Development of a Quantitative Metabolomic Approach to Study Clinical Human Fecal Water Metabolome Based on Trimethylsilylation Derivatization and GC/MS Analysis

ARTICLE *in* ANALYTICAL CHEMISTRY · AUGUST 2010

Impact Factor: 5.64 · DOI: 10.1021/ac1006552 · Source: PubMed

3 AUTHORS:

Xianfu Gao
Chinese Academy of Sciences
15 PUBLICATIONS **301** CITATIONS
SEE PROFILE

Estelle Pujos-Guillot
French National Institute for Agricultural R…
65 PUBLICATIONS **1,123** CITATIONS
SEE PROFILE

jean-louis Sebedio
Human Nutrition Unit
254 PUBLICATIONS **5,099** CITATIONS
SEE PROFILE

Anal. Chem. **2010,** *82,* 6447–6456

Development of a Quantitative Metabolomic Approach to Study Clinical Human Fecal Water Metabolome Based on Trimethylsilylation Derivatization and GC/MS Analysis

Xianfu Gao,[†] Estelle Pujos-Guillot,[‡] and Jean-Louis Sébédio*[,†]

INRA, UMR 1019, Plateforme d'Exploration du Métabolisme, Nutrition Humaine, F-63122, Saint Genès Champanelle, France, and Clermont Université, UFR Médecine, UMR 1019 Nutrition Humaine, F-63000, Clermont-Ferrand, France

Metabolomic analysis of human fecal water recently aroused increasing attention with the importance of fecal metabolome in exploring the relationships between symbiotic gut microflora and human health. In this study, we developed a quantitative metabolomic method for human fecal water based on trimethylsilylation derivatization and GC/MS analysis. Methanol was found to be the best solvent for protein precipitation and extraction of fecal water metabolome. Within the optimized linear range of sampling volume (less than 50 μL), compounds showed a good linearity with a correlation coefficient higher than 0.99. The developed method showed good repeatability for both sample preparation and GC/MS analysis with the relative standard deviations lower than 10% for most compounds and less than 20% for a few other ones. The method was further validated by studying analytical variability using a set of clinical samples as well as a pooled sample. The pH value and matrix effects were the main factors affecting the accuracy of quantitative calibration curves. The increased pH value decreased the loss of short chain fatty acids during lyophilization. Spiking fecal water to a standard mixture significantly enhanced the accuracy of quantitative calibration curves, probably due to the inhibition of volatile loss during lyophilization and the increase of compound solubility in the derivatization medium. A strategy for calibration curve preparation was proposed in order to avoid the effects of pH and matrix. Totally, 133 compounds were structurally confirmed from a set of clinical samples, and 33 of them were quantified, which demonstrates the suitability of this method for a quantitative metabolomic study of human fecal water samples.

Metabolomics aims to simultaneously measure as many small molecules as possible present in a given biological system, such as body fluid, cell, tissue, or feces samples, in order to acquire comprehensive insights on the functioning of a biological system.[1,2] It is currently considered as an effective mean for studying significant differences in metabolic profiles and screening potential biomarkers by combining an untargeted robust instrumental analysis with multivariate statistics. Recent development of metagenomics suggests that a human is a complicated "superorganism" due to the combination with numerous gastrointestinal microorganisms, which closely affect nutrient absorption, energy regulation, detoxification, or transformation of xenobiotics and the health of the gastrointestinal tract.[3−5] On the basis of the perspective of system biology, the test materials of metabolomics were spread from urine and plasma to saliva, tissue, and feces in order to obtain comprehensive metabolic information. Currently, metabolomic analysis of feces samples, especially the aqueous phase, fecal water, is receiving more attention,[6,7] because fecal metabolic compositions and variations not only reflect the status of the gut microbiome but also bridge the relationships between symbiotic microbes and the host's (human) health. The discovery of biomarkers and the characterization of metabolic phenotypes of fecal water will provide the potentiality for effective health promotion, disease prevention, and even personalized nutrition and healthcare.

A robust high-throughput and large-scale sample analysis is crucial to the outcomes of metabolomics in such a field. Mass spectrometry-based techniques (MS) and nuclear magnetic resonance spectroscopy (NMR) are developed as the major analytical tools for metabolomics. NMR is a global and nondestructive technique only requiring minimal sample pretreatment, and its application in disease metabolomic research was recently described for fecal water.[6−9] MS coupled with advanced chromato-

* To whom correspondence should be addressed. E-mail: jls@clermont.inra.fr. Phone: +33 473624677.

[†] INRA.

[‡] Clermont Université.

(1) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155–171.

(2) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181–1189.

(3) Li, M.; Wang, B. H.; Zhang, M. H.; Rantalainen, M.; Wang, S. Y.; Zhou, H. K.; Zhang, Y.; Shen, J.; Pang, X. Y.; Zhang, M. L.; Wei, H.; Chen, Y.; Lu, H. F.; Zuo, J.; Su, M. M.; Qiu, Y. P.; Jia, W.; Xiao, C. N.; Smith, L. M.; Yang, S. L.; Holmes, E.; Tang, H. R.; Zhao, G. P.; Nicholson, J. K.; Li, L. J.; Zhao, L. P. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2117–2122.

(4) Goodacre, R. *J. Nutr.* **2007**, *137*, 259S–266S.

(5) Jia, W.; Li, H. K.; Zhao, L. P.; Nicholson, J. K. *Nat. Rev. Drug Discovery* **2008**, *7*, 123–129.

(6) Jansson, J.; Willing, B.; Lucio, M.; Fekete, A.; Dicksved, J.; Halfvarson, J.; Tysk, C.; Schmitt-Kopplin, P. *PLoS One* **2009**, *4*, e6386.

(7) Monleon, D.; Morales, J. M.; Barrasa, A.; Lopez, J. A.; Vazquez, C.; Celda, B. *NMR Biomed.* **2009**, *22*, 342–348.

(8) Jacobs, D. M.; Deltimple, N.; van Velzen, E.; van Dorsten, F. A.; Bingham, M.; Vaughan, E. E.; van Duynhoven, J. *NMR Biomed.* **2008**, *21*, 615–626.

(9) Marchesi, J. R.; Holmes, E.; Khan, F.; Kochhar, S.; Scanlan, P.; Shanahan, F.; Wilson, I. D.; Wang, Y. L. *J. Proteome Res.* **2007**, *6*, 546–551.

10.1021/ac1006552
Published on Web 07/14/2010

graphic separation instruments, such as liquid chromatography (LC) and gas chromatography (GC), has become a powerful metabolomic tool, with a wide dynamic range and reproducible quantitative capabilities, for multicomponent measurement in biological samples.[10−12] GC/MS is a robust method for quantification with more sensitivity and resolution than the commonly used NMR approach and better reliability in structure identification of candidate biomarkers than LC/MS. It has been applied in metabolomic research of urinary, plasma, and tissue samples.[13−16]

Our previous work developed a robust specific metabolomic methodology for a fecal water sample based on the derivatization with ethyl chloroformate.[17] However, some potential biomarkers such as carbohydrates, bile acids, polyols, and hydroxyl acids could not be derivatized. Trimethylsilylation (TMS) is a widely applied derivatization method for providing highly polar compounds a sufficient volatility and narrowing the boiling point window so that they can be eluted at suitable temperatures without thermal decomposition and molecular rearrangement, as well as increasing ionization efficiency. Although TMS compounds are very effective and commonly used derivatization agents, to date, no methodological literature about fecal water metabolomics based on such derivatization has been published. To the best of our knowledge, current methodological publications almost omitted relatively comprehensive recoveries and quantification, especially for those compounds with volatility or low solubility in aqueous phase. In the current study, we developed a quantitative metabolomic methodology for human fecal water based on GC/MS and trimethylsilylation derivatization. The factors affecting relative and absolute quantification, such as the linear range of sampling volume, and the influences of pH and matrix of a standard mixture on the preparation of calibration curves were studied. The method was validated using a pooled fecal water sample, and 133 metabolites were structurally identified from a set of clinical fecal water samples.

EXPERIMENTAL SECTION

Chemicals and Materials. *N*,*O*-bis(Trimethylsilyl)trifluoroacetamide (BSTFA) with 1% trimethylchlorosilane (TMCS) and methoxylamine were obtained from Supelco (Bellefonte, PA). [$^{13}C_1$]-L-valine, a stable isotope-labeled internal standard, was obtained from Cambridge Isotope Laboratories (Andover, MA). All standard compounds were commercially purchased from Sigma-Aldrich, and solutions in methanol of HPLC grade or ultrapure water (Milli-Q system, Millipore, Billerica, MA) were prepared. The mixed stock solutions of methanol-soluble or water-soluble standards were prepared, respectively, by combining the two standard solutions, and their pH values were further regulated to 7. Totally, 33 representative compounds from each chemical class were chosen and mixed to produce a stock standard mixture solution (500 μg/mL).

Sixty-five human fecal water samples were obtained from healthy volunteers during the METAPROFILE human study. Written informed consent was obtained from all participants. The study protocol was approved by the Comité de Protection des Personnes Sud-Est IV (no. AU703).

Samples were prepared as follows: the homogenized stool samples (5−8 g) were ultracentrifuged at 4 °C at 50 000 rpm (171 500*g*, Beckman Ti 70.1 rotor) for 2 h, and 2 μL of NaN_3 (100 mg/mL) as an antimicrobial agent per gram of fecal water was added to the supernatants, which were stored at −80 °C as 1 mL fecal water aliquots prior to derivatization.[17]

Sample Preprocessing and Derivatization. Fecal water samples were thawed and short-time vortexed at room temperature. A 2-fold volume of ice-cold methanol was added into fecal water samples. The mixture was vortexed strongly and subsequently centrifuged at 14 000 rpm at 4 °C for 10 min. A quality control (QC) sample was prepared by pooling 65 human fecal water samples according to the proposition of Sangster et al.[18] A 60 μL supernatant (equivalent to 20 μL of fecal water) and 10 μL of [$^{13}C_1$]-L-valine (100 μg/mL) were freeze-dried, and subsequently, 80 μL of methoxylamine solution (15 mg/mL in pyridine) was added to each vial. The resultant mixture was vigorously vortex-mixed for 1 min and reacted at 37 °C for 24 h in order to inhibit the cyclization of reducing sugars and the decarboxylation of α-keto acids. Eighty microliters of BSTFA (with 1% TMCS) were added into the mixture and derivatized at 70 °C for 60 min, and a 2-fold volume of heptane was added into derivatives prior to injection. At the same time, a control derivatization sample (fecal water sample was substituted by deionized water) was prepared in order to remove the background noise produced during sample preprocessing, derivatization, and GC/MS analysis.

GC/MS Analysis. Analysis was performed on an Agilent 7890A gas chromatography system coupled to an Agilent 5975C inert XL EI/CI MSD system (Agilent, CA). A HP-5MS fused-silica capillary column (30 m × 250 μm i.d.; Agilent J & W Scientific, Folsom, CA), chemically bonded with a 5% phenyl−95% methylpolysiloxane cross-linked stationary phase (0.25 μm film thickness), was utilized to separate the derivatives. Helium was used as a carrier gas at a constant flow rate of 1 mL/min through the column. One microliter of derivatives was injected, and the solvent delay time was set to 5.5 min. The initial oven temperature was held at 60 °C for 2 min, ramped to 140 °C at a rate of 10 °C/min, to 240 °C at a rate of 4 °C/min, to 300 °C at a rate of 10 °C/min, and finally held at 300 °C for 8 min. The initial inlet gas pressure was 8.2317 psi. The temperatures of injector, transfer line, and electron impact (EI) ion source were set to 250 °C, 290 °C, and 230 °C, respectively. The electron energy was 70 eV, and mass data was collected in a full scan mode (m/z 50−600). Agilent "retention time locking" (RTL) was applied to control the reproducibility of retention times (RT). Even if RTL still produces small RT drifts, especially at the region of low column temperature

(10) Want, E. J.; Nordstrom, A.; Morita, H.; Siuzdak, G. *J. Proteome Res.* **2007**, *6*, 459–468.
(11) Pereira, H.; Martin, J.-F.; Joly, C.; Sébédio, J.-L.; Pujos-Guillot, E. *Metabolomics* **2010**, *6*, 207–218.
(12) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrom. Rev.* **2007**, *26*, 51–78.
(13) Jlye, A.; Trygg, J.; Gullberg, J.; Johansson, A. I.; Jonsson, P.; Antti, H.; Marklund, S. L.; Moritz, T. *Anal. Chem.* **2005**, *77*, 8086–8094.
(14) Lin, H. M.; Edmunds, S. J.; Helsby, N. A.; Ferguson, L. R.; Rowan, D. D. *J. Proteome Res.* **2009**, *8*, 2045–2057.
(15) Ni, Y.; Su, M. M.; Lin, J. C.; Wang, X. Y.; Qiu, Y. P.; Zhao, A. H.; Chen, T. L.; Jia, W. *FEBS Lett.* **2008**, *582*, 2627–2636.
(16) Mal, M.; Koh, P. K.; Cheah, P. Y.; Chan, E. C. Y. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 487–494.
(17) Gao, X.; Pujos-Guillot, E.; Martin, J.-F.; Galan, P.; Juste, C.; Jia, W.; Sebedio, J.-L. *Anal. Biochem.* **2009**, *393*, 163–175.
(18) Sangster, T.; Major, H.; Plumb, R.; Wilson, A. J.; Wilson, I. D. *Analyst* **2006**, *131*, 1075–1078.

or for the compounds with a bad peak shape, RTL did improve peak alignment by increasing the reproducibility of RT.

To obtain reliable molecular weight information, chemical ionization (CI) mass spectra were obtained with the same capillary column and GC parameters. Pure methane was used as reagent gas.

Linearity Range and Method Validation. The supernatant of QC sample extracted by a 2-fold volume of methanol was used to evaluate the linearity range and validate the developed method. Such supernatant was serially diluted 2, 5, 10, 20, 50, 100, 200, and 500 times, individually, with a mixture of methanol−water (2:1). The linearity range was determined using a correlation coefficient ($R > 0.99$) of plotting logarithmic signal integral area against logarithmic compound concentration. On the basis of the above linearity range, the aliquots of 30 μL of QC sample were used to determine repeatability of the analytical equipment and sample preparation. The recoveries were determined by adding 20, 30, and 40 μL of NaOH regulated standard mixture (50 μg/mL in 50% methanol, pH = 7) into the aliquot of QC sample (30 μL), respectively.

Influences of pH Regulation and Matrix on Calibration Curves. Solutions of sodium hydroxide (NaOH) or ammonia hydroxide (NH_4OH), respectively, were applied to regulate the pH of a standard mixture (identical to fecal water) so as to evaluate the impacts of matrix of a standard mixture on sample preparation and GC/MS analysis. The quantitative calibration curves were obtained by analyzing the derivatives of three following media: SM, standard mixture (without pH regulation); NSM, NaOH regulated standard mixture alone (pH = 7); NSMF, NSM spiked with fecal water (QC sample). For SM and NSM groups, the standard stocks were diluted 2, 5, 10, 20, 50, 100, 200, and 500 times with methanol−water (2:1). The diluted standard solution and the same volume of methanol−water (2:1) were added to the vial for sample preparation and analysis so as to determine the influence of pH regulation on calibration curves. To evaluate matrix effects, the same volume of the diluted NSM solution and the aliquot of QC supernatant (see the section "Sample Preprocessing and Derivatization") were added to the vial for sample preparation and analysis (NSMF). The same volume of QC supernatant alone was prepared and analyzed as control as well. All samples were treated under the same conditions.

Data Extraction and Preprocessing. Many tools for GC/MS data processing have been developed.[19−21] The conventional approach of mass spectral deconvolution is based on information present within a single chromatogram. In order to perform comparative data analysis in profiling experiments, alternative softwares, based on the extraction of mass selective peak apex intensities, have been published.[19,22] TagFinder04 software[22] allows aligning large GC/MS profiling data and automated data extracting using retention time groups of mass fragments and clusters of intensity correlated mass fragments within a small retention time window such as 10 s.

Raw GC/MS data files were transformed to nominal CDF files in MetAlign software.[23] An XCMS software package[19] and in-house scripts were applied to extract data from each nominal CDF file one by one. The resulting data of each sample was exported to a single text file. All data with time window 330−2940 s and mass window 50−600 m/z, stored in a separate file, were imported to TagFinder04 software, where retention time correction, peak alignment, mass tag correlation, clustering, and grouping were performed. The tag output was further processed in Microsoft Excel (Microsoft, Redmond, WA). Referring to the data of the control derivatization sample, all artificial noise signals produced during the processes of solvent extraction, derivatization, and GC/MS analysis were excluded. Only the information from base peak ion of each cluster (one compound) was kept as quantifier. The resulting data consisting of arbitrary peak indexes (RT-m/z pair), sample names (observations), and peak areas (variables) were exported for multivariate statistical analysis. The resolved tag spectra were exported as a "msp" format file, in order to automatically search compound information from the NIST library (http://www.nist.gov/srd/mslist.htm) using NIST MS Search 2.0 software. In contrast to XCMS software package and MZmine software,[24] TagFinder can cluster the mass fragments belonging to a compound, discard unrelated noise signals, and simultaneously decrease peak loss. For linear volume range, method validation, and calibration curves based on the QC sample, data preprocessing was performed as previously described.[17]

Compound Identification in Human Fecal Water. The "msp" format file comprising deconvoluted tag spectra was introduced to the NIST MS Search 2.0 software for automatically searching compound information from the NIST05 library and the author-constructed standard library. The searching results with match similarity larger than 80% will be accepted as candidate compounds.

However, TagFinder still produces potential risk in resolving the overlapped peaks of trimethylsilyl derivatives. The mass fragments coexisting in the overlapped peaks are usually summed in intensity, which results in low correlation with all other ions. Such mass fragments will finally be discarded from any resolved mass spectra, and consequently, a low searching probability with a high matching value will occur during the library search. Thus, it is essential to qualitatively validate these overlapped compounds which were resolved by TagFinder software. Consequently, a raw GC/MS data file of QC sample was imported to the Automated Mass Spectral Deconvolution and Identification System (AMDIS) for automatically searching against an author-constructed AMDIS standard library. The results obtained from the above two methods were confirmed as reliable candidate compounds. The molecular weights of candidate compounds obtained by searching into the library were further validated by checking CI mass spectra. The retention times of reference standards were utilized to confirm the candidate compounds. The retention indexes of remaining candidate compounds were calculated and compared with a NIST

(19) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779–787.
(20) Dixon, S. J.; Brereton, R. G.; Soini, H. A.; Novotny, M. V.; Penn, D. J. *J. Chemom.* **2006**, *20*, 325–340.
(21) Jonsson, P.; Johansson, E. S.; Wuolikainen, A.; Lindberg, J.; Schuppe-Koistinen, I.; Kusano, M.; Sjostrom, M.; Trygg, J.; Moritz, T.; Antti, H. *J. Proteome Res.* **2006**, *5*, 1407–1414.
(22) Luedemann, A.; Strassburg, K.; Erban, A.; Kopka, J. *Bioinformatics* **2008**, *24*, 732–737.
(23) Lommen, A. *Anal. Chem.* **2009**, *81*, 3079–3086.
(24) Katajamaa, M.; Miettinen, J.; Oresic, M. *Bioinformatics* **2006**, *22*, 634–636.

RI database and Internet available database such as NIST Chemistry WebBook.[25]

Multivariate Statistical Analysis. The peak table (named matrix X) obtained from the outcomes of TagFinder software (Section "Data Extraction and Preprocessing") was imported to SIMCA-P 11.0 software package (Umetrics, Umeå, Sweden), where principal component analysis (PCA) was performed.[13] All data was mean-centered and Pareto-scaled[26] prior to PCA analysis. The scores plot of PCA was used to observe clustering and trends among all samples treated by methanol, acetonitrile, their combination with chloroform, and water (control group) or between 65 clinical samples and 9 QC samples. The loadings plot was applied to evaluate the effects of different solvents on the recoveries of fecal water metabolome. R^2X represents the cumulative modeled variation in X, where the value approaching 1.0 suggests that the current model is stable and reliable. Q^2 is an estimate of the predictive ability of the model, calculated by cross-validation.[13] Cross-validation with seven cross validation groups was used throughout to determine the optimal number of principal components. R^2X and Q^2, which reveal the fraction of the total variation predicted by a principal component, were used as indicatives to assess the robustness of a pattern recognition model.

RESULTS AND DISCUSSION

Derivatization and GC/MS Analysis. The GC/MS parameters were optimized. In contrast to our previous method,[17] two main short chain fatty acids in fecal water, 2-methylbutyric acid and 3-methylbutyric acid, were strongly overlapped chromatographically by the peak of methoxylamine, regardless of how to optimize chromatographic conditions. The Agilent "retention time locking" (RTL) technique was applied in all tests to keep the stability of retention times across all samples, which helps improve the accuracy of peak alignment during data preprocessing. We found that analytical variation generally occurred in the first two runs; thus, it is recommended to perform two conditioning runs prior to formal runs.

The chemical derivatization of fecal water samples prior to GC/MS analysis is an essential step for decreasing boiling points of highly polar compounds, narrowing boiling point window, and increasing ionization of derivatives. The direct derivatization of reducing sugars generally results in the formation of side products, because the cyclic and open chain structures of reducing sugars are simultaneously present, whatever the derivatization conditions.[27] To prevent the formation of side derivatives and facilitate data processing after instrumental analysis, methoximation was performed by incubation with methoxylamine prior to derivatization. Figure 1 demonstrates that the number of derivatives of glucose, a typical reducing sugar, was significantly decreased by the increased time of methoximation. To effectively decrease the effects of multiple peaks of reducing sugars on peak finding, methoxymation at 37 °C for 24 h was chosen here. Derivatization conditions of trimethylsilylation were also investigated. Five aliquots of QC sample were derivatized at 70 °C for 30, 60, 90, 120, and 150 min, respectively. Derivatization efficiency was better at 60 min, except for single phenolic acids, but was found to be inhibited by a prolonged derivatization time. The compromised conditions for derivatization were 70 °C for 60 min.

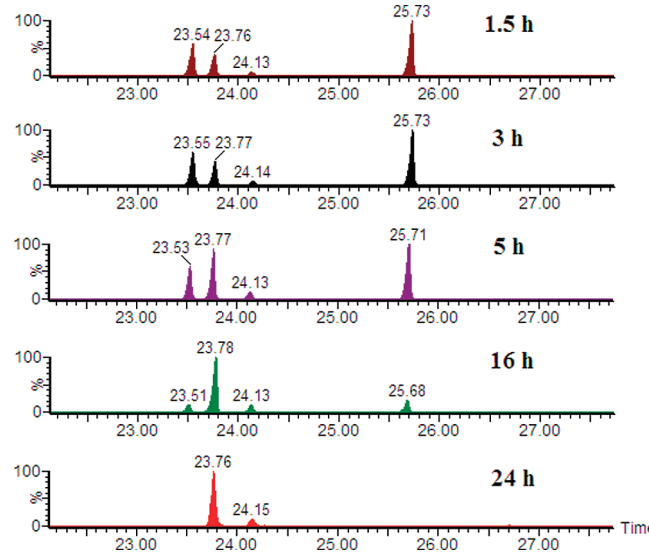

Figure 1. Effect of methoximation time on the derivatization of glucose. Methoximation was carried out at 37 °C for 1.5, 3, 5, 16, and 24 h, respectively.

Influence of Protein Precipitation on Human Fecal Water Metabolome. Human feces is a mixture of digested products, residues, and symbiotic gut microflora and their metabolites from the intestinal tract. Fecal water, as the aqueous portion of fecal constituents, is normally prepared by ultracentrifuging the fresh or freeze−thawed feces, but this method cannot remove all proteins from fecal water. There is a general consensus that sample preprocessing, such as protein precipitation, is very important for improving recoveries of small molecular metabolites in biofluids.[13] In this work, different solvents were tested. The aliquots of QC sample were treated, respectively, by a 2-fold volume of methanol, acetonitrile, methanol−chloroform (3:1), acetonitrile−chloroform (3:1), and water as control. It seems that acetonitrile exhibited a better outcome for protein precipitation as it produced more transparent supernatants than methanol. No obvious differences between acetonitrile group and acetonitrile−chloroform group were observed (data not shown). This was also found between methanol group and methanol−chloroform group. However, to ensure that protein precipitation would not result in important loss of metabolites, the recoveries of human fecal water metabolome had to be studied.

A PCA model ($R^2X = 0.912$, $Q^2 = 0.899$, two principal components) was constructed to evaluate the influence of different extraction solvents on these recoveries. No normalization to the sum of extraction ion intensities per sample was done because the different solvents may significantly affect fecal water metabolome, probably leading to a false conclusion. As a result, a two-component PCA model was generated and the scores plot is shown in Figure 2A. The separation in PC1 (Figure 2A) indicates that acetonitrile influenced to a greater extent the recovery of human fecal water, which was further strengthened by replacing 25% acetonitrile with chloroform. The loadings plot indicates what markers are important, that is, the samples located

(25) Linstrom, P. J.; Mallard, W. G. *J. Chem. Eng. Data* **2001**, *46*, 1059–1063.
(26) van den Berg, R. A.; Hoefsloot, H. C. J.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. *BMC Genomics* **2006**, *7*, 15.
(27) Asres, D. D.; Perreault, H. *Can. J. Chem.-Rev. Can. Chim.* **1997**, *75*, 1385–1392.

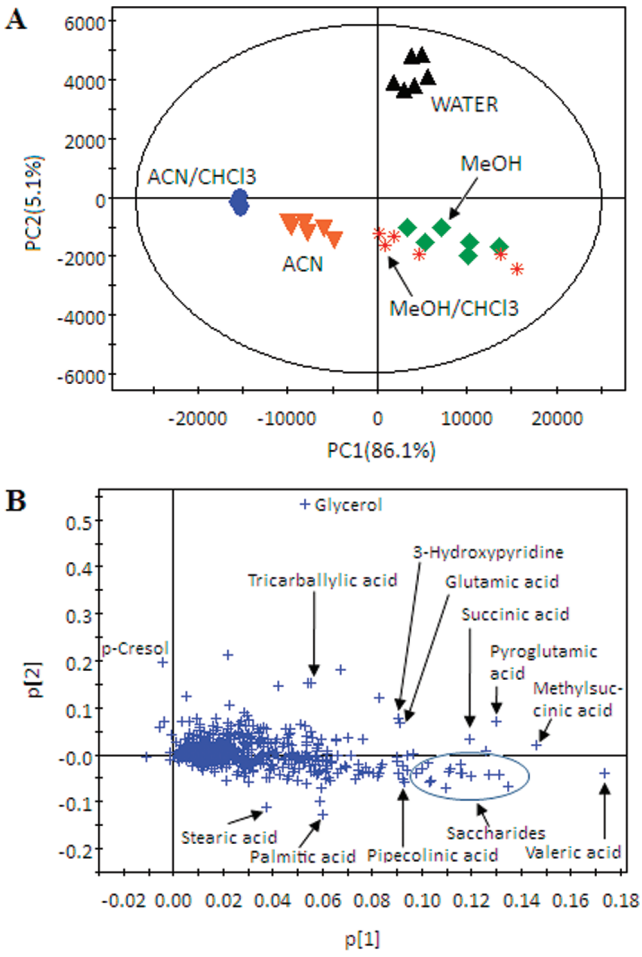

Figure 2. Influence of extraction solvents on the human fecal water metabolome. The aliquots of QC sample were treated by 2-fold volume of methanol (MeOH), acetonitrile (ACN), methanol−chloroform (3:1, MeOH/$CHCl_3$), acetonitrile−chloroform (3:1, ACN/$CHCl_3$), and deionized water (WATER) as control, respectively. PCA was carried out using raw data (no normalization).

in the corresponding region of a scores plot are enriched with the markers from the corresponding location in the loadings plot (Figure 2B).[28] A careful analysis of Figure 2B shows that methanol alone or methanol−chloroform, compared with acetonitrile and water dilution, significantly improved the extraction of almost all compounds from fecal water, in particular carboxylic acids and saccharides. It is worth noting that no obvious profile separation was found between methanol group and methanol−chloroform group. In summary, methanol alone is a simple optimal solvent for protein precipitation and maximizing peak areas of fecal water metabolome.

Qualitative Analysis of Human Fecal Water Metabolome after Derivatization. The above developed method was applied to a set of clinical human fecal water samples (65 in total) so as to qualitatively identify the compounds in the human fecal water metabolome. In this work, two deconvolution methods of mass spectra data were attempted to detect coeluting peaks (Figure 3).

TagFinder software is a free software for retrieving mass spectral tags based on the intensity correlation of mass tags within a time group across samples.[22] The accuracy of exported deconvoluted mass data is closely correlated with sample scale and variability among samples; thus, this software is especially suitable for metabolomics. After exclusion of artifact peaks, 558 chromatographic peaks were resolved from 80% of the clinical human fecal water samples. The resolved mass tags were introduced to NIST MS Search 2.0 for automatically matching NIST05 library. Since many TMS derivatives produce typical high abundance mass fragments, such as $m/z = 73$, 75, or 147, these fragments of the overlapped peaks are normally summed during peak finding and integration and, consequently, hard to be correctly clustered in TagFinder software. As a result, they will be discarded, which causes low matching probability or errors in searching against the NIST05 library.

To qualitatively validate the deconvoluted compounds, especially the overlapped compounds resolved by TagFinder software, AMDIS, which is common free software specially designed for qualitative analysis, was thereafter utilized to extract compound information directly from raw data files. To avoid possible false positives produced with AMDIS,[29] retention times were simultaneously compared with those of reference standards.

The results from NIST05 (based on TagFinder04) and AMDIS (based on raw data file) were combined in order to obtain reliable compound information (Figure 3). Totally, 133 compounds were structurally confirmed on the basis of searching an author-constructed standards library, including mass spectra and retention times, and commercially available mass spectra libraries such as NIST05. The molecular weights of these compounds were confirmed by CI mass spectra in the same GC conditions (data not shown). Table 1 shows a wide range of compounds including carboxylic acids, hydroxyl acids, carbohydrates, polyols (sugar alcohols), phenolics, bile acids, amino acids, amines, and other nitrogen-containing compounds generally found in human fecal water. Although NIST05 library provides mass spectra of 4345 TMS derivatives, the majority of compounds still cannot be qualified, partly due to lacking sufficient databases or chromatographic overlapping of compounds with similar mass fragments. While our previous publication qualitatively identified 73 compounds, among them amino acids, phenolics, and carboxylic acids, in human fecal water using ethyl chloroformate derivatization,[17] the current method allows to derivatize more kinds of compounds, such as carbohydrates, polyols, bile acids, and other nitrogen-containing heterocyclic compounds. This method has better derivatization efficiency, lower sample volume requirement, and better chromatography separation on hydroxyl acids than ethyl chloroformate derivatization.

Relative Quantification and Method Validation. In metabolomics, relative quantification based on mass intensities is usually applied to construct multivariate statistical models. The concentrations of many compounds fluctuate significantly among individuals even over 2 to 3 orders of magnitude. Sometimes, some compounds can be out of the linearity range because of the excessive concentration of fecal water, which consequently impact the outcomes of the multivariate statistical model to a greater extent. To ensure that all compounds are within the linearity range of detection, the QC sample from 65 pooled fecal water samples was applied to optimize the sampling volume. Table 2 shows that the majority of compounds exhibited a correlation coefficient higher

(28) Trygg, J.; Holmes, E.; Lundstedt, T. *J. Proteome Res.* **2007**, *6*, 469–479.

(29) Lu, H. M.; Dunn, W. B.; Shen, H. L.; Kell, D. B.; Liang, Y. Z. *TrAC, Trends Anal. Chem.* **2008**, *27*, 215–227.

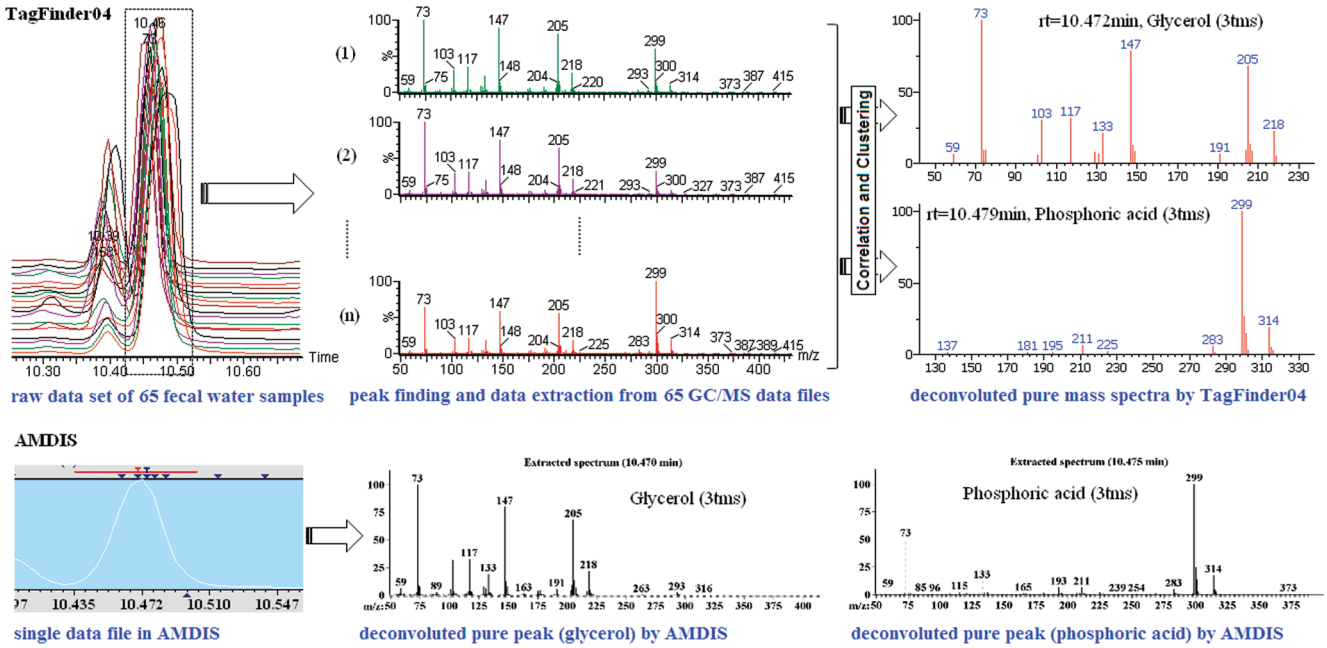

Figure 3. Data processing workflow. TagFinder04: The raw GC/MS data files were transformed to nominal CDF files and subsequently extracted and arranged to single tab files. All single tab files were imported to TagFinder04 software, where mass tags were deconvoluted according to signal intensity correlation across samples within time group, and the "msp" file consisting of mass tags was introduced to the NIST MS Search 2.0 for matching libraries. AMDIS: A raw GC/MS data file was deconvoluted by this software, and the resulting mass spectra were searched against the author-constructed AMDIS library automatically. The cross-validated compounds between TagFinder04 and AMDIS were confirmed as the reliable compounds.

than 0.99 at a wide linear range of sample volume. However, some amino acids, particularly alanine and pyroglutamic acid, showed a lower linearity range of sample volume. This study found that amino acids have lower responses and higher absolute concentrations than other compounds in fecal water samples. Too high concentrations of analytes could cause column overloading and mass detector oversaturation. Therefore, sample volume should be controlled within an appropriate range such as within 15−50 μL or less than 50 μL, and in this work, the preferable volume was found to be 20 μL for improving the reliability of relative quantification.

Method validation should be performed using a biological sample with the purpose of avoiding matrix effects.[30] These effects are one of the greatest problems in the quantification of metabolites.[31,32] As reported in Table 2, most compounds exhibited a good repeatability with relative standard deviations (RSDs) lower than 10%, except for a few compounds, especially amino acids, which showed an acceptable repeatability with RSDs between 10% and 15% (Table 2). It is worth noting that alanine and pyroglutamic acid showed bad repeatability. Unlike other aromatic compounds, *p*-cresol exhibited an excellent repeatability of GC/MS analysis but a poor repeatability of sample preparation, because *p*-cresol is prone to evaporate even in the stable fecal water matrix during the process of lyophilization. To date, there are still no generally accepted assessment criteria for repeatability of metabolomic analysis. However, as many publications and the FDA recommend single analyte tests in bioanalytical methods validation, a coefficient of variation of 15% of the nominal value is accepted as a tolerance limit, except for compounds with concentrations close to the limit of quantification (LOQ) where 20% is acceptable.[33,34]

In metabolomics, a large series of samples is generally prepared and instrumentally analyzed in batches over a long time, and therefore, analytical variability during the batch analysis should be considered. As described by Gika et al.,[35] the use of quality control (QC) samples, prepared from the study samples, enabled the performance of the system to be effectively monitored for the drift of important analytical parameters such as detector responses. Figure 4 shows the scores plot of PCA (normalized and Pareto-scaled data) of 65 human fecal water samples and 9 QC samples (the pooled sample from 65 human fecal water samples). All samples were prepared and analyzed in 3 batches (3 QC samples per batch). As can be seen, the QC samples (triangles) are tightly clustered in the middle of the plot. It is known that the closer the QC samples locate on the scores plot the more reproducible the performance of the method should be. This figure illustrates that the variability within the QCs was significantly lower than the variability among the individual fecal water samples. Thus, the variability in analytical system did not hide the intrinsic biological variability. Current method validation showed a good linearity and repeatability of sample preparation and GC/MS analysis and was, therefore, suitable for relative quantitative metabolomics study of human fecal water samples.

(30) Taylor, P. J. *Clin. Biochem.* **2005**, *38*, 328–334.
(31) Dunn, W. B.; Ellis, D. I. *TrAC, Trends Anal. Chem.* **2005**, *24*, 285–294.
(32) Ewald, J. C.; Heux, S.; Zamboni, N. *Anal. Chem.* **2009**, *81*, 3623–3629.
(33) FDA. *Guidance for industry, Bioanalytical method validation*; Food and drug administration, Center for Drug Evaluation and research (CDER): Rockville, MD, 2001.
(34) Viswanathan, C. T.; Bansal, S.; Booth, B.; DeStefano, A. J.; Rose, M. J.; Sailstad, J.; Shah, V. P.; Skelly, J. P.; Swann, P. G.; Weiner, R. *AAPS J.* **2007**, *9*, E30−E42.
(35) Gika, H. G.; Macpherson, E.; Theodoridis, G. A.; Wilson, I. D. *J. Chromatogr., B* **2008**, *871*, 299–305.

Table 1. Compounds Structurally Identified from 65 Human Fecal Water Samples by GC/MS

no.	compounds	RT	main fragments (m/z)
	Amino Acids		
1	L-alanine[a]	7.91	116; 73; 147; 190; 59
2	α-aminobutyric acid[a]	8.94	130; 73; 147; 204; 59
3	L-valine[a]	9.59	144; 73; 218; 147; 100
4	L-leucine[a]	10.40	158; 73; 147; 232; 59
5	L-isoleucine[a]	10.74	158; 73; 218; 147; 232
6	L-proline[a]	10.78	142; 73; 216; 147; 59
7	glycine[a]	10.93	174; 147; 86; 248; 100
8	L-serine[a]	11.84	204; 73; 218; 147; 100
9	L-threonine[a]	12.34	73; 218; 117; 291; 147
10	β-alanine[a]	12.97	174; 73; 248; 147; 290
11	homoserine[a]	13.44	218; 73; 103; 128; 147
12	β-aminoisobutyric acid[a]	13.59	174;73;248;304;147
13	sarcosine	13.64	116; 73; 147; 190; 218
14	L-aspartic acid[a]	13.91	232; 73; 218; 100; 147
15	L-methionine[a]	14.75	176; 73; 61; 128; 147
16	L-pyroglutamic acid[a]	14.86	156; 73; 147; 230; 258
17	L-cysteine[a]	15.60	220; 73; 218; 100; 147
18	L-glutamic acid[a]	17.00	246; 73; 128; 147; 156
19	L-phenylalanine[a]	17.12	218; 73; 192; 147; 91
20	L-ornithine[a]	21.57	142; 73; 174; 200; 420
21	L-lysine[a]	23.94	73; 174; 217; 156; 92
22	L-tyrosine[a]	24.38	218; 73; 280; 179; 147
	Amines		
23	ethanolamine[a]	10.37	174; 73; 186; 147; 100
24	putrescine[a]	19.56	174; 73; 214; 86; 59
25	cadaverine[a]	21.83	174; 73; 375; 59; 86
26	dopamine[a]	23.56	174; 73; 86; 103; 217
	Other *N*-Compounds		
27	*N*-acetyl glucosamine	28.04	73; 75; 147; 103; 205
28	spermine[a]	29.90	174; 73; 200; 201; 172
29	spermidine[a]	31.25	73; 144; 116; 174; 156
30	3-hydroxypyridine[a]	8.45	152; 167; 73; 45; 78
31	nicotinic acid[a]	10.77	180; 136; 106; 78; 51
32	2,4-dihydroxypyrimidine[b]	11.44	241; 99; 256; 73; 147
33	4,5-dimethyl-2,6-dihydroxypyrimidine	11.95	269; 284; 270; 147; 127
34	5-methyl-2,4-dihydroxypyrimidine[b]	12.52	73; 255; 113; 147; 270
35	thiazolidine-4-carboxylic acid	12.90	73; 160; 75; 59; 147
36	indole[a]	13.20	189; 174; 73; 79; 75
37	aminomalonic acid[a]	13.86	73; 147; 218; 232; 320
38	4-aminobutyric acid[a]	14.97	174; 73; 147; 304; 59
39	cytosine[a]	15.01	240; 254; 255; 256; 258
40	5-aminovaleric acid[a]	17.07	174; 73; 82; 147; 318
41	5-hydroxyindole[b]	20.45	277; 73; 189; 262
42	citrazinic acid	22.59	356; 73; 371; 147; 253
43	pantothenic acid[a]	25.69	73; 103; 291; 117; 157
44	xanthine[b]	26.41	73; 353; 368; 147; 294
45	indole-3-acetic acid[a]	29.68	73; 290; 407; 392; 202
	Polyols (Polyhydric Alcohols)		
46	1,3-propanediol[a]	7.17	147; 73; 130; 115; 66
47	glycerol[a]	10.48	73; 147; 205; 117; 133
48	1,2,5-trihydroxypentane	13.66	143; 73; 147; 71; 233
49	1,2,3,4-tetrahydroxybutane[b]	14.80	73; 217; 103; 147; 205
50	1,2,5,6-hexanetetrol	19.54	73; 129; 147; 103; 245
51	glucitol[b]	24.78	73; 319; 202; 205; 147
52	inositol[a]	28.14	73; 147; 318; 217; 305
	Carbohydrates		
53	d-(−)-arabinose[a]	18.31	73; 103; 217; 307; 147
54	d-(−)-ribose[a]	18.67	73; 307; 217; 103; 147
55	D-fructose[a]	23.31	73; 103; 217; 147; 307
56	D-glucose[a]	23.76	73; 319; 205; 147; 160
57	mannose[a]	23.91	319; 73; 147; 205; 160
58	galactose[a]	24.32	73; 319; 205; 147; 217
59	maltose[b]	39.26	73; 204; 361; 147; 217
60	gentiobiose[b]	39.49	73; 204; 361; 147; 217
	Bile Acids		
61	cholesterol[a]	42.78	129; 73; 75; 329; 368
62	deoxycholic acid[a]	44.50	255; 73; 428; 75; 266
63	cholic acid[a]	44.75	73; 253; 75; 147; 426
	Inorganic Acids		
64	hydrogen sulfide	5.73	163; 73; 178; 147; 75
65	phosphoric acid[a]	10.52	299; 73; 133; 314; 207
	Monocarboxylic Acids		
66	valeric acid[a]	6.00	75; 159; 73; 117; 132
67	4-methylvaleric acid[a]	6.87	75; 73; 173; 117; 55
68	2-(methoxyimino)propanoic acid[b]	7.12	73; 174; 59; 89; 115
69	hexanoic acid[a]	7.39	75; 73; 173; 117; 131
70	heptanoic acid[a]	8.82	75; 73; 187; 117; 89
71	cyclohexanecarboxylic acid[a]	9.31	73; 75; 185; 55; 129
72	octanoic acid[a]	10.25	73; 75; 201; 117; 129
73	pipecolinic acid[a]	11.84	156; 73; 230; 147; 158
74	decanoic acid[a]	13.41	73; 75; 117; 229; 132
75	palmitic acid[a]	26.41	313; 73; 117; 75; 129
76	linoleic acid[a]	30.03	337; 73; 75; 67; 81
77	oleic acid[a]	30.14	75; 73; 117; 339; 129
78	stearic acid[a]	30.71	341; 73; 117; 75; 129
	Di(multi)-carboxylic Acids		
79	maleic acid[b]	10.94	147; 73; 245; 45; 75
80	succinic acid[a]	10.97	147; 73; 75; 247; 129
81	methylsuccinic acid[a]	11.18	147; 73; 75; 261; 217
82	fumaric acid[a]	11.49	245; 73; 147; 75; 143
83	glutaric acid[a]	12.48	147; 73; 75; 55; 261
84	adipic acid[a]	14.42	75; 73; 111; 147; 275
85	3-methylhexanedioic acid	15.03	73; 75; 125; 289; 147
86	suberic acid[a]	18.60	75; 73; 187; 55; 129
87	tricarballylic acid[a]	19.56	73; 75; 147; 377; 185
88	azelaic acid[b]	20.86	73; 75; 317; 117; 129
89	sebacic acid[a]	23.10	73; 75; 129; 331; 55
90	dodecanedioic acid[a]	24.19	73; 75; 55; 217; 117
	Hydroxyl Acids		
91	lactic acid[a]	7.26	147; 73; 117; 191; 219
92	glycolic acid[a]	7.49	147; 73; 177; 205; 133
93	3-hydroxybutyric acid[a]	8.75	147; 73; 117; 191; 233
94	2-hydroxyhexanoic acid[a]	9.87	73; 159; 75; 147; 103
95	4-hydroxyvaleric acid[b]	10.18	73; 117; 147; 75; 157
96	5-hydroxyvaleric acid[b]	11.28	147; 73; 75; 247; 55
97	glyceric acid[a]	11.35	73; 147; 189; 292; 103
98	5-hydroxyhexanoic acid[a]	11.71	117; 73; 75; 171; 147
99	3,4-dihydroxybutyric acid[a]	13.12	73; 147; 223; 189; 321
100	2-hydroxycyclohexanecarboxylic acid[b]	13.24	147; 81; 273; 129; 183
101	trans-4-hydroxycyclohexanecarboxylic acid	13.99	73; 81; 129; 273; 75
102	malic acid[a]	14.24	73; 147; 233; 245; 189
103	*cis*-4-hydroxycyclohexanecarboxylic acid	14.56	73; 75; 129; 81; 273
104	2,3,4-trihydroxybutyric acid[a]	15.93	73; 147; 292; 220; 205
105	2-hydroxyglutaric acid[a]	16.04	73; 129; 147; 247; 75
106	4-hydroxycyclohexylacetic acid[a]	16.73	73; 75; 170; 129; 81
107	gluconic acid	17.97	73; 204; 103; 147; 133
108	D(−)-quinic acid[b]	22.97	73; 345; 147; 255; 204
109	glucuronic acid[b]	24.95	73; 333; 160; 147; 189
110	glucaric acid[b]	25.38	73; 333; 147; 292; 305
	Phenolics		
111	phenol[b]	7.13	151; 166; 121; 92; 65
112	*p*-cresol[a]	8.60	165; 180; 91; 135; 75
113	benzoic acid[a]	10.04	179; 105; 77; 135; 75
114	phenylacetic acid[a]	10.79	73; 75; 164; 193; 91
115	pyrocatechol[a]	11.11	73; 254; 151; 239; 136
116	3-phenylpropionic acid[a]	12.73	104; 75; 207; 73; 91
117	2-hydroxybenzoic acid[b]	14.58	73; 267; 135; 91; 193
118	pyrogallol[a]	15.31	239; 73; 342; 133; 327
119	3-hydroxybenzoic acid[a]	15.65	267; 73; 282; 223; 193
120	4-hydroxyphenylethanol[b]	15.78	179; 73; 282; 193; 267
121	3-hydroxyphenylacetic acid[a]	16.62	73; 164; 147; 296; 281
122	4-hydroxybenzoic acid[a]	17.11	267; 223; 73; 193; 282
123	4-hydroxyphenylacetic acid[a]	17.32	73; 179; 296; 252; 281
124	3-(3-hydroxyphenyl)propionic acid[a]	19.24	205; 192; 73; 75; 310
125	3-(4-hydroxyphenyl)propionic acid[a]	20.00	179; 192; 73; 310; 75
126	3,4-dihydroxyphenylethanol	20.31	73; 267; 370; 103; 89
127	2,5-dihydroxybenzoic acid[a]	20.55	355; 73; 356; 267; 223
128	3,4-dihydroxybenzoic acid[a]	21.53	193; 73; 370; 355; 311
129	3,4-dihydroxyphenylacetic acid[a]	21.80	73; 384; 179; 267; 207
130	2,5-dihydroxyphenylacetic acid[a]	22.42	73; 384; 341; 147; 252
131	3,4-dihydroxyphenylpropionic acid[a]	24.43	179; 73; 398; 267; 280
132	3,4,5-trihydroxybenzoic acid[b]	24.90	281; 458; 443; 179; 355
133	caffeic acid[a]	28.67	73; 396; 219; 207; 307

[a] These compounds were validated by a reference substance. [b] These compounds were validated by retention indexes (RI).

Table 2. Linearity and Repeatability of Sample Preparation and GC/MS Analysis Based on the Pooled Fecal Water Sample[a]

compounds	linearity: sampling (μL)	linearity: correlation coefficient (R^2)	repeatability (RSD, %): sampling (μL)	repeatability (RSD, %): GC/MS
Carboxylic Acids				
cyclohexanecarboxylic acid	1.25−100	0.993	7.77	1.09
decanoic acid	6.67−75	0.994	5.29	2.54
fumaric acid	1.25−75	0.991	9.16	5.43
glutaric acid	1.25−75	0.995	6.95	4.73
adipic acid	2.5−100	0.997	6.95	2.50
hexanoic acid	5−100	0.992	5.64	3.90
methylsuccinic acid	0.5−100	0.993	3.23	3.51
4-methylvaleric acid	2.5−100	0.990	9.85	3.14
myristic acid	2.5−100	0.995	7.31	4.10
nonanoic acid	5−50	0.989	12.83	6.72
palmitic acid	1.25−100	0.998	13.25	2.43
pimelic acid	1.25−75	0.991	4.03	18.77
suberic acid	12.5−100	0.998	10.53	5.03
succinic acid	0.5−100	0.990	5.70	5.44
tricarballylic acid	1.25−100	0.997	14.58	8.57
valeric acid	0.5−100	0.995	4.78	1.59
Amino Acids				
alanine	1.25−50	0.994	19.55	12.81
β-alanine	1.25−75	0.998	5.13	3.95
cysteine	6.67−50	0.994	7.06	13.33
glutamic acid	0.5−75	0.990	8.15	11.93
glycine	1.25−75	0.998	6.93	3.90
isoleucine	1.25−75	0.998	7.86	7.06
leucine	1.25−75	0.994	4.27	5.75
lysine	5−50	0.996	4.37	7.38
methionine	6.67−75	0.991	3.91	11.88
ornithine	12.5−50	0.995	3.48	7.38
proline	1.25−50	0.993	9.71	14.05
pyroglutamic acid	2.5−75	0.991	11.78	14.31
serine	0.5−50	0.998	3.25	12.15
threonine	1.25−50	0.994	3.82	6.47
tyrosine	1.25−100	0.991	5.56	3.21
valine	1.25−75	0.995	4.14	6.43
Amines				
cadaverine	1.25−75	0.992	5.03	1.85
ethanolamine	1.25−75	0.994	6.23	6.25
Polyols				
glucitol	1.25−100	0.995	4.72	1.90
glycerol	5−100	0.994	2.74	2.79
inositol	2.5−100	0.997	4.04	2.92
Carbohydrates				
arabinose	0.5−100	0.997	2.48	1.12
fructose	0.5−100	0.998	4.64	2.00
galactose	0.5−100	0.998	7.78	1.24
glucose	0.5−100	0.998	2.89	1.24
mannose	0.5−100	0.992	2.57	1.43
ribose	0.5−100	0.998	2.60	2.29

compounds	linearity: sampling (μL)	linearity: correlation coefficient (R^2)	repeatability (RSD, %): sampling (μL)	repeatability (RSD, %): GC/MS
Hydroxyl Acids				
2,4-dihydroxybutyric acid	0.5−100	0.997	3.87	12.09
3,4-dihydroxybutyric acid	1.25−75	0.996	6.61	2.35
glyceric acid	1.25−100	0.991	2.71	3.46
glycolic acid	2.5−100	0.991	5.97	5.55
3-hydroxybutyric acid	1.25−100	0.989	4.68	5.76
4-hydroxycyclohexylacetic acid	2.5−100	0.997	3.90	2.15
cis-4-hydroxycyclohexanecarboxylic acid	0.5−100	0.997	4.95	2.17
2-hydroxyglutaric acid	1.25−100	0.995	5.34	5.21
5-hydroxyhexanoic acid	2.5−100	0.991	3.43	0.69
4-hydroxyvaleric acid	2.5−100	0.992	2.58	6.58
5-hydroxyvaleric acid	1.25−100	0.990	4.38	1.83
malic acid	1.25−100	0.997	2.30	5.16
2,3,4-trihydroxybutyric acid	2.5−100	0.987	9.86	4.71
Phenolics				
benzoic acid	2.5−100	0.987	3.19	2.50
caffeic acid	2.5−100	0.997	7.27	1.62
p-cresol	1.25−75	0.994	11.38	1.67
1,2-dihydroxybenzene	12.5−100	0.999	4.53	3.28
2,5-dihydroxybenzoic acid	2.5−100	0.994	3.63	4.26
3,4-dihydroxybenzoic acid	1.25−100	0.992	4.40	1.68
2,5-dihydroxyphenylacetic acid	0.5−100	0.995	5.14	4.49
3,4-dihydroxyphenylpropionic acid	2.5−100	0.994	1.98	7.49
3-Hydroxybenzoic acid	2.5−75	0.991	4.68	10.69
4-hydroxybenzoic acid	1.25−100	0.991	9.17	5.22
3-hydroxyphenylacetic acid	1.25−75	0.991	2.88	3.91
4-hydroxyphenylacetic acid	1.25−100	0.996	3.86	5.10
3-(4-hydroxyphenyl)propionic acid	1.25−100	0.997	5.61	1.81
phenylacetic acid	0.5−100	0.995	6.14	0.97
3-phenylpropionic acid	0.5−100	0.997	3.82	1.42
Other Compounds				
4-aminobutyric acid	1.25−100	0.996	6.08	3.19
aminomalonic acid	2.5−100	0.992	14.66	4.84
5-aminovaleric acid	1.25−100	0.994	5.80	2.09
2,4-dihydroxypyrimidine	0.5−100	0.998	3.82	8.33
4,5-dimethyl-2,6-dihydroxypyrimidine	0.5−100	0.998	3.63	3.61
3-hydroxypyridine	1.25−100	0.997	4.01	6.70
5-methyl-2,4-dihydroxypyrimidine	6.67−75	0.994	2.70	7.75
nicotinic acid	12.5−100	0.999	5.41	5.01
pantothenic acid	0.5−100	0.999	8.68	4.52
pipecolinic acid	0.5−50	0.995	7.64	3.60
D(−)-quinic acid	0.5−100	0.997	2.73	5.11
1,2,5-trihydroxypentene	0.5−100	0.999	4.42	1.89
Inorganic Compounds				
hydrogen sulfide	0.5−100	0.991	7.67	3.20
deoxycholic acid	2.5−100	0.990	8.21	6.72

[a] Linearity range (μL) was determined by plotting the logarithmic signal intensity against the logarithmic sample volume; R^2: linearity correlation coefficient within the linear range of sample volume.

Absolute Quantification. *Influence of pH of Standard Mixture.* A standard mixture (SM) of compounds commonly found in fecal water was prepared to establish calibration curves for absolute quantification. A preliminary experiment showed that some short and long chain fatty acids had weak responses and consequently resulted in a bad detection. In order to clarify which step(s), protein precipitation, lyophilization, or derivatization, may affect the metabolome, two basic solutions, sodium hydroxide (NaOH) and ammonium hydroxide (NH_4OH), were utilized to regulate the pH of the reference standard mixture to the same value as the pooled human fecal water sample. The pH regulation of standard mixture by NaOH and NH_4OH solutions significantly decreased the loss of short chain fatty acids, but NaOH regulation exhibited a better effect for short chain fatty acids (data not shown). Short chain fatty acids are an important class of water-soluble metabolites which are prone to be lost during lyophilization. Long chain fatty acids hardly dissolve in water and pyridine (reaction medium). For absolute quantification, therefore, free fatty acids are not adequate for constructing calibration curves. Fatty acid sodium salts distinctly decreased

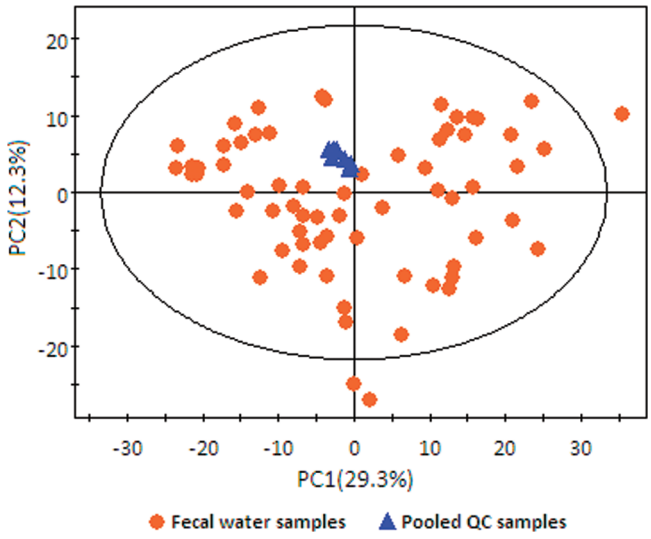

Figure 4. PCA of 65 human fecal water samples and 9 pooled quality control samples. The data were normalized to the total integrated area (1000) per sample and further mean-centered and Pareto-scaled prior to PCA analysis.

the volatile loss during lyophilization and increased the solubility in water, methanol, and the derivatization medium.

The recovery was further studied on the basis of the NaOH-regulated standard mixture (NSM, pH = 7). Recoveries of sample preparation were found to be satisfactory (98.0%−117.8%) for short chain fatty acids, dicarboxylic acids, nonaromatic amino acids, and monosaccharides. However, the recoveries of some other compounds such as long chain fatty acids, tyrosine, and *p*-cresol exceeded 100%−200%. This result suggests the variable matrix effects, which can induce errors in recovery determination. Thus, a simple pH regulation of the reference standard mixture actually cannot provide an identical matrix property to that of fecal water.

Study of Matrix Effects. Matrix effects are an important problem for absolute quantification when preparing standard calibration curves. The significant matrix difference between biological samples and a standard mixture probably results in obvious quantitative differences. To mimic fecal water matrix when preparing quantitative calibration curves, a series of diluted NSM was added into the aliquots of fecal water sample (NSMF group). The signals from the fecal water sample were subtracted from those of NSMF group prior to preparing calibration curves. Four types of calibration curve models were obtained from SM, NSM, and NSMF, as illustrated in Figure 5.

As previously discussed, the pH regulation to neutrality significantly decreased the volatile loss of short chain fatty acids during lyophilization. Figure 5A demonstrates that the NSM group showed a much better signal intensity and linearity for short chain fatty acids than those of the SM group. The simple pH regulation by NaOH can ensure the quality of calibration curves for short chain fatty acids.

The NSM procedure did not give better calibration curves for long chain fatty acids and aromatic compounds than SM, while NSMF significantly increased their linearity correlation coefficients (*R*) and slopes (Figure 5B). The weak solubility (e.g., long chain fatty acids and caffeic acid) and high volatility (e.g., *p*-cresol and 3-phenylpropionic acid) may explain this effect. Dicarboxylic acids (except succinic acid), hydroxyl acids, and heterocyclic com-

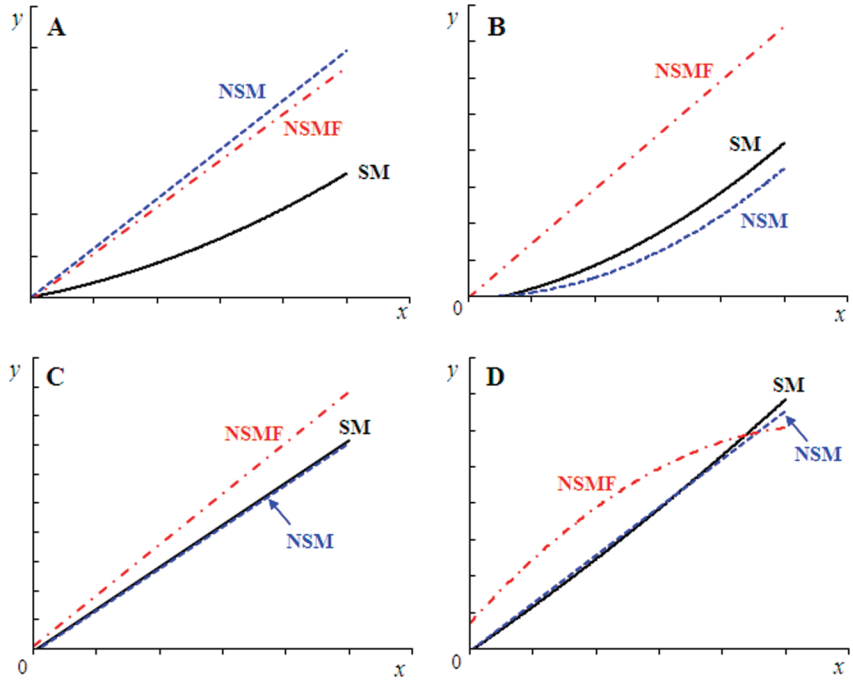

Figure 5. Influence of matrix effects on calibration curves: (A) short chain fatty acids, including valeric acid, 4-methylvaleric acid, and hexanoic acid; (B) long chain fatty acids and aromatic compounds, including palmitic acid, stearic acid, phenylamine, *p*-cresol, phenylacetic acid, 3-phenylpropionic acid, 4-hydroxybenzoic acid, 4-hydroxyphenylacetic acid, 3-(4-hydroxyphenyl)propionic acid, and caffeic acid; (C) dicarboxylic acids, hydroxyl acids, and other heterocyclic compounds, including methylsuccinic acid, hexanedioic acid, fumaric acid, pipecolinic acid, cyclohexanecarboxylic acid, glycolic acid, malic acid, 3-hydroxybutyric acid, and 3-hydroxypyridine; (D) monosaccharides and compounds containing an amino group, except aromatic amino acids, including arabinose, ribose, mannose, galactose, alanine, leucine, proline, cadaverine, 4-aminobutyric acid, 5-aminovaleric acid, and succinic acid (exception). SM, standard mixture without pH regulation; NSM, NaOH regulated standard mixture (identical pH to fecal water); NSMF, NSM spiked with fecal water; *x*, concentration (μg/mL), *y*, mass signal intensity.

Table 3. Quantification of 33 Main Compounds in the Human Fecal Water Samples (n = 17, Mean ± Standard Deviation)

compounds	quantifier (m/z)	calibration curve	concentration (μg/mL)
alanine	116	NSM	396.6 ± 209.5
leucine	158	NSM	264.3 ± 185.4
4-methylvaleric acid	75	NSM	197.0 ± 67.5
succinic acid	147	NSM	192.9 ± 215.2
mannose	73	NSM	182.0 ± 97.8
ribose	73	NSM	169.8 ± 106.1
arabinose	73	NSM	156.3 ± 98.7
phenylacetic acid	73	NSMF	151.9 ± 89.8
valeric acid	75	NSM	146.4 ± 89.2
galactose	73	NSM	134.5 ± 71.0
proline	142	NSM	109.4 ± 83.8
palmitic acid	73	NSMF	108.2 ± 154.5
hexanoic acid	173	NSM	74.8 ± 70.6
malic acid	73	NSMF	72.8 ± 50.8
4-aminobutyric acid	174	NSM	59.5 ± 53.7
3-phenylpropionic acid	104	NSMF	42.7 ± 37.0
phenylalanine	218	NSMF	41.8 ± 65.7
5-aminovaleric acid	174	NSM	30.7 ± 29.8
cyclohexanecarboxylic acid	73	NSMF	26.3 ± 63.3
cadaverine	174	NSM	22.2 ± 24.5
hexanedioic acid	73	NSMF	21.5 ± 13.2
methylsuccinic acid	147	NSMF	20.0 ± 27.7
glycolic acid	147	NSMF	13.3 ± 4.8
stearic acid	341	NSMF	12.8 ± 21.4
p-cresol	165	NSMF	10.3 ± 10.7
fumaric acid	245	NSMF	9.8 ± 5.9
4-hydroxyphenylacetic acid	73	NSMF	8.7 ± 6.7
pipecolinic acid	156	NSMF	7.9 ± 9.6
3-hydroxypyridine	152	NSMF	5.2 ± 3.6
3-(4-hydroxyphenyl)propionic acid	179	NSMF	4.3 ± 5.0
4-hydroxybenzoic acid	267	NSMF	4.1 ± 4.3
caffeic acid	219	NSMF	2.7 ± 2.4
3-hydroxybutyric acid	147	NSMF	2.3 ± 1.0

pounds exhibited a good linearity when preparing calibration curves by SM, NSM, or NSMF methods. However, the NSMF procedure generally showed a higher or even significantly higher intensity than those of the other two methods (Figure 5C). Human fecal water is a stable electrolyte system with good buffer capacity and pH value between 6 and 7. The majority of compounds coexist as neutral salts, which increase boiling points and solubility of compounds in an aqueous system. Losses of compounds with high volatility and/or weak solubility from the SM or NSM procedure were avoided by spiking fecal water as matrix (NSMF group).

Interestingly, as shown in Figure 5D, NSMF demonstrates an upward convex curve with significantly elevated signal intensity for monosaccharides and the compounds containing amino groups, except aromatic amino acids. Such a curve was not observed for SM and NSM groups. NSM exhibited a better linearity than SM and was, therefore, the best selection.

This experiment suggests that the volatility and solubility of compounds probably are two main factors to explain the matrix effects, which are essential when absolute quantification is considered. For this reason, the NSM method is appropriate for the absolute quantification of short chain fatty acids, monosaccharides, and the compounds containing an amino group, except for aromatic amino acids. However, for the other compounds, the NSMF method will provide more appropriate calibration curves for quantification.

On the basis of the categorized calibration curves, 33 compounds identified in the clinical human fecal water samples were quantified. Table 3 shows that the major components present at more than 100 μg/mL were alanine, leucine, proline, 4-methylvaleric acid, valeric acid, palmitic acid, succinic acid, arabinose, ribose, mannose, galactose, and phenylacetic acid. Although many factors, such as genetic background, diet, physiological status, and gut microflora, may cause a distinct interindividual variability of fecal metabolome, current work aims to principally present a reliable absolute quantitative method using TMS derivatization by the family of chemicals, because of a very wide range of compounds with varied chemical and physical properties in human fecal water.

CONCLUSIONS

Human fecal water is a physicochemically complex mixture of varieties of metabolites. In this study, a quantitative metabolomic protocol focusing on a human fecal water sample was developed on the basis of trimethylsilylation derivatization and GC/MS analysis. The pooled sample from a set of clinical human fecal water samples was utilized in this study. Methanol was the best solvent for protein precipitation and the extraction of fecal water metabolome. The signal response was linear when sampling volume was lower than 50 μL. Almost all fecal water compounds exhibited a good linearity with correlation coefficient higher than 0.99. The method exhibited a good repeatability for both sample preparation and GC/MS analysis with the RSDs lower than 10% for most compounds and less than 20% for a few other ones. The variability of the analytical system was significantly lower than that among the biological samples. Totally, 133 compounds were structurally identified from a set of clinical human fecal water samples. The pH and matrix effects of a standard mixture significantly affected the accuracy of quantitative calibration curves and, thus, was the main influencing factor for absolute quantification. The increased pH value significantly inhibited the volatile loss of short chain fatty acids during lyophilization. The calibration curves prepared from the standard mixture with pH regulation (~7) by NaOH solution are necessary for the absolute quantification of short chain fatty acids, monosaccharides, and the compounds containing an amino group (except aromatic amino acids). However, for the other compounds, it is recommended to construct calibration curves by spiking fecal water to a pH-regulated standard mixture in order to avoid matrix effects. This may suggest that fecal water provided a stable electrolyte system with good buffer capacity and pH value between 6 and 7, which decreased the volatile loss of standard compounds during lyophilization and improved their solubility in the derivatization medium. The proposed method showed a good linearity and repeatability of sample preparation and GC/MS analysis and was promising for quantitative metabolomic study of human fecal water samples. In the perspective of a multiplatform approach, it would be interesting to consider metabolomic analysis of the solid residue after fecal water extraction by LC/MS, in order to obtain complementary information.

ACKNOWLEDGMENT

X.G. was thankful to INRA, France, for providing a bursary award. This work was partly funded by the ANR METAPROFILE.

Received for review March 12, 2010. Accepted June 28, 2010.

AC1006552