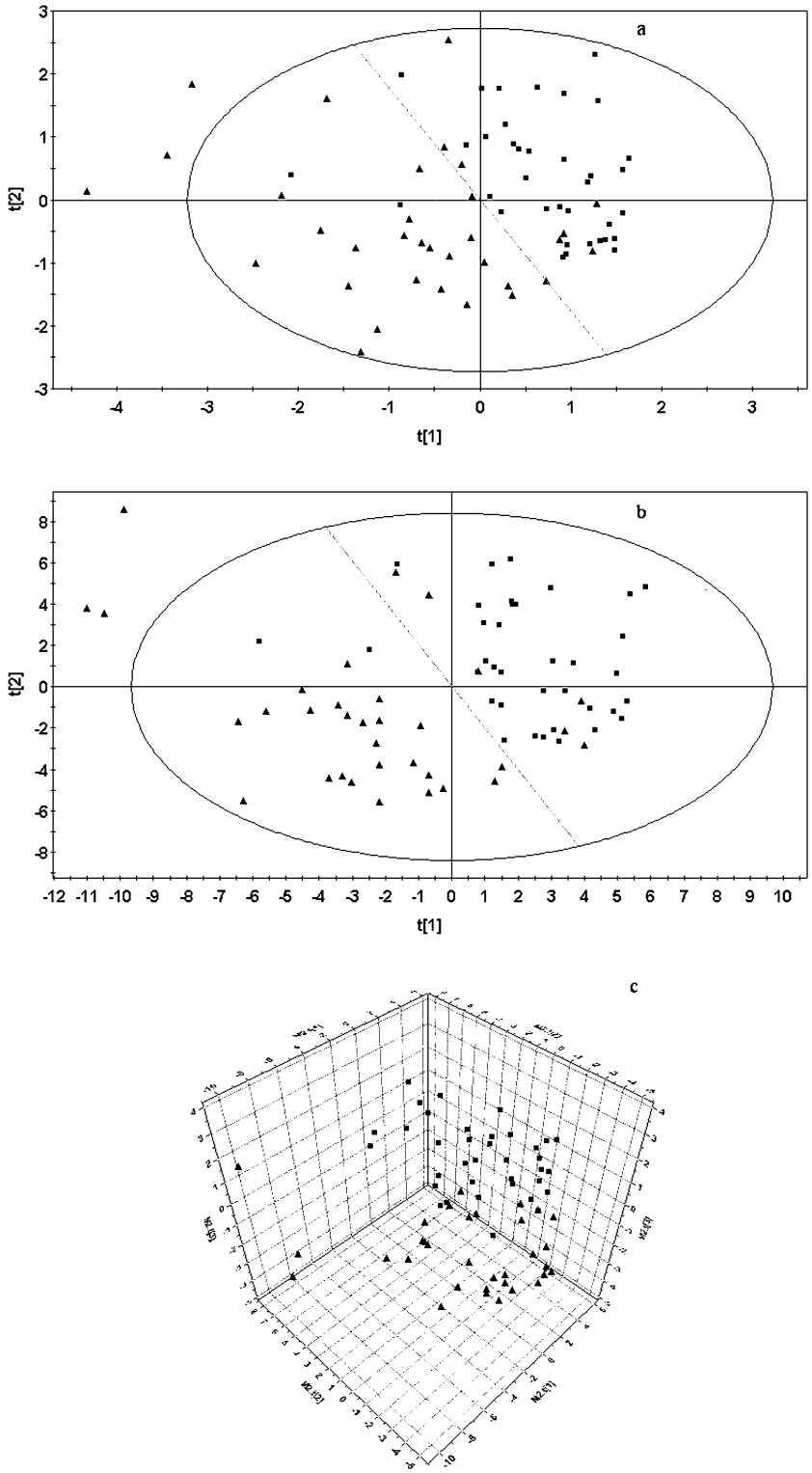

Figure 3. Scores from PLS-DA models classifying DM2 and control. (a) Mean-centered only (Ctr) model; (b) autoscaled (Uv) model; (c) 3D with Uv scaled model. The solid ellipse denotes the 95% significance limit. (■) Control; (▲) DM2.

data sets generated within metabonomics require more sophisticated statistical data analysis. Consequently, various adaptations of principal components-, partial least-squares-, and neural network-based methods have been used to optimize the classification of toxicity or diseases.[42] In the present study, PLS-DA was used to investigate the difference between the DM2 and control. Variable scaling is an integral part of multivariate analysis as it regulates the relative importance of each variable in the subsequent mode.[43]

(41) Holmes E.; Antti, H. *Analyst* **2002**, *127*, 1549−1557.

(42) Jackson, J. E. *A User's Guide to Principal Components*; Wiely: New York, 1991.

(43) Hsu, F. F. *J. Mass Spectrom.* **2000**, *35*, 596−606.

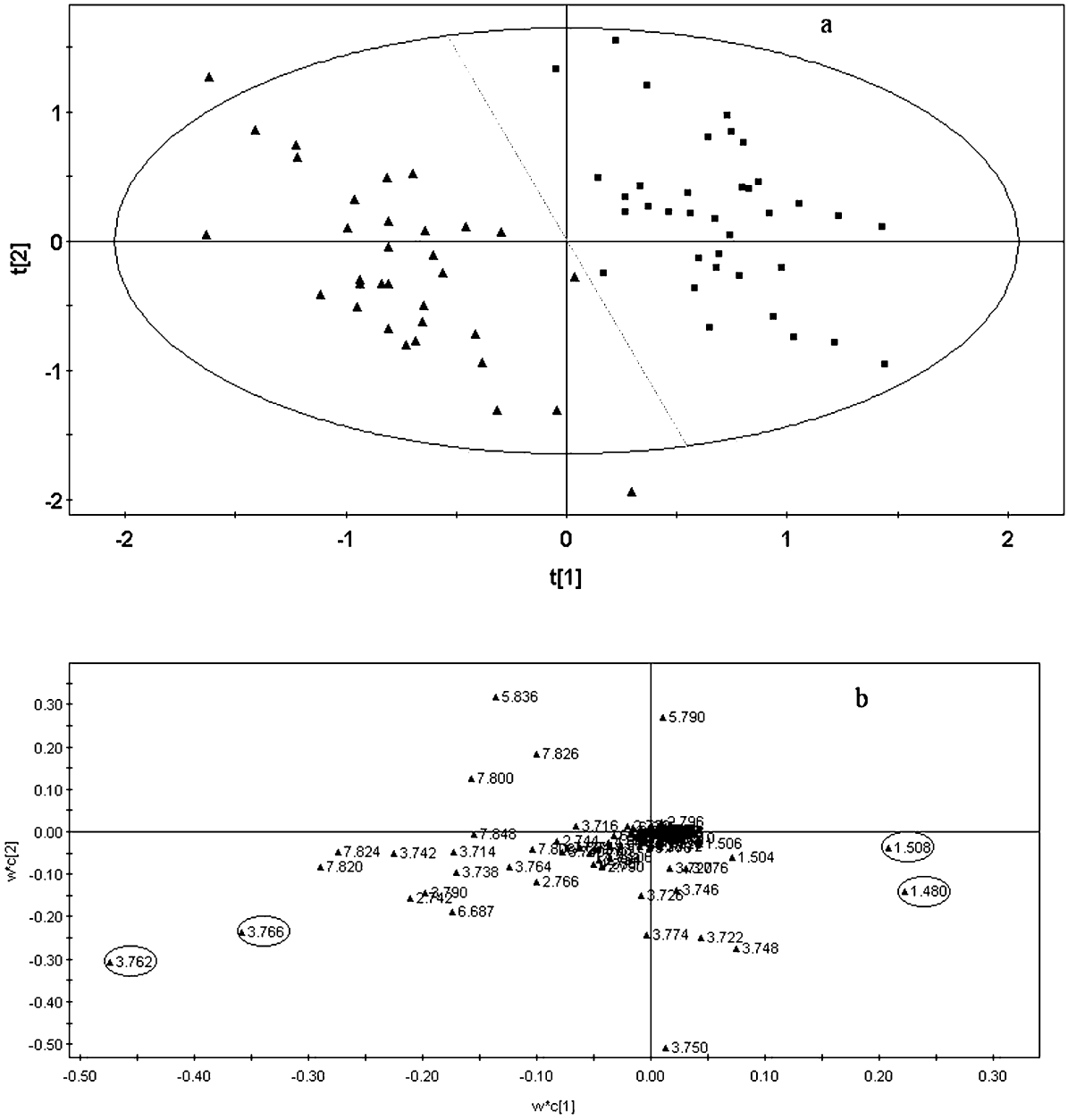

Figure 4. Scores and loadings plots from PLS-DA models classifying DM2 and control with OSC-filtered. (a) Scores plot; the solid ellipse denotes the 95% significance limit. (■) Control; (▲) DM2. (b) Corresponding loadings plot; possible biomarkers are marked with a circle; in loadings figure, data *x*,*y*: *x* represents the different classes of phospholipid species, *and y* represents the *m*/*z* value (rounding values).

Uv is a common practice in the absence of prior information about a variable. The class separation in the PLS-DA model could improve progressively from using just mean-centered data (Ctr) (Figure 3a), after the application of Uv scaling (Figure 3b) and OSC scaling (Figure 4a) to the data set. The living form of mapping a spatial phospholipids profile with Uv scaling was shown in three-way (3D) data matrices (Figure 3c). In other words, that PLS-DA can be used to separate the two classes (control vs DM2).

To determine the effect of different data preprocessing on the predictive ability of the resulting modes, approximately two-thirds of the samples (the "training set") were selected at random to construct a PLS-DA model that could then be used to predict the class membership of the remaining one-third of samples (the "test set"). The classification results are shown in Table 2. From Table 2 it can be seen that the Ctr model showed relatively poor predictive ability with 78.3% prediction rate. Although Uv scaling produced the same prediction rate and specificity comparing to OSC-filtered data, its recognition rate (89.2%) and correct rate (87.0%) were both lower than that of the OSC (95.6% recognition rate, 91.3% correct rate). In a word, the OSC could be used to optimize the class separation; it obviously improved the performance of multivariate pattern recognition analysis and enhanced the predictive power of the model.

Potential Biomarkers. The loadings plot (Figure 4b) indicated the most influential ions that were responsible for separation between sample classes; the ions having the greatest influence on the PLS-DA scores plot were those furthest away from the main cluster of ions. In turn, these compounds might be likely candidates for biomarkers. Figure 4b indicated that the possible biomarkers are those with m/z 508, 480, 766, and 762 (rounding

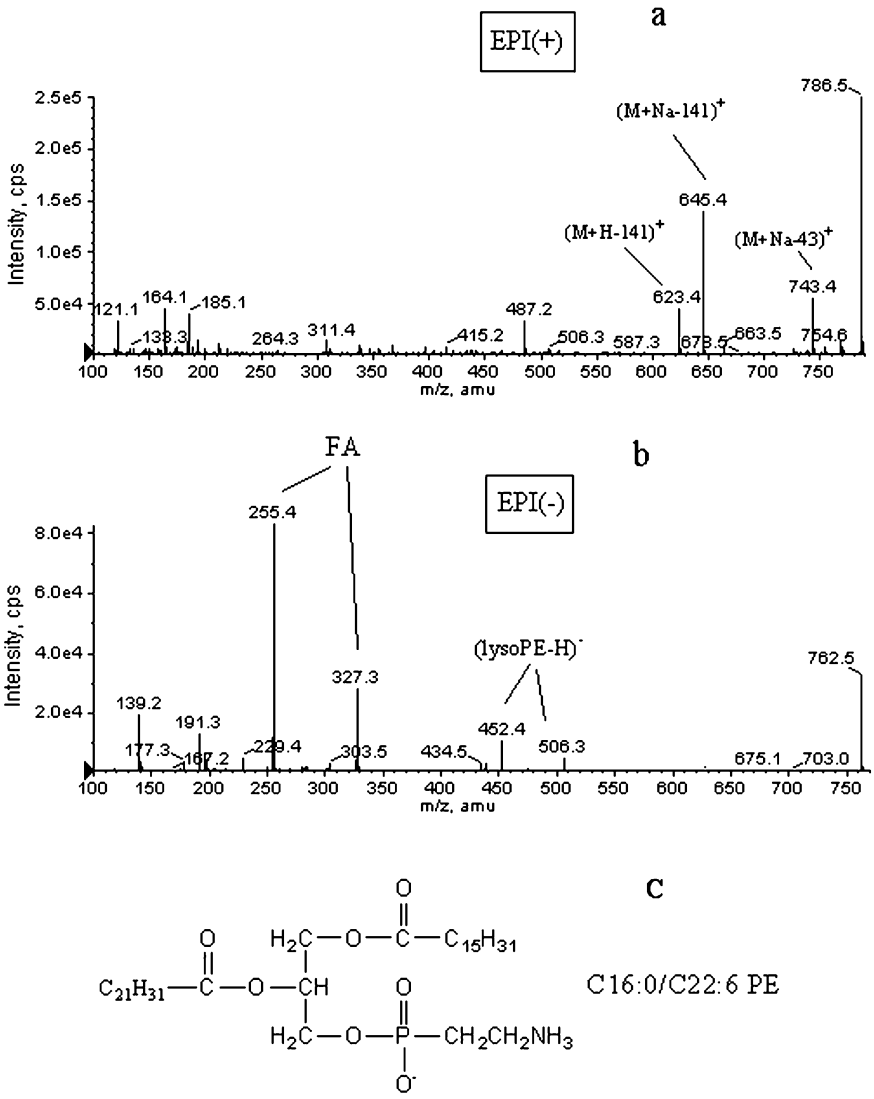

Figure 5. (a) EPI spectrum of *m*/*z* 786.5 representing the $[M + Na]^-$ ion of PE, (b) EPI spectrum of *m*/*z* 762.5 representing the $[M - H]^-$ ion of PE, and (c) structures of this compound. FA- - -fatty acid residues, lysoPE-H- - -PE has lost one of the two fatty acid residues.

Table 2. Classification of the Plasma Phospholipid Data by PLS-DA[a] Method

method	recognition rate	prediction rate	sensitivity	specificity	correct rate
Ctr	87.0% (40/46)	78.3 (18/23)	90.0% (27/30)	88.6% (31/35)	84.1% (58/69)
UV	89.2% (41/46)	82.6% (19/23)	93.3% (28/30)	91.4% (32/35)	87.0% (60/69)
OSC/Ctr	95.6% (44/46)	82.6% (19/23)	93.9% (31/33)	91.4% (32/35)	91.3% (63/69)

[a] Recognition rate is the correct classification of the training set. Prediction rate is the rate of the correct classification of the predicting set. Sensitivity is the number of true positives classified as positive. Specificity is the number of true negatives classified as negative.[51] Ctr, mean-centered unscaled model; Uv, unit variance scaled; OSC/Ctr, OSC-filtered model with no post-OSC scaling.

values) in the negative-ion mode; their identification can be carried out by EPI experiments in the positive-ion or negative-ion mode.

As an example, Figure 5 showed the positive and negative EPI spectra of the compound with *m*/*z* 762.5 in the negative-ion mode. Based on its retention time, it might be glycerophosphoethanolamine (GPE). We first reviewed its corresponded EPI spectra in the positive-ion mode. The CAD tandem mass spectrum of the $(M + H)^+$ at *m*/*z* 764.5 (figure not shown) contained a major ion at *m*/*z* 623.4 $([M + H - 141])^+$ representing the characterization of GPE species, which further confirmed that it was a GPE.[31] But the spectrum did not contain abundant ions to identify fatty acid substituents. In contrast, the CAD tandem mass spectrum of the $(M + Na)^+$ ion at *m*/*z* 786.5 exhibited abundant fragment ions at *m*/*z* 743.4 ($[M + Na - 43]^+$), 645.4 ($[M + Na - 141]^+$), and 623.4 ($[M + H - 141]$), representing neutral loss of aziridine, $(HO)_2P(O)(OCH_2CH_2NH_2)$ and $(NaO)(HO)P(O)(OCH_2CH_2NH_2)$, respectively (Figure 5a). The lower ratio of *m*/*z* 743.4 ($[M + Na - 43]^+$) to 645.4 ($[M + Na - 141]^+$) in diacyl phosphatidylethanolamine (Figure 5b) than the ratio of ($[M + Na - 43]^+$) ion to ($[M + Na - 141]^+$) ion in plasmenylethanolamine indicated that the compound was diacyl phosphatidylethanolamine.[45] These ions were also indicative of the polar headgroup and were diagnostic for the GPE species. The characterization of PE as sodiated adduct was similar to that of lithiated adduct.[43] The fragment ions at *m*/*z*

Table 3. Phospholipid Molecular Species Identification as Potential Biomarkers for Classifying the DM2 Patients from the Control

	negative ion		positive ion		
PL class	quasi-molecular ion	*m/z*	quasi-molecular ion	*m/z*	composition
PE	$[M - H]^-$	762.5	$[M + Na]^+$	786.6	C16: 0/C22: 6
PE	$[M - H]^-$	766.5	$[M + Na]^+$	788.5	C18: 0/C20: 4
lyso-PC	$[M - CH_3]^-$	480.4	$[M + H]^+$	496.4	C16: 0
lyso-PC	$[M - CH_3]^-$	508.4	$[M + H]^+$	524.4	C18: 0

487.2 and 415.2 represented a combined loss of aziridine and the *sn*-1 ($C_{15}H_{31}COOH$) or *sn*-2 ($C_{21}H_{31}COOH$) fatty acid. These ions gave information about the fatty substituents. Figure 5b showed the negative EPI spectrum of diacyl PE species. The fragment ions at *m/z* 506.3 and 434.5 corresponded to dehydrated C22:6 lyso-PE and C16:0 lyso-PE ions (PE lost both one of the two fatty acid residues and one water), respectively. Additionally, deprotonated C16:0 lyso-PE ion (at *m/z* 452.4) was also observed. The fragment ions detected at *m/z* 327.3 and 255.4 corresponded to C22:6 and C16:0 fatty acid residues (carboxylate anion fragments) of $[M - H]^-$ at *m/z* 762.5, respectively. The position of the acyl chains in the glycerol backbone of the phospholipid molecule is obviously important for their degree of dissociation. This phenomenon has previously been reported by several publications.[45–49] There has been a discrepancy about which carboxylate anion yields the most intense peak in the product ion spectrum in these reports. One report stated that the intensity of the fatty acid fragment in the *sn*-2 position was approximately twice that of the fatty acid in the *sn*-1 position,[45] and another report stated that there was no preferential loss of the fatty acid moiety from either *sn*-1 or *sn*-2 position.[46] Hvattum et al.[49] reported that the abundance ratio of the carboxylate anions relates to many factors, such as collision energy, the phospholipid class, and the fatty acids attached to the *sn*-2 position. In this paper, we adopt the idea that the phospholipids isolated from animals most often contain a saturated fatty acid in *sn*-1 position and an unsaturated fatty acid in *sn*-2 position.[50] Therefore, the $[M - H]^-$ at *m/z* 762.5 was identified as C16:0/C22: 6 diacyl PE. Figure 5c showed the possible structures of this compound.

The identification of other potential biomarkers was also performed by positive or negative EPI spectra, Table 3 listed the composition of the phospholipid species as potential biomarkers for classifying the DM2 patients from the control.

CONCLUSIONS

LC/MS and multivariate statistical analysis has been successfully used for the analysis of phospholipid profiling. Both unsupervised PCA and supervised PLS-DA have been applied to classify DM2 and control. For an unsupervised technique, such as PCA, the excessive irrelevant variables could result in error during the comparison of samples. In contrast, feature extraction of variables is beneficial to separate different groups. The utilization of supervised techniques such as PLS-DA makes use of any knowledge of the sample class and maximizes the separation between the two classes. Using the supervised PLS-DA algorithm with Uv scaling and OSC technique to the data set, it was found that the separation of different classes has been highly improved, particularly with OSC. The application of LC/MS coupled to PLS-DA of data with OSC scaling makes it possible to classify DM2 and control and, further, to discover potential biomarkers that can be identified by MS/MS.

The combination of this multivariate approach and LC/MS/MS data-mining metabolite identification program results in a very powerful tool for metabonomics research. These data clearly show that LC/MS offers a viable alternative to, and may be complementary to, proton NMR spectroscopy for applications in disease metabolite profiling.

ACKNOWLEDGMENT

The studies have been supported by the high-tech R & D plan (2003AA223061), National Basic Research Program (2003CB716003) of State Ministry of Science and Technology of China, and the foundation (20425516) for Distinguished Young Scholars from National Natural Science Foundation of China.

Received for review December 23, 2004. Accepted March 16, 2005.

AC0481001

(44) Jonsson, P.; Gullberg, J.; Nordström, A.; Kusano, M.; Kowalaczyk, M.; Sjöström, M.; Moritz, T. *Anal. Chem.* **2004**, *76*, 1738−1745.
(45) Han, X.; Gross, R. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 1202−1210.
(46) Kerwin, J. L.; Tuininga, A. R.; Ericsson, L. H. *J. Lipid Res.* **1994**, *35*, 1102.
(47) Smith, P. B. W.; Snyder, A. P.; Harden, C. S. *Anal. Chem.* **1995**, *67*, 1824.
(48) Chen, S. *Lipids* **1997**, *32*, 85.
(49) Hvattum, E.; Hagelin, G.; Larsen, Å. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1405.
(50) Yorek, M. A. *Phospholipids Handbook*; Cevc, G., Ed,; Marcel Dekker: New York 1993; p 745.
(51) Gehrke, C. W., Kuo, K. C., Eds. *Chromatography and Modification of Nucleosides, Part C*; Elsevier: Amsterdam, 1990.