Extraction and GC/MS Analysis of the Human Blood Plasma Metabolome

ARTICLE *in* ANALYTICAL CHEMISTRY · JANUARY 2006

Impact Factor: 5.64 · DOI: 10.1021/ac051211v · Source: PubMed

8 AUTHORS, INCLUDING:

Johan Trygg
Umeå University
125 PUBLICATIONS **7,325** CITATIONS
SEE PROFILE

Annika Johansson
Umeå University
13 PUBLICATIONS **983** CITATIONS
SEE PROFILE

Pär Jonsson
Umeå University
30 PUBLICATIONS **1,754** CITATIONS
SEE PROFILE

Henrik Antti
Umeå University
89 PUBLICATIONS **5,895** CITATIONS
SEE PROFILE

Anal. Chem. **2005,** *77,* 8086−8094

Extraction and GC/MS Analysis of the Human Blood Plasma Metabolome

Jiye A,[†,‡] Johan Trygg,[§] Jonas Gullberg,[||] Annika I. Johansson,[||] Pär Jonsson,[§] Henrik Antti,[§] Stefan L. Marklund,[†] and Thomas Moritz*,[||]

Department of Medical Biosciences, Clinical Chemistry, Umeå University, SE-90185, Umeå, Sweden, Research Group of Chemometrics, Department of Chemistry, Umeå University, SE-90187, Umeå, Sweden, Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Umeå University, SE-90187, Umeå, Sweden, and Department of Clinical Pharmacology, the First Affiliated Hospital of Nanjing Medical University, 210029, Nanjing, China

Analysis of the entire set of low molecular weight compounds (LMC), the metabolome, could provide deeper insights into mechanisms of disease and novel markers for diagnosis. In the investigation, we developed an extraction and derivatization protocol, using experimental design theory (design of experiment), for analyzing the human blood plasma metabolome by GC/MS. The protocol was optimized by evaluating the data for more than 500 resolved peaks using multivariate statistical tools including principal component analysis and partial least-squares projections to latent structures (PLS). The performance of five organic solvents (methanol, ethanol, acetonitrile, acetone, chloroform), singly and in combination, was investigated to optimize the LMC extraction. PLS analysis demonstrated that methanol extraction was particularly efficient and highly reproducible. The extraction and derivatization conditions were also optimized. Quantitative data for 32 endogenous compounds showed good precision and linearity. In addition, the determined amounts of eight selected compounds agreed well with analyses by independent methods in accredited laboratories, and most of the compounds could be detected at absolute levels of ~0.1 pmol injected, corresponding to plasma concentrations between 0.1 and 1 μM. The results suggest that the method could be usefully integrated into metabolomic studies for various purposes, e.g., for identifying biological markers related to diseases.

Most common diseases are caused by complex interactions between genetic factors, diet, other life style factors, and the environment. All these factors may influence the spectrum and concentrations of metabolites and other low molecular weight compounds (LMC) in tissues and bodily fluids. The LMC that are involved in or affected by disease processes may serve as disease biomarkers. Compounds such as carbohydrates, nucleic acids, amino acids, lipids, various hormones, and phenolics have usually been individually measured in studies of diseases. Recent advances in NMR, GC/MS, or LC/MS methodology, however, have improved the analysis of LMC and thereby enabled more global metabolomic approaches for identifying novel markers for specific diseases, understanding more about the biology, as well as life style and dietary factors behind the disease.[1,2,3] Although it has been applied in relatively few studies as yet, metabolomic analysis has been shown to provide useful information in such diverse areas as clinical diagnostics of coronary heart disease,[4] analysis of genetic disorders,[5] responses to toxins,[6] and characterization of the effects of environmental stressors.[7]

Blood plasma contains a wide variety of chemically diverse LMC, which also vary widely in concentrations and stabilities and are commonly noncovalently bound to proteins. The development of protocols for deproteinization, extraction, and analysis is therefore challenging. Since large numbers of samples are generally analyzed in studies of diseases, single procedures allowing the study of a wide variety of LMC, such as metabolomic GC/MS techniques, are desirable. However, no thoroughly optimized method for analyzing the blood plasma metabolome by GC/MS has been presented to date. Thus, in the present study, we applied experimental design theory (design of experiment; DOE)[8] to optimize a method to extract the blood plasma metabolome and analyze it by GC/MS. For this purpose, the deproteinization and extraction efficiency of five organic solvents for more than 500 detected components in blood plasma, and the effects of varying several other experimental parameters, were tested to optimize the protocols. The data were evaluated by multivariate

* Corresponding author. E-mail: thomas.moritz@genfys.slu.se.

† Department of Medical Biosciences, Clinical Chemistry, Umeå University.

‡ First Affiliated Hospital of Nanjing Medical University.

§ Research Group of Chemometrics, Department of Chemistry, Umeå University.

|| Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Umeå University.

(1) Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. *Trends Biotechnol.* **2004**, *22*, 245−252.

(2) Van der Greef, J.; Stroobant, P.; van der Heijden, R. *Curr. Opin. Chem. Biol.* **2004**, *8*, 559−565.

(3) Nicholson, J. K.; Holmes, E.; Lindon, J. C.; Wilson, I. D. *Nat. Biotechnol.* **2004**, *22*, 1268−1274.

(4) Brindle, J. T.; Antti, H.; E. Holmes, 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.

(5) Griffin, J. L.; Williams, H. J.; Sang, E.; Clarke, K.; Rae, C.; Nicholson, K. J. *Anal. Biochem.* **2001**, *293*, 16−21.

(6) Holmes, E.; Nicholls, A. W.; Lindon, J. C.; Connor, S. C.; Connelly, J. C.; Haselden, J. N.; Damment, S. J. P.; Spraul, M.; Neidig, P.; Nicholson, J. K. *Chem. Res. Toxicol.* **2000**, *13*, 471−478.

(7) Viant, M. R.; Rosenblum, E. S.; Tieerdema, R. S. *Environ. Sci. Technol.* **2003**, *37*, 4982−4989.

(8) Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wikstrom, C.; Wold, S. *Design of Experiments*; Umetrics AB: Umeå, Sweden, 2000.

10.1021/ac051211v CCC: $30.25
Published on Web 11/08/2005

statistical tools including principal component analysis (PCA) and partial least-squares projections to latent structures (PLS). Using the suggested protocol, large amounts of information can be acquired, with high reproducibility, on the blood plasma metabolome.

EXPERIMENTAL SECTION

Reference Compounds, Stable Isotope-Labeled Internal Standards, Reagents, and Human Plasma Sources. The reference compounds were purchased from Sigma (St. Louis. MO), Merck (Darmstadt, Germany), Aldrich (Steinhein, Germany), and Serva (Heidelberg, Germany). The compounds and reagents were all of analytical grade except where stated otherwise. The stable isotope-labeled internal standard compounds (IS), [$^{13}C_5$]-proline, [$^{2}H_4$]-succinic acid, [$^{13}C_5$,^{15}N]-glutamic acid, [1,2,3-$^{13}C_3$]-myristic acid, [$^{2}H_7$]-cholesterol, and [$^{13}C_4$]-disodium α-ketoglutarate were purchased from Cambridge Isotope Laboratories (Andover, MA); [$^{13}C_{12}$]-sucrose, [$^{13}C_4$]-palmitic acid, and [$^{2}H_4$]-butanediamine·2HCl were from Campro (Veenendaal, The Netherlands); and [$^{2}H_6$]-salicylic acid was from Icon (Summit, NJ). *N*-Methyl-*N*-trimethylsilyltrifluoroacetamide (MSTFA) plus 1% trimethylchlorosilane (TMCS) and pyridine (silylation grade) were purchased from Pierce Chemical Co. Stock solutions of the reference compounds and IS were prepared either in Milli-Q water or in methanol at the same concentration, 0.5 $\mu g\ \mu L^{-1}$.

Fresh EDTA anticoagulated blood was collected from 30 healthy volunteers. Blood plasma was prepared by centrifuging at 1600*g* for 10 min at 4 °C. A pool was prepared of all plasma samples, which was stored in aliquots at −80 °C. Plasma was thawed by incubation at 37 °C for 15 min and vortex-mixed before use.

GC/TOFMS Analysis. A 1-μL aliquot of derivatized sample was injected splitless by an Agilent 7683 Series autosampler (Agilent, Atlanta, GA) into an Agilent 6980 GC equipped with a 10 m × 0.18 mm i.d. fused-silica capillary column chemically bonded with 0.18-μm DB5-MS stationary phase (J&W Scientific, Folsom, CA). The injector temperature was set at 270 °C. Helium was used as carrier gas at a constant flow rate of 1 mL min^{-1} through the column. For every analysis, the purge time was set to 60 s at a purge flow rate of 20 mL min^{-1} and an equilibration time of 1 min. The column temperature was initially kept at 70 °C for 2 min and then increased from 70 to 320 °C at 30 °C min^{-1}, where it was held for 2 min. The column effluent was introduced into the ion source of a Pegasus III TOFMS (Leco Corp., St Joseph, MI). The transfer line temperature was set at 250 °C and ion source temperature at 200 °C. Ions were generated by a 70-eV electron beam at a current of 2.0 mA. Masses were acquired from m/z 50 to 800 at a rate of 30 spectra s^{-1}, and the acceleration voltage was turned on after a solvent delay of 170 s.

Analysis of GC/MS Data. To evaluate the extraction protocols, nonprocessed MS files from GC/TOFMS analysis were exported in NetCDF format to MATLAB software 6.5 (Mathworks, Natick, MA), where all data pretreatment procedures, such as baseline correction, chromatogram alignment, time-window setting, and multivariate curve resolution (MCR)[9] were performed using custom scripts.

(9) Jonsson, P.; Johansson, A.; Gullberg, J.; Trygg, J.; A, J.; Grund, B.; Marklund, S.; Sjöström, M.; Antti, H.; Moritz, T. *Anal. Chem.* **2005**, *77*, 5635−5642.

ChromaTOF 2.00 software (Leco Corp.) was used for automatic peak detection and calculating peak areas of IS and specific compounds. Automatic peak detection and mass spectrum deconvolution with the ChromaTOF software were performed with peak width set to 2 s. Peaks with lower signal-to-noise (S/N) ratios than 10 were rejected. To obtain accurate peak areas for the IS and specific compounds, two unique quantification masses for each component were specified and the data were reprocessed.

Mass spectra of all detected compounds were compared with spectra in the NIST library 2.0 (as of January 31, 2001), the in-house mass spectra library database established by Umeå Plant Science Center, or the mass spectra library maintained by the Max Planck Institute in Golm (http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html).

Experimental Design 1: Extraction Solvents. The ability of five organic solvents (methanol, ethanol, acetonitrile, acetone, chloroform) to extract LMC from blood plasma was investigated. The total solvent volume for each experiment was fixed at 800 μL, and the volume of each solvent included was varied between 0 and 800 μL, except for chloroform, which was limited to a maximum of 200 μL to avoid the formation of two-phase systems. A D-optimal experimental design[8] was constructed with 24 different extraction/deproteinization mixtures of solvents, resulting in a total (including replicates) of 41 experiments (Table 1). The following steps were involved in the experiments. First, each IS (20 μg in an individually prepared stock solution) was placed in a glass tube. The solution was evaporated to dryness, 2 mL of plasma was added to the residue, and the tube was vigorously vortex-mixed for 2 min, ultrasonicated to dissolve compounds in a Transsonic T460 ultrasonic bath (Elma, Germany) at 35 kHz and room temperature for 2 × 3 min, and then incubated at 37 °C for 30 min. The plasma was then filtered through a 0.22-μm filter (Millipore Corp.) to remove undissolved particles. A 200-μL portion of the pooled plasma was added to 800 μL of the solvent mixture stipulated by the experimental design in an Eppendorf tube. The solution was mixed for 10 s, kept on ice for 10 min, and vigorously extracted at a frequency of 30 Hz for 3 min using a MM301 vibration Mill (Retsch GmbH & Co. KG, Haan, Germany). After 10 min on ice, the solution was centrifuged at 19600*g* for 10 min at 4 °C. Then 200 μL of the resulting supernatant was transferred to a GC vial and evaporated to dryness in a Speed-vac Concentrator (Savant Instrument, Framingdale, NY). A 30-μL sample of methoxyamine (15 $\mu g\ \mu L^{-1}$) in pyridine was added to each GC vial, and the resultant mixture was vigorously vortex mixed for 10 min. Methoxymation was carried out at room temperature for 16 h. The solution was then vortex mixed again for 10 min after adding 30 μL of MSTFA with 1% TMCS as catalyst. After silylation for 1 h, 40 μL of heptane was added to each GC vial. All samples were prepared and analyzed according to two different, randomized orders, one for sample preparation and the other for GC/MS analysis.

Experimental Design 2: Experimental Parameters. To optimize the extraction, methoxymation, and silylation steps (and, thus, the overall experimental protocol), the effects of varying the following experimental factors were investigated: methanol volume; extraction duration; temperature and duration of the incubations before and after extraction; and temperature and duration of the methoxymation and silylation. The experimental setup was

Table 1. Experimental Design for Investigating the Extraction Efficiency of Different Organic Solvents

		amount, μL					
ID no	run order	methanol	ethanol	acetonitrile	acetone	chloroform	plasma
N1[a]	20, 14, 18	800	0	0	0	0	200
N2	6	0	800	0	0	0	200
N3	31	0	0	800	0	0	200
N4	11	0	0	0	800	0	200
N5	24	600	0	0	0	200	200
N6	18	0	600	0	0	200	200
N7[b]	3	0	0	600	0	200	200
N8[c]	19, 12, 21	0	0	0	600	200	200
N9	1	0	0	0	735	65	200
N10	23	0	0	535	265	0	200
N11	15	0	0	265	535	0	200
N12	30	0	535	0	265	0	200
N13	25	0	265	0	535	0	200
N14	16	0	535	265	0	0	200
N15	27	0	265	535	0	0	200
N16[a]	29, 9, 17	665	0	0	0	135	200
N17	4	535	0	0	265	0	200
N18	5	265	0	0	535	0	200
N19	13	535	0	265	0	0	200
N20	22	265	0	535	0	0	200
N21	32	535	265	0	0	0	200
N22	8	265	535	0	0	0	200
N23	7	0	235	235	235	100	200
N24[d]	26, 2, 33, 10	200	200	200	200	0	200

[a] Triplicates. [b] Due to the formation of two phases in N7, the original ratio of acetonitrile to chloroform was changed from 600:200 to 770:30. [c] Triplicates; due to the formation of two phases in N8 the original ratios of acetone to chloroform was changed from 600:200 to 700:100. [d] Four replicates.

Table 2. Experimental Design for Optimizing the Extraction and Derivatization Conditions

expt no.	methanol, μL	incubation[a] °C, min	extraction min	incubation[b] °C, min	oximation °C, h	silylation °C, h
N1	700	0, 10	1	0, 10	20, 16	20, 1
N2	700	70, 30	1	0, 10	70, 1; 20,15	20, 1
N3	700	0, 10	3	0, 10	70, 1; 20,15	70, 1
N4	700	70, 30	3	0, 10	20, 16	70, 1
N5	700	0, 10	1	−20, 120	70, 1; 20,15	70, 1
N6	700	70, 30	1	−20, 120	20, 16	70, 1
N7	700	0, 10	3	−20, 120	20, 16	20, 1
N8	700	70, 30	3	−20, 120	70, 1; 20,15	20, 1
N9	900	0, 10	1	0, 10	20, 16	70, 1
N10	900	70, 30	1	0, 10	70, 1; 20,15	70, 1
N11	900	0, 10	3	0, 10	70, 1; 20,15	20, 1
N12	900	70, 30	3	0, 10	20, 16	20, 1
N13	900	0, 10	1	−20, 120	70, 1; 20,15	20, 1
N14	900	70, 30	1	−20, 120	20, 16	20, 1
N15	900	0, 10	3	−20, 120	20, 16	70, 1
N16	900	70, 30	3	−20, 120	70, 1; 20,15	70, 1
N17	800	0, 10	2	0, 10	20, 16	20, 1
N18	800	0, 10	2	0, 10	20, 16	20, 1
N19	800	0, 10	2	0, 10	20, 16	20, 1
N20	800	0, 10	2	0, 10	20, 16	20, 1

[a] Temperature and duration before extraction. [b] Temperature and duration after extraction.

based on a 2^{6-2} fractional factorial design[8] (Table 2). A 100-μL sample of plasma was added to 700, 800, or 900 μL of methanol, according to the design, and water was added to bring the total volume to 1000 μL. The solution was mixed for 10 s, kept on ice for 10 min or 70 °C for 30 min, and then vigorously extracted for 1, 2, or 3 min. After 10 or 120 min on ice, the tubes were centrifuged at 19600*g* for 10 min at 4 °C. A 200-μL sample of supernatant was then transferred to a GC vial and evaporated to dryness. Methoxymation was carried out either at room temperature for 16 h or first at 70 °C for 1 h and then at room temperature for 15 h. The samples were subsequently trimethylsilylated at 70 °C or room temperature for 1 h. After adding 40 μL of heptane (containing 0.5 μg of methyl stearate as an internal standard) and vortex mixing for 10 s, 1 μL of the derivatized sample was analyzed by GC/TOFMS.

Experimental Design 3: Methanol Volume. The results of our studies suggested that methanol alone was the best of the tested solvents for extracting LMC from blood plasma. To determine the optimal proportion of methanol, 13 experiments were carried out in which the methanol volume was varied at five

levels, keeping all other experimental conditions constant. A total of 700, 750, 800, 850, or 900 μL was added to 100 μL of plasma and water to a total a volume of 1000 μL. The solution was vortex mixed for 10 s, kept on ice for 10 min, and then vigorously extracted for 2 min. After 120 min on ice, the samples were centrifuged at 19600*g* for 10 min at 4 °C. A 200-μL aliquot of supernatant was transferred to a GC vial and evaporated to dryness. Methoxymation was carried out at room temperature for 16 h. Finally, the samples were trimethylsilylated at room temperature for 1 h, after which 30 μL of heptane (containing 0.5 μg of methyl stearate as internal standard) was added. A 1-μL sample of each derivatized sample was analyzed by GC/TOFMS.

Linearity and Precision. Pooled plasma was diluted with water to relative concentrations of 0.063, 0.125, 0.250, 0.500, and 1.000 (v/v, plasma/plasma+water). A 200-μL sample of the diluted plasma, including 2 μg of water-soluble IS, was then added to 800 μL of methanol, including 2 μg of methanol-soluble IS in an Eppendorf tube. Extraction and derivatization was performed according to experimental design 3. After GC/TOFMS analysis, the peak areas of 32 LMC and 10 IS were integrated. The ratio of each LMC and a corresponding IS was calculated, and linear correlation coefficients were calculated in the analyzed concentration interval. Precision was investigated by calculating the relative standard deviation (RSD) at three plasma dilutions: 0.063, 0.250, and 1.000 (v/v).

Quantification of Endogenous Plasma Compounds. The endogenous concentrations of 10 compounds known to be present in human plasma (proline, glutamic acid, phenylalanine, methionine, lysine, tryptophan, citric acid, *myo*-inositol, creatinine, uric acid) were determined by the standard addition method.[10] First, aqueous stock solutions of the 10 compounds were prepared. Then a series of four 1:1 (v/v) plasma/water preparations were made, including 2 μg of each water-soluble IS. Differing amounts of the stock solutions were also added to three of these preparations (and none to the fourth, which provided a control), so that the response of each compound (normalized by the IS) when analyzed by GC/MS increased by roughly 20, 50, and 100% compared to the control. A 200-μL sample of each of the plasma preparations were then added to 800 μL of methanol, including 2 μg of methanol-soluble IS, in an Eppendorf tube. After extraction and derivatization according to experimental design 3, and GC/TOFMS analysis, the peak areas for each compound and IS were integrated. The ratio of each LMC and a corresponding IS was calculated as the response. The concentrations of the 10 compounds were calculated by linear regression, where the endogenous concentration of the compounds was determined as the *y*-intercept of the regression curves. Concentrations of proline, glutamate, palmitate, and nonesterified cholesterol were also calculated directly from the corresponding IS with known concentration. For comparison, a number of analyses were carried out at accredited clinical chemistry laboratories. Plasma amino acids concentrations of the six amino acids were analyzed on a Biotronic LC 3000 amino acids analyzer (Eppendorf Biotronic, Hamburg, Germany. Huddinge University Hospital, Stockholm, Sweden) calibrated with a standard amino acid solution (Sigma, product no. A9906). Creatinine and urate were analyzed on a Vitros 950 apparatus (Ortho Clinical Diagnostics, Umeå University Hospital, Umeå, Sweden) calibrated with Vitros Performance controls.

Limit of Detection (LOD). To determine the LOD for the 12 endogenous compounds, and the minimum extraction volume that could be used without adversely affecting the analysis, plasma dilutions were prepared at a concentration of 0.100 (plasma/plasma+water; v/v). To extract LMC, 40 μL of the diluted plasma was then added to 160 μL of methanol. After extraction and centrifugation, 100.0, 50.0, 25.0, 10.0, 5.0, 2.5, and 1.0 μL of the supernatants were transferred to GC vials and evaporated to dryness. A 15-μL sample of methoxyamine (15 μg/μL) in pyridine and 15 μL of MSTFA with 1% TMCS were added for methoxymation and trimethylsilylation, respectively, according to experimental design 3. Finally, 20 μL of heptane was added to each GC vial, and 1 μL of the derivatized sample was analyzed by GC/TOFMS. The peak areas for specific quantification masses for each of the 12 compounds were determined, and their LODs (S/N = 5) were calculated.

Multivariate Data Analysis and Design of Experiments. All data analysis and modeling were done using Modde 7.0 or Simca 11 Beta software (Umetrics, Umeå, Sweden). Modde 7.0 was also used to generate design matrixes for the different experimental studies (DOE). PCA and PLS were used to relate the design matrix **X** (e.g., the different solvent mixtures) to the peak areas of all resolved GC/TOFMS peaks (**Y** matrix). Cross-validation[11] with seven cross-validation groups was used throughout to determine the number of components. The following statistics for the regression models are discussed throughout this paper. R^2X is the cumulative modeled variation in **X**, R^2Y is the cumulative modeled variation in **Y**, and Q^2Y is the cumulative predicted variation in **Y**, according to cross-validation. The range of these parameters is 0−1, where 1 indicates a perfect fit.

RESULTS

Using design-of-experiment procedures, together with multivariate analysis, we have investigated how different parameters (choice of extraction solvents, derivatization, physical conditions) affect the extraction and derivatization of LMC from human blood plasma. The experimental work was divided into four sections: (A) screening for solvents affecting extraction efficiency; (B) investigation of how different extraction and derivatization conditions affect the detection of LMC; (C) fine-tuning the proportion of methanol for optimal analysis of LMC; (D) investigation of the linearity, limits of detection, and precision of the suggested extraction method.

All samples were analyzed by GC/TOFMS, and data were processed either by a recently developed hierarchical multicurve resolution (H-MCR) method[9] or ChromaTOF software.

Optimization of the Extraction Solvent. A D-optimal design was set to investigate how five commonly used solvents (methanol, ethanol, acetonitrile, acetone, chloroform) for protein precipitation affect the efficiency of extracting LMC from human blood plasma. The GC/TOFMS data obtained from the 41 samples were processed using the MCR method, and 501 peaks were resolved, 80 of which were identified (see Supporting Information Table S1). A PCA model of the 501 resolved peaks from the 41 samples

(10) Musfeld, C.; Biollaz, J.; Belaz, N.; Kesselring, U. W.; Decosterd L. A. *J. Pharm. Biomed. Anal.* **2001**, *24*, 391−404.

(11) Wold, S. *Technometrics* **1978**, *20*, 397−405.

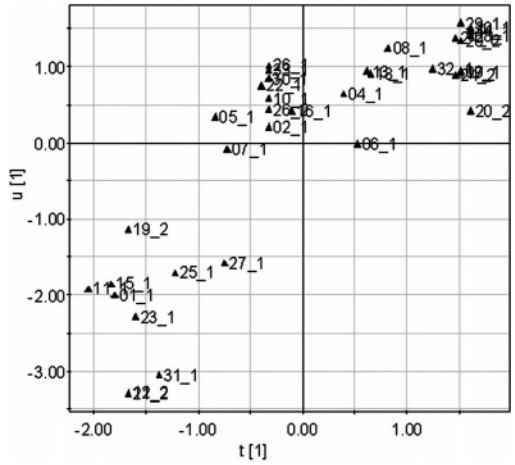

Figure 1. Determination of optimal solvents for extracting LMC from human blood plasma. The PLS scores plot (t1−u1) shows the inner correlation structure between the X matrix (i.e., the experimental conditions/solvents; plots labeled by run order, see Table 1) and the Y matrix (i.e., the areas of the resolved GC/TOFMS peaks). Approximately 30% of the variation in the resolved peak areas can be explained by the model. The lower left quadrant of the plot includes experiments that produced the lowest overall peak areas, while the upper right quadrant includes experiments that yielded the best results in terms of peak areas.

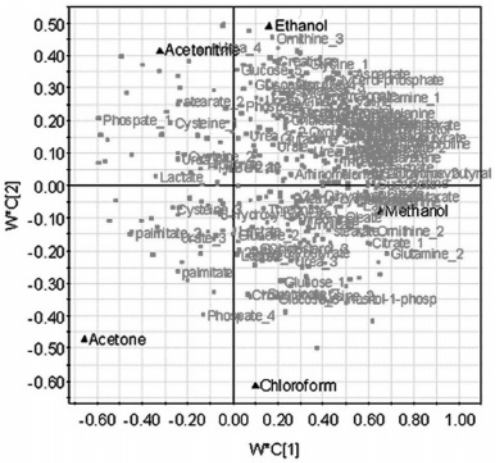

Figure 2. Determination of optimal solvents for extracting LMC from human blood plasma. The PLS loading plot summarizes the influence and correlation structure between variables in both the **X** matrix (▲, design variables) and **Y** matrix (■, the areas of the resolved GC/TOFMS peaks). Methanol is strongly positively correlated with a majority of the resolved peak areas.

showed that 5 samples were outliers (data not shown). These outliers were found to be due to the formation of two phases in some solvent mixtures including chloroform. These samples were removed from further modeling, and the GC/TOFMS data processing was repeated, as the outliers could influence the deconvolution step. A new PCA model with the remaining 36 samples was calculated, and no additional outliers were found. It is common practice to normalize the resolved GC/MS profiles using the added internal standards in order to remove any systematic differences in the data. However, in this study, the influence of the different experimental conditions was investigated. Hence, normalization should only have been done if the internal standards were not affected by the different solvents; otherwise, much of the variation between the different experiments would have been lost, leading to erroneous conclusions. A PLS model relating the design matrix **X** (i.e., experimental conditions) to the 10 added internal standards (**Y** matrix) was calculated to determine the possible effects of the different solvents on the quantification of the internal standards. A PLS model with one significant component ($R^2X = 0.3$, $R^2Y = 0.4$, $Q^2Y = 0.3$), according to cross-validation, was found. This showed that the added IS peak areas were systematically influenced by the different solvents. Hence, no normalization was applied in this study.

A two-component PLS model was generated between the design matrix **X** and the 501 resolved peak areas (**Y** matrix) of the 36 samples, that explained, according to the R^2X, R^2Y, and Q^2 parameters (0.6, 0.3, and 0.2, respectively), ~30% of the variation in the resolved peak areas. No correlation with retention time order was found for the resolved peak areas (data not shown). In Figure 1, the lower left quadrant of the plots includes experiments (Table 1) that produced the lowest overall peak areas. Analogously, the upper right quadrant includes experiments that yielded the best results in terms of peak areas.

A PLS loading plot is shown in Figure 2. The majority of the resolved peaks are strongly and positively correlated with methanol. The peak areas of only a small minority of the detected compounds (e.g., ornithine, phosphate, and cholesterol) are enhanced by acetonitrile-, acetone-, ethanol-, or chloroform-containing solvents. Peaks closer to the center in the loading plot (Figure 2) are less affected by the composition of the solvent mixture. Hence, the results of this experimental study show that a solvent consisting of methanol alone is the best of the tested options to maximize the peak area for the majority of the 501 resolved peaks.

Optimization of Extraction and Derivatization Parameters. After screening for the best solvent for extracting LMC from blood plasma, a 2^{6-2} fractional factorial design was constructed for optimizing the extraction and derivatization conditions (Table 2). The experimental factors investigated were the following: methanol volume; extraction duration; temperature and duration of the incubation before and after extraction; and temperature and duration of the methoxymation and silylation. After processing the GC/TOFMS data from the 20 samples using the MCR method, 518 resolved peaks were obtained. A PLS model was calculated between the design matrix **X** (i.e., experimental conditions) and the peak areas of the 10 internal standards (**Y** matrix) to determine whether the internal standards are systematically affected by the different experimental conditions or not. A one-component PLS model was found to give strong correlations ($R^2X = 0.2$, $R^2Y = 0.8$, $Q^2Y = 0.5$). The most important factor in this model is the amount of methanol (700−900 μL), which is negatively correlated with the peak areas of all internal standards. This systematic influence of the experimental conditions on the internal standards suggests that no normalization using these internal standards should be performed, as in the solvent design (see above). Hence, the only normalization performed was against methyl stearate, which was added after the derivatization as a GC/MS internal standard and was not therefore affected by any of the experimental conditions investigated.

A PLS model between the design matrix **X** (experimental conditions) and the **Y** matrix (all 518 resolved peaks) was

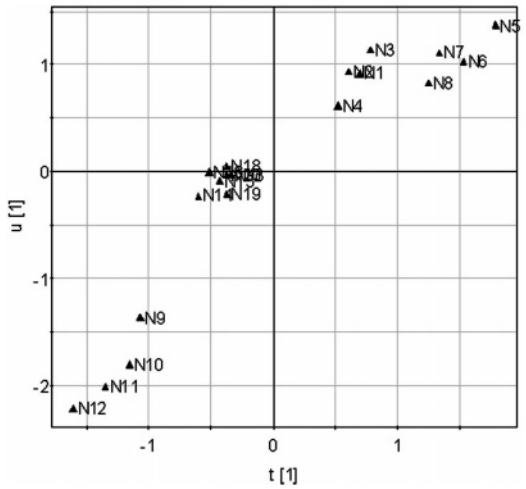

Figure 3. Results from the optimization of extraction and derivatization conditions, experimental design 2. See Table 2. The PLS score plot shows a strong inner correlation structure between the various experimental conditions (**X** matrix) and the areas of the resolved GC/TOFMS peaks (**Y** matrix). The upper right and lower left quadrant show the experimental conditions that yielded the highest and lowest overall GC/TOFMS peak areas, respectively.

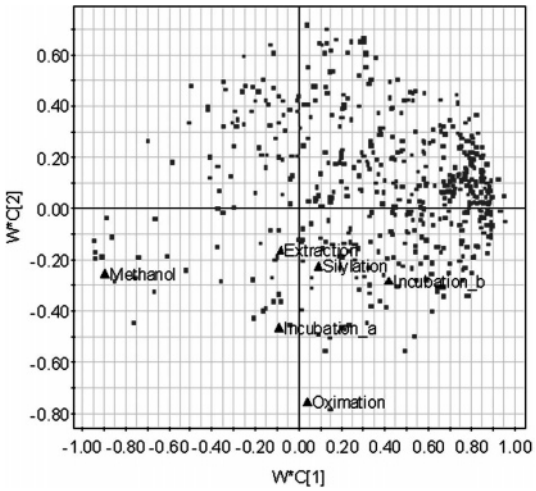

Figure 4. Results from the optimization of extraction and derivatization conditions. The PLS loading plot shows the correlation structure between the experimental settings, **X** (▲) and resolved peaks, **Y** (■). The most important factor influencing the peak areas for the resolved components is the amount of methanol (700−900 μL), which is negatively correlated to the majority of the resolved peaks.

calculated. The resulting two-component PLS model ($R^2X = 0.4$, $R^2Y = 0.4$, $Q^2Y = 0.2$) reveals a strong inner relation between X and Y (Figure 3). The three groupings in the figure are due to the experimental design settings (Table 2; −1, 0, +1). According to the model, experiments 5−8 resulted in the highest overall peak areas.

The PLS loading plot shows the correlation structure between the X and Y variables (Figure 4). Two main factors influence the peak areas: the amount of methanol and the methoxymation conditions, both of which are negatively correlated to the majority of the resolved peaks. This means that the amount of methanol should be kept at 700 μL rather than 900 μL (when the total extraction volume is 1000 μL, including 100 μL of plasma) and the methoxymation step should be performed for 16 h at room temperature. The model also suggests that the remaining factors

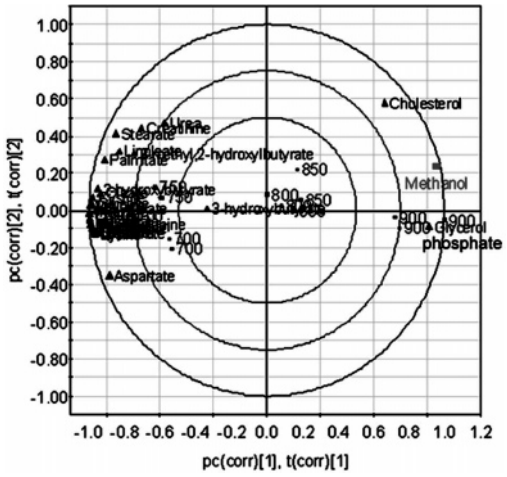

Figure 5. Optimization of the methanol volume for extracting LMC from human blood plasma. The PLS biplot shows the correlation between peak areas for 32 endogenous metabolites and the methanol/plasma+water ratio (tested ratios included 700:300, 750:250, 800:200, 850:150, and 900:100. Higher ratios of methanol favor extraction of lipophilic compounds, such as cholesterol.

(extraction time, silylation, incubation steps) do not strongly affect the peak areas (Table 2, Figure 4). This was also confirmed in a parallel experiment where the effect of extraction temperature were investigated. The results show that low extraction temperature (−20 °C) has no effect or only slightly positive effect (acetone) on the peak area for the resolved peaks compared to extraction at room temperature (data not shown).

Optimization of the Methanol Volume. The optimization of extraction and derivatization conditions showed that 700 μL of methanol in a total volume of 1000 μL including 100 μL of plasma is optimal for extracting LMC (Figure 4). However, to investigate how the proportion of methanol affects specific metabolites more closely, 13 new experiments were carried out in which the proportion of methanol was varied at five levels while all other experimental conditions were kept constant. After processing the GC/TOFMS data using the ChromaTOF software, the peak areas of 32 endogenous metabolites were obtained. A two-component PLS model ($R^2X = 0.9$, $R^2Y = 0.9$, $Q^2Y = 0.9$) was generated between the extracted endogenous metabolites and the proportion of methanol used in the extraction phase. In Figure 5, the PLS biplot shows that a higher ratio of methanol is preferable for hydrophobic compounds, but negative for hydrophilic compounds. Comparison of the peak areas (data not shown) of most of the compounds, which are highly water-soluble, with those of cholesterol, which have low water solubility, suggests that an acceptable compromise between efficient extraction of polar and nonpolar metabolites can be reached by using a methanol volume of 800 μL (8:1:1, MeOH/H_2O/plasma v/v/v).

Linearity and Precision. Thirty-two of the identified endogenous compounds were selected to investigate the precision and linearity of the method. These compounds covered a wide span of GC retention times, and they belong to different classes of compounds with various physicochemical properties, such as organic acids, amino acids, fatty acids, lipids, and carbohydrates. With the optimized extraction procedure, it was possible to detect and quantify the corresponding endogenous metabolite for 4 out of the 10 isotope-labeled IS (palmitic acid, glutamic acid, proline,

Table 3. RSDs and Linearity Correlation Coefficients for 32 Endogenous Compounds in Human Blood Plasma Analyzed by GC/TOFMS[a]

compound	RSD (%) at concns (v/v) of 0.063	0.250	1.000	correlation coefficient	corresponding IS
alanine	15.0	9.4	4.7	0.9996	[$^{13}C_5$]-proline
2-hydroxybutyrate	12.8	6.1	4.8	1.0000	[$^{2}H_4$]-succinic acid
3-hydroxybutyrate	12.5	11.4	4.1	0.9998	[$^{13}C_5$]-proline
3-methyl, 2-hydroxybutyrate	11.3	6.4	9.0	0.9999	[$^{2}H_4$]-succinic acid
valine	12.2	19.4	6.8	0.9997	[$^{13}C_5$]-proline
leucine	13.6	15.9	9.9	0.9942	[$^{13}C_5$]-proline
urea	4.7	2.4	10.4	0.9927[b]	[$^{2}H_4$]-butanediamine
proline	5.4	7.5	10.9	0.9963[b]	[$^{13}C_5$]-proline
glycerate	11.2	14.2	7.4	1.0000	[$^{2}H_4$]-succinic acid
glycine	16.1	14.9	16.9	0.9925	[$^{13}C_5$]-proline
serine	8.6	10.7	10.4	0.9996	[$^{13}C_5$]-proline
threonine	6.9	6.7	8.2	0.9999	[$^{13}C_5$]-proline
methionine	11.1	15.1	14.4	0.9984	[$^{13}C_5$,^{15}N]-glutamic acid
creatinine	29.1	23.8	11.5	0.9975	[$^{13}C_5$,^{15}N]-glutamic acid
aspartate	25.8	23.8	22.4	0.9898	[$^{13}C_5$,^{15}N]-glutamic acid
phenylalanine	17.8	11.5	8.4	0.9996	[$^{2}H_6$]-salicylic acid
glutamate	12.8	6.8	6.9	0.9996	[$^{13}C_5$,^{15}N]-glutamic acid
laurate	23.1	16.5	19.9	0.9991	[1,2,3-$^{13}C_3$]-myristic acid
glycerol phosphate isomer	11.0	23.8	3.7	0.9956	[1,2,3-$^{13}C_3$]-myristic acid
citrate	7.3	3.5	6.6	0.9999	[$^{13}C_5$,^{15}N]-glutamic acid
glucose	14.9	13.0	22.5	0.9731[b]	[$^{13}C_{12}$]-sucrose
lysine	26.3	20.9	8.6	0.9961	[$^{13}C_5$,^{15}N]-glutamic acid
tyrosine	12.8	13.0	10.4	0.9995	[$^{13}C_5$, ^{15}N]-glutamic acid
palmitate	7.3	6.6	1.1	0.9998	[$^{13}C_4$]-palmitic acid
myo-inositol	19.6	8.6	7.4	0.9997	[$^{13}C_{12}$]-sucrose
urate	13.8	7.7	8.2	0.9689[b]	[$^{13}C_4$]-palmitic acid
tryptophan	19.7	18.6	10.7	1.0000	[$^{13}C_5$,^{15}N]-glutamic acid
linoleate	4.6	11.0	10.5	0.9999	[$^{13}C_4$]-palmitic acid
oleate	8.1	7.7	5.4	0.9998	[$^{13}C_4$]- palmitic acid
stearate	8.3	9.6	5.8	0.9988	[$^{13}C_4$]-palmitic acid
cystine	16.3	17.8	17.1	0.9942	[$^{13}C_5$,^{15}N]-glutamic acid
cholesterol, nonesterified	12.6	17.7	28.9	0.9500[b]	[$^{2}H_7$]-cholesterol

[a] The plasma dilutions cover concentrations from 0.063 to 1.000 (v/v, $n = 5$). [b] The correlation coefficients improve to 0.9971, 0.9989, 0.9997, 0.9998, and 1.0000 for glucose, cholesterol, urea, urate, and proline, respectively, when data obtained from experiments with the highest concentration of plasma are excluded from the linear regression.

cholesterol). The remaining 28 metabolites were quantified using an IS with similar chemical properties or retention times (Table 3). The precision of the analysis was calculated as the relative standard deviation of the peak area for each metabolite corrected by the peak area of the IS. The RSD of most of the 32 compounds was less than 20%, and less than 10% for many of them, at the three plasma concentrations (Table 3). Generally the lowest concentrations of plasma resulted in the highest RSD.

The linearity of the response was determined by analyzing plasma at five different concentrations and found to be generally high for most of the compounds investigated. Exceptions were compounds such as glucose and cholesterol, which occur in high concentrations in plasma (Table 3). For these compounds, the precision of the analysis was especially low at high plasma concentrations. Therefore, linearity was improved (to >0.997) by excluding data from experiments in which the highest concentration of plasma was used.

Validation of Method and Determination of Detection Limits. The method developed for the global analysis of LMC in blood plasma was validated by determining the absolute concentrations of 12 endogenous metabolites and comparing the results for 6 amino acids, urate, and creatinine from our method with those from accredited laboratories using validated methods (Table 4). The determined concentrations using the developed extraction and analytical method by GC/TOFMS were fairly consistent with the results from these laboratories, with the exception for lysine and creatinine. To some extent this might be due to the higher RSD for those compounds (Table 3). Furthermore, the accuracy of the suggested protocol for metabolomic analysis of LMC in plasma was also shown by the high correlation between the concentrations of glutamate determined using an amino acid analyzer and our GC/TOFMS method using the standard addition method with or without corresponding stable isotope-labeled internal standard method.

The detection limits of the 12 compounds were also estimated (Table 4) and found, in most cases to be in the range of 0.1−1 μM in human blood plasma. This corresponds to an absolute detection limit as low as 0.1 pmol for some of the compounds. By comparing the results when 20, 50, 100, and 200 μL of plasma were extracted in 80, 200, 400, and 800 μL of methanol, it was also shown that reproducible results could be obtained with very low volumes of plasma (data not shown). No significant differences were found among the results obtained with the different extraction volumes of plasma.

DISCUSSION

In metabolomics, the goal is to analyze as many metabolites as possible from a biological specimen. This is a considerably

Table 4. Concentrations and Detection Limits for Endogenous Compounds in Human Blood Plasma Analyzed According to the Developed Method for Global Analysis of LMC

endogenous compounds	concns, plasma pool μM	limit of detection	
		pmol	μM
proline	289.1[a,c]	1.16	0.14
glutamate	42.4[a,c]	4.24	0.53
phenylalanine	55.8[a,c]	1.12	0.14
tryptophan	55.7[a,c]	1.11	0.14
methionine	28.8[a,c]	0.58	0.07
lysine	127.3[a,c]	2.55	0.32
citrate	318.6[a]	3.19	0.40
myo-inositol	24.5[a]	0.98	0.12
creatinine	106.5[a,c]	10.65	1.33
urate	331.5[a,c]	6.63	0.83
proline	253.0[b,c]	1.01	0.13
glutamate	43.5[b,c]	4.35	0.54
palmitate	125.8[b]	2.52	0.31
cholesterol, nonesterified	2109.7[b]	42.19	5.27

[a] Concentrations calculated with standard addition method. [b] Concentration calculated by standard addition method including corresponding stable isotope labeled internal standard. [c] Concentrations of proline, glutamate, phenylalanine, tryptophan, methionine, lysine, creatinine, and urate were 261, 44, 61, 58, 29, 196, 75, and 276 μM, respectively, as determined using validated methods in accredited laboratories.

more complex task than analyzing a limited number of specific compounds. For instance, it is practically impossible to obtain high accuracy and reproducibility for all metabolites. Blood plasma contains several thousand compounds, and the number can vary with diet, gut microbial flora, physical activity, genetic background, and environmental factors. Therefore, it is not possible to add stable isotope-labeled internal standards for all detected compounds, which would increase the accuracy and reproducibility of the analysis. Furthermore, estimation of the concentration of compounds with high accuracy and reproducibility requires high extraction efficiency. This is impossible in practice for all of the metabolites, since plasma contains metabolites that differ widely in chemical nature and amounts, and they are commonly noncovalently bound to proteins. Therefore, the goal in this investigation was to develop a method that is likely to extract as many classes of compounds as possible with high efficiency and reproducibility.

To date, very little has been published on optimization of protocols for metabolomic analysis. However, O'Hagan et al.[12] recently described the use of an automated optimization strategy based on evolutionary computing for optimizing GC separations of the metabolomes of human serum and yeast. In addition, Nicholson and co-workers investigated how the analysis of plasma by NMR was affected by using different deproteinization methods.[13] In the present investigation, we used DOE to optimize both extraction and derivatization parameters for the analysis of human blood plasma. We have recently shown that by using DOE it is possible to develop a reliable protocol for metabolomic analysis of *Arabidopsis*.[14] DOE introduces variation to the system in a systematic way, allowing the effects of this variation to be analyzed using regression models. The interpretation of these models will then be reliable since there is little or no correlation between the factors varied in the design. Furthermore, by using DOE, we were able to obtain an overview of the problems associated with optimizing the extraction or derivatization protocol for metabolomic analysis of plasma, with a relatively limited number of experiments.

Global extraction of LMC can be enhanced by deproteinization instead of other conventional methods, for instance, liquid−liquid extraction or solid-phase extraction. Several methods can be used to precipitate and remove protein in plasma, such as the addition of organic solvents or salts or changes of pH using strong acids.[15−17] However, strong acids adversely affect chemically unstable compounds,[18] and too much salt in samples can impair the MS analysis. Therefore, we chose to use organic solvents to optimize the extraction protocol. In the present study we investigated mainly five common organic solvent (methanol, ethanol, acetonitrile, acetone, chloroform). However, it should be mentioned that other organic solvent might give slightly different results. The results from the present investigation suggest that methanol is the optimal solvent for metabolomic analysis of plasma (Figures 1 and 2). Methanol was used by O'Hagan et al.[12] in the analysis of human serum and has also been the choice of a published investigation of the *Escherichia coli* metabolome.[17] However, our findings that methanol was optimal for extraction and deproteinization conflict with those of Daykin et al.,[13] who recommended the use of acetonitrile for NMR analysis. This apparent discrepancy could be due to the limited set of metabolites that dominate the NMR spectra. In the present investigation, use of acetonitrile resulted in poor, and less reproducible, extraction of most metabolites (Figure 1) than other solvents (data not shown). The choice of extraction solvent always involves a compromise between the extraction efficiency for polar and nonpolar compounds. We here suggest the use 800 μL of methanol (total extraction volume 1000 μL), even though the results from the optimization of the methanol volume show that 700 and 750 μL result in somewhat higher peak areas for most compounds and that 900 μL leads to slightly higher responses of, e.g., cholesterol.

To perform rapid extractions of metabolites from plasma we used a Retsch mill, which is usually employed to grind hard materials of plants with beads to increase the efficiency.[14] However, it was found also useful for analysis of metabolite profiles in plasma as it can rapidly and efficiently homogenize plasma extracts. By applying the present parameter setting, the number of peaks detected in the present investigation was more than 500. The number detected depends heavily on the software settings, such as the signal/noise threshold and the baseline value. For example, using the Leco ChromaTof software, the number of peaks detected could be varied from 400 to 1100 by using different settings (data not shown). The recently developed H-MCR data

(12) O'Hagan, S.; Dunn, W. B.; Brown, M.; Knowles, J. D.; Kell, D. B. *Anal. Chem.* **2005**, *77*, 290−303.
(13) Daykin, C. A.; Foxall, P. J. D.; Connor, S. C.; Lindon, J. C.; Nicholson, J. K. *Anal. Chem.* **2002**, *304*, 220−230.
(14) Gullberg, J.; Jonsson, P.; Nordström, A.; Sjöström M.; Moritz, T. *Anal. Biochem.* **2004**, *331*, 283−295.
(15) Vanholder, R.; Hoefliger, N.; De Smet, R.; Ringoir, S.; Vogeleere, P. *Kidney Int.* **1992**, *41*, 1707−1712.
(16) Tyagi, R. K.; Azrad, A.; Degani, H.; Salomon, Y. *Magn. Reson. Med.* **1996**, *35*, 194−200.
(17) Maharjan, R. P.; Ferenci, T. *Anal. Biochem.* **2003**, *313*, 145−154.
(18) de Koning, W.; Van Dam, K. *Anal. Biochem.* **1992**, *204*, 118−123.

processing method was used for the data processing instead of the GC/MS software.[9] One of its main features is that chromatographic data are simultaneously processed and aligned, resulting in higher proportions of correctly detected peaks and reductions in the calculation time. All 500 resolved peak areas, including a subset of identified peaks, were used as response variables for the solvent and derivatization designs.

By adding stable isotope-labeled internal standards, it is possible to increase the accuracy and reproducibility of GC/MS analyses. The 10 internal standards chosen here represent different classes of compounds, e.g., amines, amino acids, fatty acids, sterols, and disaccharides. With this strategy we can determine accurate levels for quite large numbers of compounds. However, it should be emphasized that the differences in relative levels between groups are generally of most interest in metabolomic studies.[5,19] Furthermore, by adding a large number of internal standards, any differences in recovery between different classes of compounds and samples should be detected. This is a simple quality check of the analysis, as the peak areas for the different internal standards should not vary between the various samples in the same kind of body fluid. In the present investigation, the precision of the determinations of the 32 chosen metabolites ranged from 2 to 29%. This is similar to the precision achieved in other metabolomic approaches[14,20] and is in our opinion acceptable. Earlier studies have shown that the variation due to chromatographic and mass spectrometry conditions is ~3.5%.[14] Although the method errors could be decreased by improving the chromatographic system and peak area integration, most errors originate from pipetting and, to some extent, nonreproducible variations in homogenization/extraction/derivatization. However, implementing measures to improve the extraction efficiency, e.g., further extractions with other solvents, would not necessarily be advantageous since they would also increase the cycle times, and ease and rapidity are key concerns in metabolomic analysis.

(19) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155−171.
(20) Fiehn, O.; Kopka, J,; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. *Nat. Biotechnol.* **2000**, *18*, 1157−1161.

CONCLUSIONS

From the results of the designs, an extraction method was developed that can be briefly described as follows. First, 100 μL of blood plasma is extracted with a 900-μL mixture of methanol and water (8:1 v/v) containing all the internal standards, followed by centrifugation. A 200-μL aliquot of the supernatant is then transferred to a GC/MS vial and evaporated to dryness. Prior to GC/MS analysis, the samples are methoxymated at room temperature for 16 h (with 30 μL of 15 mg mL^{-1} methoxyamine in pyridine) and trimethylsilylated with 30 μL of MSTFA for 1 h. The method is rapid, reproducible, and results in low detection limits when combined with GC/TOFMS analysis. We believe that the method can also be used for LC/MS analysis and metabolomic/metabonomic studies for identifying biological markers related to diseases. Since small plasma volumes can be used, it could also be useful for studies involving small experimental animals such as mice.

ACKNOWLEDGMENT

This study was financially supported by the Swedish Research Council, Wallenberg Consortium North (WCN), the Kempe Foundation, EU strategic funding, and Strategic Research Funding (SSF). We thank Krister Lundgren for his help with the GC/TOFMS analyzes.

SUPPORTING INFORMATION AVAILABLE

Identified compounds in human blood plasma. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review July 8, 2005. Accepted October 7, 2005.

AC051211V