RESEARCH PAPER

CrossMark

Balancing metabolome coverage and reproducibility for untargeted NMR-based metabolic profiling in tissue samples through mixture design methods

Hong Zheng[1] · Zhitao Ni[1] · Aimin Cai[1] · Xi Zhang[1] · Jiuxia Chen[1] · Qi Shu[1] · Hongchang Gao[1]

Abstract

Untargeted metabolomics attempts to acquire a comprehensive and reproducible set of small-molecule metabolites in biological systems. However, metabolite extraction method significantly affects the quality of metabolomics data. In the present study, we calculated the number of peaks (NP) and coefficient of variation (CV) to reflect metabolome coverage and reproducibility in untargeted NMR-based metabolic profiling of tissue samples in rats under different methanol/chloroform/water (MCW) extraction conditions. Different MCW extractions expectedly generated diverse characteristics of metabolome. Moreover, the classic MCW method revealed tissue-specific differences in the NP and CV values. To obtain high-quality metabolomics data, therefore, we used mixture design methods to optimize the MCW extraction strategy by maximizing the NP value and minimizing the CV value in each tissue sample. Results show that the optimal formulations of MCW extraction were 2:2:8 (ml/mg tissue) for brain sample, 2:4:6 (ml/mg tissue) for heart sample, 1.3:2:8.7 (ml/mg tissue) for liver sample, 4:2:6 (ml/mg tissue) for kidney sample, 2:3:7 (ml/mg tissue) for muscle sample, and 2:4:6 (ml/mg tissue) for pancreas sample. Therefore, these findings demonstrate that different tissue samples need a specific optimal extraction condition for balancing metabolome coverage and reproducibility in the untargeted metabolomics study. Mixture design method is an effective tool to optimize metabolite extraction strategy for tissue samples.

Keywords Metabolomics · Metabolite extraction · Metabolome reproducibility · Optimization · Tissue-specific

Introduction

Metabolomics is the end-point of omics cascade that aims to analyze a comprehensive set of metabolites in bio-samples and explore their changes related to genomic and proteomic perturbations [1]. Currently, metabolomics plays an essential role in the field of biological and medical sciences [2]. In general, there are two complementary approaches for metabolomics studies, namely, targeted and untargeted metabolomics. The untargeted method focuses on comprehensive metabolites without bias; however, the targeted method detects a single or a panel of specific metabolites. Each method possesses its advantages, for example, the targeted approach can achieve more sensitive and quantitative analysis. In an untargeted study, by measuring global metabolic profiling, novel metabolites related to disease mechanisms or biomarkers are more likely to be discovered. Thus, the untargeted approach is usually performed prior to the targeted investigation.

Metabolite extraction is a vital step in metabolomics [3]. Unlike classic targeted strategy, untargeted metabolomics requires a fast and reproducible sample preparation method. Additionally, this method has to cover a wide range of metabolites. Choi and Verpoorte revealed that "What you see is what you extract" [4]. Thus, for ensuring the reliability of a metabolomics study, it is essential to develop an effective and reproducible protocol of metabolite extraction [3]. Many researchers have paid attention to this issue and tried to optimize sample preparation for improving metabolomics data quality. Kim et al. optimized sample preparation methods including

---

✉ Hongchang Gao
gaohc27@wmu.edu.cn

[1] Institute of Metabonomics & Medical NMR, School of Pharmaceutical Science, Wenzhou Medical University, Wenzhou 325035, China

sampling and extraction based on GC-MS to facilitate a reliable and accurate metabolomics analysis for *Saccharomyces cerevisiae* [5]. By optimizing combination of different solvents, the coverage of metabolome can be greatly expanded in bio-fluid [6, 7], cell [8, 9], and tissue samples [10, 11]. For achieving a high efficiency and reproducible preparation protocol of esophageal tissue, solvent extraction and tissue disruption methods were optimized by Wang et al. [12]. In addition, Naz et al. [13] optimized sample treatment and analytical method to improve the coverage and linearity of metabolome in lung tissue for a multiplatform approach. The intensity and reproducibility of brain tissue metabolome can also be improved by optimizing tissue lysis and metabolite extraction methods using NMR, LC-MS, and GC-MS [14]. Therefore, metabolite extraction optimization will be a significant and ongoing work in the field of metabolomics.

The most popular strategy for tissue metabolite extraction is the two-phase extraction system including methanol, chloroform, and water; the upper and lower phases contain polar and non-polar metabolites, respectively [15]. However, the final ratio of methanol/chloroform/water (MCW) extraction [16, 17] is miscellaneous, which brings us an inevitable problem, that is, how to set this ratio for a specific tissue sample? A design of experiment (DoE) approach is a solution to overcome this issue. DoE is an optimization method that makes possible to simultaneously evaluate the importance of each factor and their interactions, model the factor-response relationship with a minimal number of experiments, and optimize factors setting to achieve a desirable response [18]. Therefore, the aim of this work was to use the DoE method as a tool to optimize the final ratio of MCW extraction for maximizing the metabolome coverage and reproducibility in brain, heart, liver, kidney, muscle, and pancreas samples of rats and to evaluate tissue-specific differences in metabolite extraction.

Materials and methods

Chemicals

Methanol and chloroform were of analytical grade and purchased from Aladdin (Shanghai, China). Distilled water was obtained from a Millipore Direct-Q3 UV system (MilliPore, Boston, MA, USA). Please note that methanol and chloroform are harmful; therefore, these solvents must be handled in a fume hood and avoid skin contact or inhalation.

Animals

Eight-week-old Sprague-Dawley rats (male; $n = 20$; body weight = 225 ± 12 g) were purchased from the SLAC Laboratory Animal Co. Ltd. (Shanghai, China). All rats were housed in the specific pathogen-free colony under a controlled condition (room temperature = 22 °C; light/dark cycle = 12 h/ 12 h, lights on at 7:00 a.m.) at the Laboratory Animal Center of Wenzhou Medical University (Wenzhou, China). Rats were given free access to standard rat chow and tap water. This study was conducted in accordance with the "Guide for the Care and Use of Laboratory Animals" and approved by the Institutional Animal Care and Use Committee of Wenzhou Medical University.

Tissue sample collection

All rats were sacrificed by decapitation after a 12-h fasting. Brain, heart, liver, kidney, muscle, and pancreas tissues were extracted, immediately frozen in liquid nitrogen, and stored at − 80 °C until use.

Metabolite extraction optimization and NMR-based metabolomics analysis

To ensure consistency for the extraction optimization study, brain, heart, liver, kidney, muscle, and pancreas tissues obtained from 20 rats were correspondingly mixed and homogenized using a handheld homogenizer. Polar metabolome were extracted using the MCW extraction method. The general procedure of MCW method includes the following: the frozen tissue was weighed into a tube and then ice-cold methanol and distilled water were subsequently added to the tube; the mixture was homogenized using a handheld homogenizer; ice-cold chloroform was subsequently added to the mixture and vortex-mixed for 10 s; the mixture was stood on ice for 15 min and centrifuged at 10,000×g for 15 min at 4 °C; lastly, the supernatant was transferred to a new tube, lyophilized for 24 h, and kept at − 80 °C until analysis.

The lyophilized extract was reconstituted in the tube with 0.6 ml of D_2O (99.5%) containing 0.05% of sodium trimethylsilyl propionate-d_4 (TSP) and transferred to a 5 mm NMR tube for metabolomics analysis. ^{1}H NMR metabolic profiling of tissue extract was measured by using a Bruker AVANCE III 600 MHz NMR spectrometer (Bruker BioSpin, Rheinstetten, Germany). A standard single-pulse sequence with water signal pre-saturation ("zgpr," Bruker) was performed at 25 °C. The main acquisition parameters were set as follows: scans, 256; acquisition time, 2.65 s per scan; data points, 64 K; spectral width, 12,000 Hz; relaxation delay, 6 s. Then, the ^{1}H NMR spectra were referenced to the TSP peak at 0.0 ppm and preprocessed using auto-phase and auto-baseline corrections using Topspin software (v2.1 pl4, Bruker BioSpin, Germany).

During the procedure of MCW extraction, we speculate that the ratio of methanol, chloroform, and water will greatly impact on the data quality of metabolome. In the present study, metabolite extraction was carried out with different

solvent mixtures defined using a mixture design approach. The mixture design method is an efficient DoE approach that is able to determine the optimal mixture proportion of different constituents for achieving a desired outcome [19]. For untargeted metabolomics, the desired result is to acquire more abundant and reproducible metabolic information from biological samples. Hence, in this study, we also used the mixture design method to optimize the final ratio of MCW extraction for maximizing the metabolome coverage and reproducibility in different tissues of rats, including brain, heart, liver, kidney, muscle, and pancreas. The effects of three independent variables, namely methanol (X1), chloroform (X2), and water (X3), on the number of peaks (NP, Y1) and coefficient of variation (CV, Y2) were analyzed using the mixture design method under SAS software (design of experiments, SAS 9.2, SAS Institute Inc., Cary, NC, USA). Ten design points labeled from E1 to E10 were listed in Table 1. The range of independent variables was set to 2–8 ml/g wet tissue weight, but note that this range can be user-defined. The NP value was automatically counted as peaks with the intensity above 10,000 from each NMR spectrum using Topspin software (v2.1 pl4, Bruker BioSpin, Germany). In addition, each experiment was conducted in five times and the CV value was calculated as the ratio of the standard deviation to the mean of the NP value to indicate data reproducibility using Microsoft Excel 2010 (Microsoft Corp., Redmond, WA). The threshold of peak intensity and the CV value can also be user-defined, but please note that these settings should be fixed for all the experiments.

The statistic method was applied to find the best fitted model of three independent variables using SAS 9.2 software. The significances of three independent variables and their interactions were evaluated by using ANOVA and P value below 0.05 was considered statistically significant.

Multivariable analysis

After preprocessing, all NMR spectra from the same tissue were aligned using the "*icoshift*" method [20] in MATLAB (R2012a, The Mathworks Inc., Natick, MA, USA). The spectral region from 0.0 to 10.0 ppm excluding the residual water signals from 4.7 to 5.0 ppm was subdivided and integrated to binning data with a size of 0.01 ppm for multivariable analysis. Principal component analysis (PCA) based on the Pareto-scaled binning data was performed to examine an overview change in metabolic pattern among different MCW extractions using SIMCA 12.0 software (Umetrics, Umeå, Sweden).

Metabolite identification and quantification

The NMR signal was assigned in accordance with the Human Metabolome Database [21] and Chenomx NMR Suite 7.1 (Chenomx, Alberta, Canada). Moreover, two-dimensional (2D) ^{1}H-^{1}H correlated spectroscopy (COSY) and ^{13}C-^{1}H heteronuclear single quantum coherence (HSQC) experiments were performed to analyze the representative samples for verifying our tentative assignments. For quantification of specific metabolite, its peak area was manually integrated under Topspin software and calculated on the basis of its peak area by reference to the internal standard TSP concentration. Then, the relative concentration of metabolite was Pareto-scaled and visualized as heatmaps, and cluster analysis was performed by Ward's method and Euclidean distance using MetaboAnalyst 3.0 [22].

Results and discussion

Overview of the tissue-specific effect of MCW extraction on metabolic profiling

Metabolic profiling aims to get extensive metabolome coverage in biological samples, which has been used as a promising tool to explore the biomarker and mechanism of disease [23–25]. However, many factors may influence metabolic profiling, such as sample collection and storage [14], sample extraction and reconstitution [26] as well as sample analytical method [27]. In this study, to investigate the effect of extraction procedures on metabolic profiling in different tissues, we examined differences in metabolic patterns among different MCW extractions using PCA. As demonstrated in PCA score plots (Fig. 1), a clear clustering of samples was observed according to the extraction method. E9 and E10 with a higher ratio of methanol to water were clearly separated from the rest of extraction procedures along PC1 for brain (Fig. 1A), liver
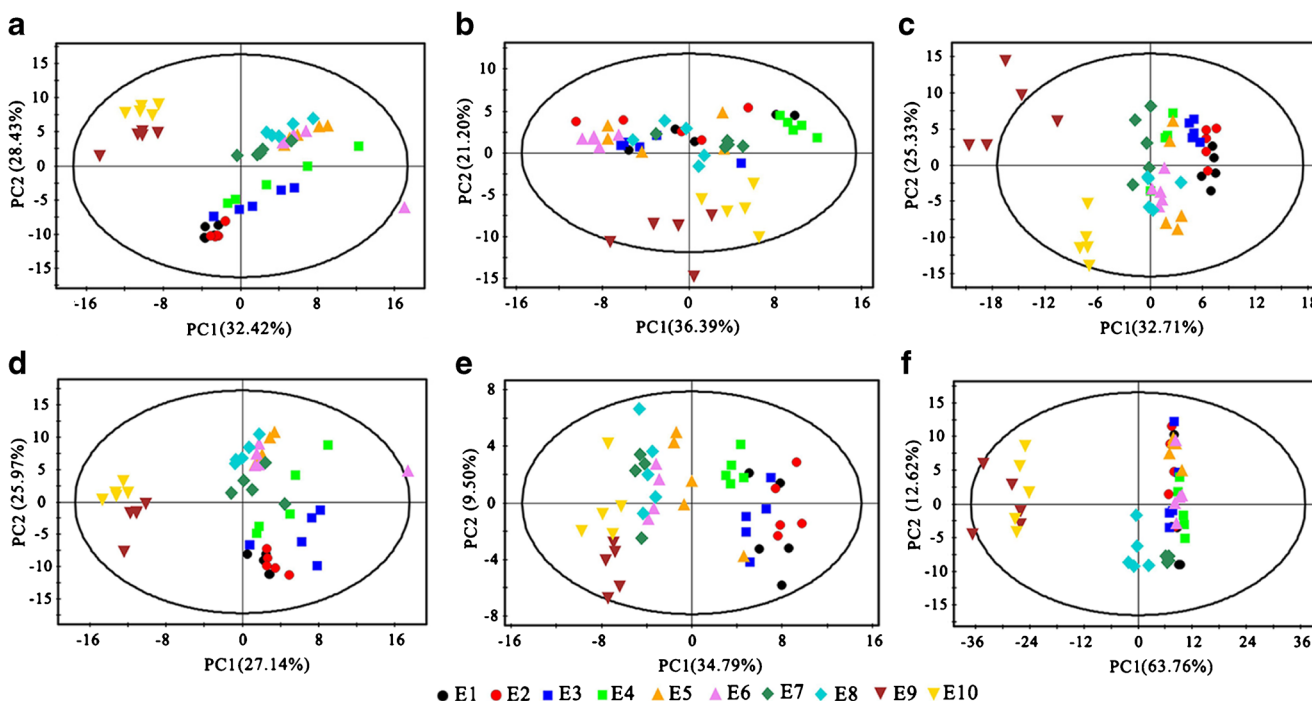
Table 1 Mixture design for optimization of methanol/chloroform/water extraction

Run	Coded factor			Uncoded factor (ml/g)		
	M[a]	C[b]	W[c]	M	C	W
E1	0	0	1	2	2	8
E2	0	0.33	0.67	2	4	6
E3	0	0.67	0.33	2	6	4
E4	0	1	0	2	8	2
E5	0.33	0	0.67	4	2	6
E6	0.33	0.33	0.33	4	4	4
E7	0.33	0.67	0	4	6	2
E8	0.67	0	0.33	6	2	4
E9	0.67	0.33	0	6	4	2
E10	1	0	0	8	2	2

[a] Methanol

[b] Chloroform

[c] Water

Fig. 1 PCA score plots of untargeted NMR-based metabolic profiling in tissue samples under different methanol/chloroform/water (MCW) extraction conditions. Tissue sample: **A** brain; **B** heart; **C** liver; **D** kidney; **E** muscle; **F** pancreas. Extraction condition (M/C/W, ml/g tissue): E1, 2:2:8; E2, 2:4:6; E3, 2:6:4; E4, 2:8:8; E5, 4:2:6; E6, 4:4:4; E7, 4:6:2; E8, 6:2:4; E9, 6:4:2; E10, 8:2:2

(Fig. 1C), kidney (Fig. 1D), and pancreas (Fig. 1F) tissues and PC2 for heart tissue (Fig. 1B). Additionally, along PC1, a clear separation between E1–E4 and E5–E10 was observed in muscle tissue (Fig. 1E). Overall, PCA results reveal that MCW extraction had a tissue-specific effect on metabolic profiling.

Impact of MCW extractions on metabolome coverage and reproducibility in different tissues

The desirable property of untargeted metabolic profiling is expected to be not only wide-ranging, but also stable [28]. In the present study, we calculated the number of peaks (NP) and coefficient of variation (CV) to reflect these two properties under different MCW extractions. The effects of different MCW extractions on the NP and CV values are presented using a radar chart method, which is commonly applied to visualize one attribute related with multiple factors, as shown in Fig. 2. The radar chart consists of ten equiangular spokes and each spoke represents one of the MCW extraction conditions. The axis of a spoke indicates the range of the variable (the NP or CV value) and the axis's origin (the center of the graphic) represents the minimum value of the variable and the end the maximum value (Fig. 2). In brain tissue, all MCW extractions studied herein achieved a NP value above 100 (Fig. 2A), whereas the CV values exceeded 8% for E2, E5, E6, and E8 (Fig. 2G). E1 and E2 possessed a higher value of NP (approximately 120) than other MCW extractions in heart tissue, as shown in Fig. 2B. Yet, relative to E1, E2 had a lower

CV value (Fig. 2H). As can be seen from Fig. 2C, a relatively higher NP value (> 160) was observed for E1, E2, E3, E5, E6, and E8 in liver tissue; moreover, in these extraction conditions, E6 had a lower CV value (< 6%, Fig. 2I). Figure 2D reveals that the NP value reached or exceeded 200 for all MCW extractions in kidney tissue. In addition, MCW extractions excepting E2 and E7 showed a CV value below 6% (Fig. 2J). In muscle tissue, the NP value around 120 was obtained from E1, E2, E3, E5, E6, E8, and E9 (Fig. 2E), and E3 and E9 exhibited a lower CV value (< 6%) than other extraction conditions (Fig. 2K). According to Fig. 2F, E1, E2, E5, and E6 had a higher NP value (> 100) relative to other extractions in pancreas tissue. However, the CV value above 10% was observed from E5 (Fig. 2L). It can be concluded that MCW extraction caused an obvious tissue-specific effect on metabolome coverage and reproducibility. Therefore, optimizing MCW extraction is a necessary step for improving the data quality of metabolome. Meanwhile, the tissue-specific effect on metabolite extraction also needs to be noticed.

Optimization and validation of MCW extractions using mixture design methods

Table 2 lists P values for the effects of methanol, chloroform, water, and their interaction on the NP and CV values in different tissues analyzed by mixture design approaches, and the main effects were illustrated in Fig. 3 for the NP value and Fig. 4 for the CV value. As shown in Fig. 3, the NP values

Fig. 2 Changes in the number of peaks (NP) and coefficient of variation (CV) under different MCW extraction conditions in untargeted NMR-based metabolic profiling. Red line, NP; blue line, CV. Tissue sample: **A, G** brain; **B, H** heart; **C, I** liver; **D, J** kidney; **E, K** muscle; **F, L** pancreas. Extraction condition (M/C/W, ml/ g tissue): E1, 2:2:8; E2, 2:4:6; E3, 2:6:4; E4, 2:8:8; E5, 4:2:6; E6, 4:4:4; E7, 4:6:2; E8, 6:2:4; E9, 6:4:2; E10, 8:2:2. The radar chart consists of ten equiangular spokes, representing ten extraction conditions designed in this study (E1–E10). The axis of a spoke indicates the range of the variable (the NP or CV value), at which the axis's origin (the center of the graphic) represents the minimum value of the variable and the end the maximum value
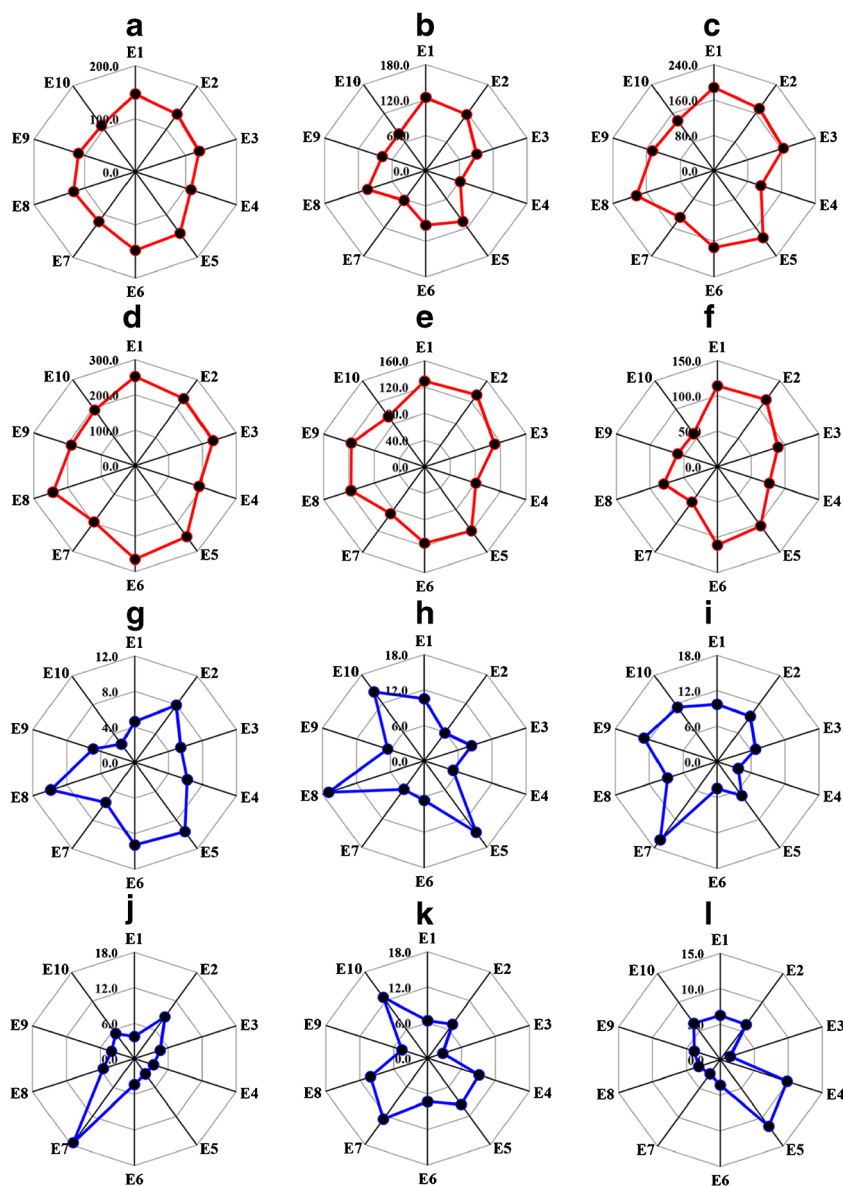
were reduced by increasing methanol and chloroform levels in all tissue samples, which showed statistically significant (Table 2). Overall, an increased trend of the NP values was observed with the increase of water content for all tissues, since we focused on the extraction of polar metabolites in this study. However, for liver and kidney tissues, excessive water addition could result in the reduction of the NP values ($P <$ 0.0001; Fig. 3C; Electronic Supplementary Material (ESM) Fig. S1D). A significant interaction effect of methanol and chloroform was detected on the NP value in liver tissue ($P =$ 0.03, Table 2). Moreover, heart ($P = 0.04$) and liver ($P = 0.02$) tissues exhibited a significant interaction effect of chloroform and water on the NP value (Table 2).

For untargeted metabolomics, the NP value was expected to be maximal, but the CV value to be minimal. In brain tissue, a continuous decrease in the CV value was found with the

increase of chloroform content ($P = 0.006$, Fig. 4A). Figure 4A also reveals that the CV value increased first, and then decreased, by increasing methanol ($P = 0.07$) and water ($P = 0.01$) levels (Table 2). In addition, a significant interaction effect of methanol and water on the CV value was obtained in brain tissue ($P = 0.004$, Table 2). The increasing trend in the CV value was detected in heart tissue with increasing methanol ($P = 0.002$) and water (P = 0.006) contents, as shown in Fig. 4B and Table 2. However, for chloroform, the CV value was dramatically reduced, but then slightly increased ($P = 0.03$, Fig. 4B). As shown in Fig. 4C and Table 2, in liver tissue, methanol addition resulted in a significant increase of the CV value ($P = 0.03$), whereas the CV value was reduced firstly and then increased with the increase of water content ($P = 0.03$). Yet, there was no significant difference for chloroform ($P = 0.18$). Figure 4D illustrates that

Table 2 Statistical results from mixture design methods in different tissue samples

Factor	Brain		Heart		Liver		Kidney		Muscle		Pancreas	
	NP[a]	CV[b]	NP	CV	NP	CV	NP	CV	NP	CV	NP	CV
M[c]	< 0.0001	0.07	< 0.0001	0.002	< 0.0001	0.03	< 0.0001	0.10	0.0003	0.03	0.007	0.17
C[d]	< 0.0001	0.006	< 0.0001	0.03	< 0.0001	0.18	< 0.0001	0.04	0.0008	0.05	0.003	0.05
W[e]	< 0.0001	0.01	< 0.0001	0.006	< 0.0001	0.03	< 0.0001	0.05	0.0001	0.10	0.0005	0.06
M·C[f]	0.22	0.32	0.94	0.07	0.12	0.19	0.59	0.50	0.19	0.64	0.88	0.29
M·W[g]	0.19	0.004	0.42	0.18	0.03	0.27	0.07	0.74	0.70	0.85	0.33	0.88
C·W[h]	0.38	0.20	0.04	0.44	0.02	0.81	0.19	0.40	0.10	0.36	0.24	0.22

[a] Number of peaks;

[b] Coefficient of variation

[c] Methanol

[d] Chloroform

[e] Water

[f] Interaction effect of methanol and chloroform

[g] Interaction effect of methanol and water

[h] Interaction effect of chloroform and water

the CV value increased first and then decreased with increasing of chloroform content in kidney tissue ($P = 0.04$, Table 2). A continuous decline in the CV value was observed in other two solvents, but no significant difference for methanol ($P = $

0.10, Fig. 4D). In addition, we found that the CV value was notably increased with the addition of methanol in muscle tissue ($P = 0.03$, Fig. 4E) and water in pancreas tissue ($P = 0.06$, Fig. 4F). With increasing of chloroform level, the CV
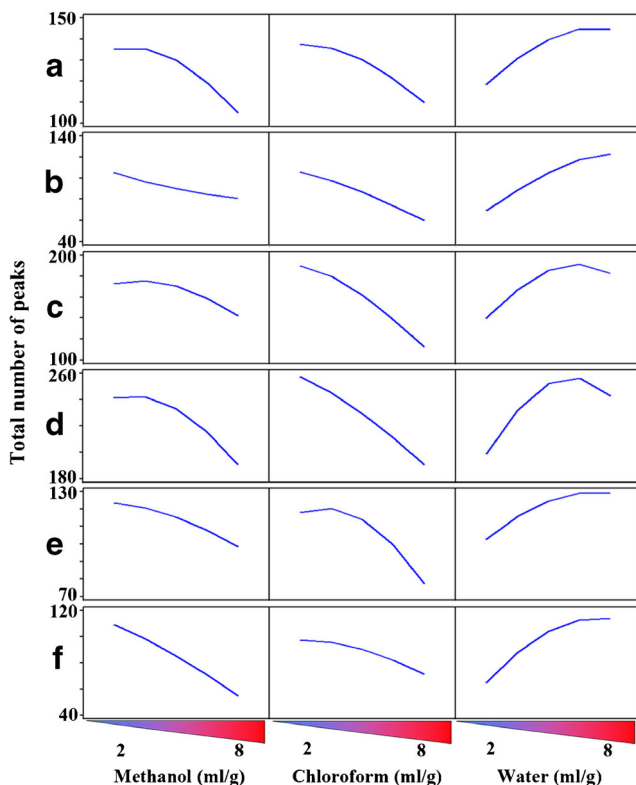

Fig. 3 Changes in total number of peaks in untargeted NMR-based metabolic profiling of tissue samples as a function of methanol, chloroform, and water contents in extraction solvent systems. Tissue sample: **A** brain; **B** heart; **C** liver; **D** kidney; **E** muscle; **F** pancreas
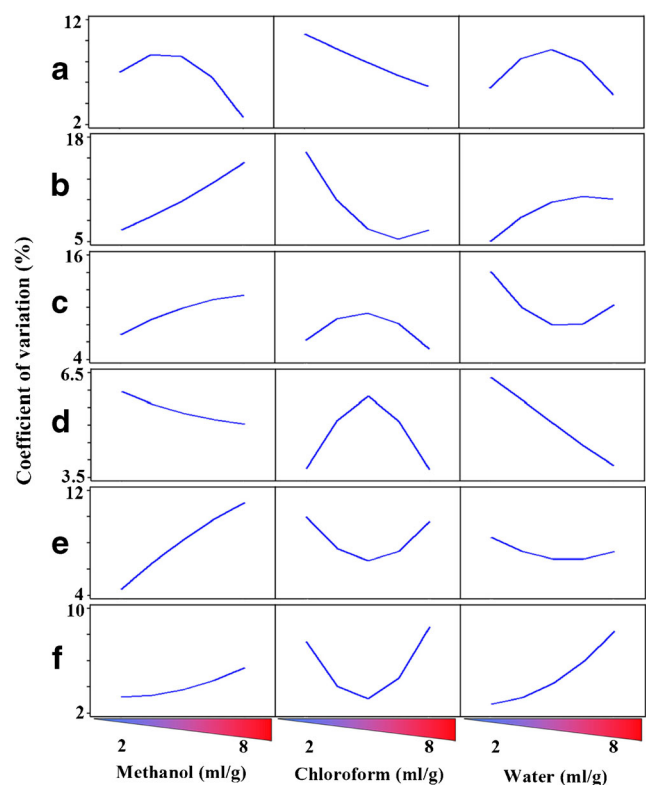
Fig. 4 Changes in coefficient of variation in untargeted NMR-based metabolic profiling of tissue samples as a function of methanol, chloroform, and water contents in extraction solvent systems. Tissue sample: **A** brain; **B** heart; **C** liver; **D** kidney; **E** muscle; **F** pancreas

value decreased first and then increased in both muscle ($P = 0.05$, Fig. 4E) and pancreas tissues ($P = 0.06$, Fig. 4F). However, changes in the CV values were not statistically significant for water in muscle tissue ($P = 0.10$, Fig. 4E) and methanol in pancreas tissue ($P = 0.17$, Fig. 4F). Taken together, tissue-specific differences obviously existed in metabolite extraction, especially for polar metabolome reproducibility.

To balance metabolome coverage and reproducibility, we then used mixture design methods to optimize the MCW extraction strategy by maximizing the NP value and minimizing the CV value. Moreover, a contour plot was applied for the quantitative analysis of each factor on the response [29]. Figure 5 shows the contour plots that indicate changes in the NP and CV values as a function of methanol, chloroform, and water in different tissue samples. We found that the NP value (red line) gradually increased with increasing of water proportion in the MCW solvent system, excepting in liver (Fig. 5C) and kidney (Fig. 5D) samples. For the CV value, however, different tissue samples display different variation tendencies, as illustrated in Fig. 5. In this study, we attempted to maximize the NP value and meanwhile ensure the CV value below 8%, but it should be noted that this selection criterion can be user-defined. Finally, as listed in Table 3, the optimal extraction formulation of MCW was selected for each tissue sample, including 2:2:8 (ml/mg tissue) in brain sample, 2:4:6 (ml/mg tissue) in heart sample, 1.3:2:8.7 (ml/mg tissue) in liver sample, 4:2:6 (ml/mg tissue) in kidney sample, 2:3:7 (ml/mg tissue) in muscle sample, and 2:4:6 (ml/mg tissue) in pancreas sample. These MCW extraction formulations were expected to get the NP and CV values of 146.60 and 4.62 for brain tissue, 117.60 and 5.85 for heart tissue, 193.20 and 7.63 for liver tissue, 248.80 and 3.16 for kidney tissue, 130.30 and 4.88 for muscle tissue, and 116.60 and 6.11 for pancreas tissue, respectively (Table 3). Subsequently, the optimal extraction method of MCW was validated by five independent experiments, and the final results were listed in Table 3. The optimal MCW formulations can achieve the NP and CV values of 143.20 and 3.15 for brain tissue, 109.80 and 5.36 for heart tissue, 173.40 and 6.87 for liver tissue, 258.00 and 3.24 for kidney tissue, 133.80 and 2.33 for muscle tissue, and 124.80 and 6.08 for pancreas tissue, respectively. The validation results indicate that the actual results were close or even better than the predicted results, suggesting a reliable optimization of metabolite extraction.

Metabolite extraction is undoubtedly a key step for metabolomics analysis, so researchers have worked on optimizing this step for NMR-based tissue metabolomics, such as liver [15], muscle [30], brain [31], and human vein tissue [32]. In the present study, we suggest that mixture design method can be an efficient tool to assist the optimization of metabolite extraction in metabolomics research. This method has been widely used to study the relationship between the proportions of different variables and responses and to obtain an optimum formulation for a desired outcome [29]. Therefore, it could be applied to find the optimal extraction medium for better characterization of metabolic profiling. In this study, for example, we systematically evaluated the impact of different solvent extraction systems on the metabolome coverage and reproducibility using the mixture design method and achieved the optimal solvent formulation for metabolite extraction in various animal tissues, including brain, heart, liver, kidney, muscle, and pancreas.
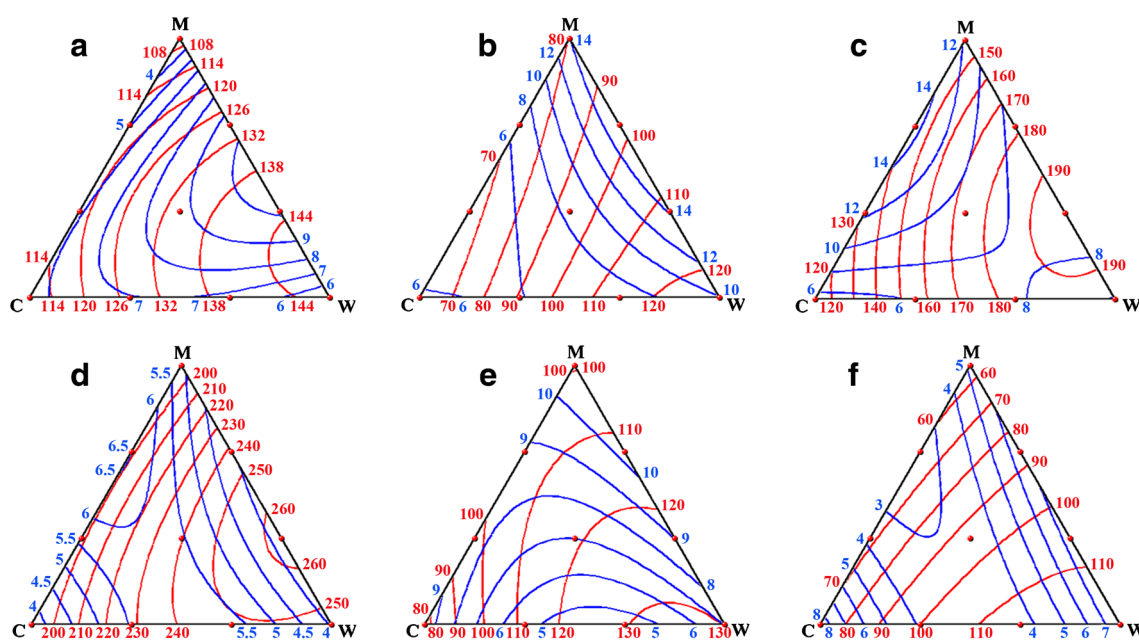

Fig. 5 The contour plots indicating changes in the number of peaks and coefficient of variation as a function of methanol (M), chloroform (C), and water (W) in different tissue samples. Tissue sample: A brain; B heart; C liver; D kidney; E muscle; F pancreas

Table 3 Summary of MCW extractions optimized by mixture design methods in different tissue samples

Sample	Optimal extraction			Prediction		Validation		Identification	
	M[a]	C[b]	W[c]	NP$_P$[d]	CV$_P$[e]	NP$_V$[f]	CV$_V$[g]	M[h]	ID%[i]
Brain	2	2	8	146.60	4.62	143.20	3.15	44	87.99
Heart	2	4	6	117.60	5.85	109.80	5.36	45	91.99
Liver	1.3	2	8.7	193.20	7.63	173.40	6.87	45	89.10
Kidney	4	2	6	248.80	3.16	258.00	3.24	49	85.85
Muscle	2	3	7	130.30	4.88	133.80	2.33	31	90.77
Pancreas	2	4	6	116.60	6.11	124.80	6.08	37	91.54

[a] Methanol (ml/g tissue)

[b] Chloroform (ml/g tissue)

[c] Water (ml/g tissue)

[d] Number of peaks predicted by the mixture design method

[e] Coefficient of variation predicted by the mixture design method

[f] Number of peaks validated by five independent experiments

[g] Coefficient of variation validated by five independent experiments

[h] Number of identified metabolites

[i] Percentage of identified peaks in total peaks

Metabolite identification and quantification

To assign metabolic signals, the HMDB and Chenomx NMR Suite were applied for preliminary identification and 2D NMR for further validation. As listed in ESM Table S1, a total of 68 metabolites were identified in tissue samples, and the detailed assignments of metabolites were shown in 1D (Fig. 6A, ESM Figs. S1–S5) and 2D (ESM Figs. S6–S11) NMR spectra. In this study, we identified 44, 45, 45, 49, 31, and 37 metabolites from brain, heart, liver, kidney, muscle, and pancreas tissues in

healthy rats, which account for 87.99, 91.99, 89.10, 85.85, 90.77, and 91.54% of total NMR signals in corresponding tissue samples, respectively (Table 3).

To further evaluate the quality of metabolite quantification, the main metabolites were quantified from NMR spectra and illustrated as heatmaps in Fig. 6B for brain, ESM Fig. S12A for heart, Fig. S12B for liver, Fig. S12C for kidney, Fig. S12D for muscle, and Fig. S12E for pancreas. According to Fig. 6B, the concentrations of brain metabolites extracted by E1, E2, E5, and optimal conditions were higher than other extraction
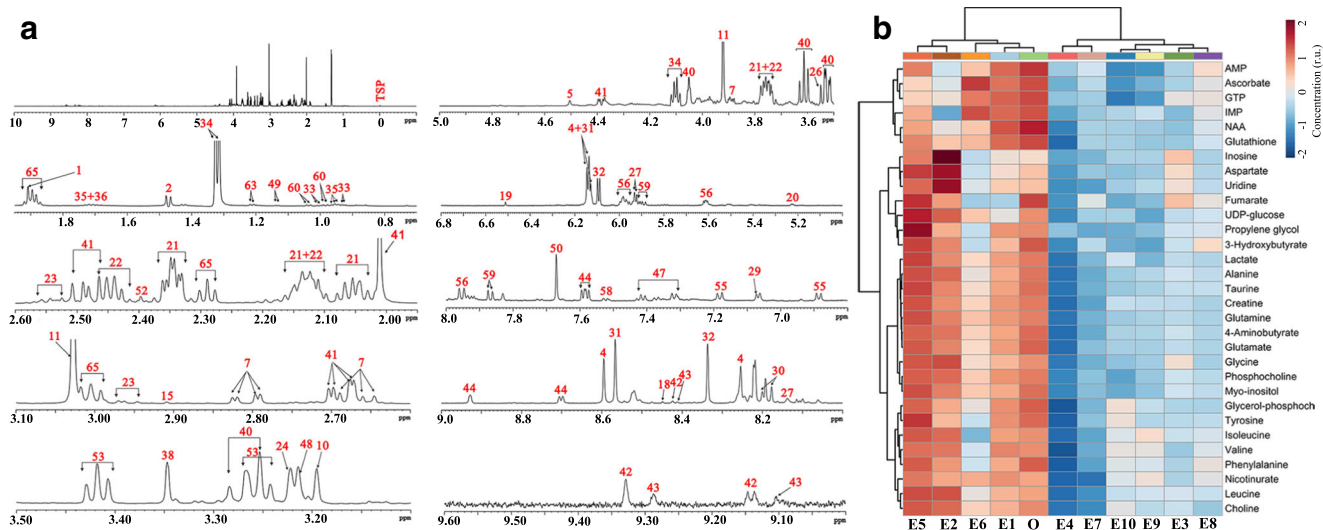

Fig. 6 Metabolite identification and quantification in NMR-based metabolic profiling of brain tissue sample. **A** Annotations of metabolites; the numbers correspond to the metabolites in ESM Table S1. **B** Heatmap showing the concentrations of main metabolites in brain tissue under different MCW extraction conditions. Extraction condition (M/C/W, ml/g tissue): E1, 2:2:8; E2, 2:4:6; E3, 2:6:4; E4, 2:8:8; E5, 4:2:6; E6, 4:4:4; E7, 4:6:2; E8, 6:2:4; E9, 6:4:2; E10, 8:2:2; O (optimal condition), 2:2:8

methods. For heart tissue, E1 and optimal conditions had higher concentrations of metabolites relative to other extraction methods (ESM Fig. S12A). ESM Fig. S12B reveals that most metabolites concentrations in liver tissue were higher under E1, E2, E5, E6, and optimal conditions. The optimal extraction condition exhibited relatively higher concentrations of most metabolites in kidney tissue as compared with other conditions (ESM Fig. S12C). In addition, higher concentrations of most metabolites were extracted from muscle tissue under E1, E2, E5, E6, E8, E9, and optimal conditions as well as from pancreas tissue under E1, E2, E5, E6, and optimal conditions, as shown in ESM Fig. S12D and Fig. S12E, respectively. Taken together, the optimal extraction condition can extract metabolites at higher concentration in all tissue samples used in this study, although it is not a unique optimal condition.

Conclusions

In this study, we found that different MCW extractions obviously affected the characteristics of metabolite profiles in various tissue samples. Moreover, MCW method had tissue-specific differences in the number of peaks (NP) and coefficient of variation (CV), which reflect metabolome coverage and reproducibility, respectively. We then used mixture design methods to optimize the MCW extraction strategy by maximizing the NP value and minimizing the CV value for each tissue sample. The validation results demonstrate that our strategy can improve metabolome coverage and reproducibility. Therefore, different tissue samples need a specific optimal extraction condition, and the mixture design method can be used as an effective tool to optimize extraction of tissue sample for untargeted metabolic profiling. Several perspectives based on the present study can be proposed for further research: (1) the mixture design method can also be used to optimize lipid profiling in organic phase of the MCW extraction for lipidomics; (2) this method is not limited to tissue samples, but also for biofluids; (3) since dichloromethane has a relatively lower toxicity than chloroform, it is recommended to use dichloromethane as the extraction solvent.

Compliance with ethical standards

References

1. Patti GJ, Yanes O, Siuzdak G. Innovation: metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol. 2012;13(4):263–9.
2. Doerr A. Global metabolomics. Nat Methods. 2017;14:32.
3. Mushtaq MY, Choi YH, Verpoorte R, Wilson EG. Extraction for metabolomics: access to the metabolome. Phytochem Anal. 2014;25(4):291–306.
4. Choi YH, Verpoorte R. Metabolomics: what you see is what you extract. Phytochem Anal. 2014;25(4):289–90.
5. Kim S, Lee DY, Wohlgemuth G, Park HS, Fiehn O, Kim KH. Evaluation and optimization of metabolome sample preparation methods for *Saccharomyces cerevisiae*. Anal Chem. 2013;85(4): 2169–76.
6. Tulipani S, Llorach R, Urpi-Sarda M, Andres-Lacueva C. Comparative analysis of sample preparation methods to handle the complexity of the blood fluid metabolome: when less is more. Anal Chem. 2012;85(1):341–8.
7. Sitnikov DG, Monnin CS, Vuckovic D. Systematic assessment of seven solvent and solid-phase extraction methods for metabolomics analysis of human plasma by LC-MS. Sci Rep. 2016;6:38885.
8. García-Cañaveras JC, López S, Castell JV, Donato MT, Lahoz A. Extending metabolome coverage for untargeted metabolite profiling of adherent cultured hepatic cells. Anal Bioanal Chem. 2016;408(4):1217–30.
9. Ibáñez C, Simó C, Palazoglu M, Cifuentes A. GC-MS based metabolomics of colon cancer cells using different extraction solvents. Anal Chim Acta. 2017;986:48–56.
10. Römisch-Margl W, Prehn C, Bogumil R, Röhring C, Suhre K, Adamski J. Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. Metabolomics. 2012;8(1):133–42.
11. Want EJ, Masson P, Michopoulos F, Wilson ID, Theodoridis G, Plumb RS, et al. Global metabolic profiling of animal and human tissues via UPLC-MS. Nat Protoc. 2013;8(1):17–32.
12. Wang H, Xu J, Chen Y, Zhang R, He J, Wang Z, et al. Optimization and evaluation strategy of esophageal tissue preparation protocols for metabolomics by LC–MS. Anal Chem. 2016;88(7):3459–64.
13. Naz S, García A, Barbas C. Multiplatform analytical methodology for metabolic fingerprinting of lung tissue. Anal Chem. 2013;85(22):10941–8.
14. Diémé B, Lefèvre A, Nadal-Desbarats L, Galineau L, Hounoum BM, Montigny F, et al. Workflow methodology for rat brain metabolome exploration using NMR, LC-MS and GC-MS analytical platforms. J Pharm Biomed Anal. 2017;142:270–8.
15. Wu H, Southam AD, Hines A, Viant MR. High-throughput tissue extraction protocol for NMR-and MS-based metabolomics. Anal Biochem. 2008;372(2):204–12.
16. Bligh EG, Dyer WJ. A rapid method of total lipid extraction and purification. Can J Biochem Physiol. 1959;37(8):911–7.
17. Folch J, Lees M, Sloane Stanley GH. A simple method for the isolation and purification of total lipids from animal tissues. J Biol Chem. 1957;226(1):497–509.

18. Zheng H, Clausen MR, Dalsgaard TK, Mortensen G, Bertram HC. Time-saving design of experiment protocol for optimization of LC-MS data processing in metabolomic approaches. Anal Chem. 2013;85(15):7109–16.

19. Montgomery DC. Design and analysis of experiments. John Wiley & Sons; 2017.

20. Savorani F, Tomasi G, Engelsen SB. icoshift: a versatile tool for the rapid alignment of ^{1}D NMR spectra. J Magn Reson. 2010;202:190–202.

21. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0-the human metabolome database in 2013. Nucl Acids Res. 2012;41(D1):801–7.

22. Xia J, Sinelnikov IV, Han B, Wishart DS. (2015). MetaboAnalyst 3.0-making metabolomics more meaningful. Nucl Acids Res. 2015;43:251–7.

23. Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. Cell. 2008;134(5):714–7.

24. Vinayavekhin N, Homan EA, Saghatelian A. Exploring disease through metabolomics. ACS Chem Biol. 2009;5(1):91–103.

25. Suhre K. Metabolic profiling in diabetes. J Endocrinol. 2014;221(3):75–85.

26. Lindahl A, Sääf S, Lehtiö J, Nordström A. Tuning metabolome coverage in reversed phase LC–MS metabolomics of MeOH extracted samples using the reconstitution solvent composition. Anal Chem. 2017;89(14):7356–64.

27. Contrepois K, Jiang L, Snyder M. Optimized analytical procedures for the untargeted metabolomic profiling of human urine and plasma by combining hydrophilic interaction (HILIC) and reverse-phase liquid chromatography (RPLC)–mass spectrometry. Mol Cell Proteomics. 2015;14(6):1684–95.

28. Yang W, Chen Y, Xi C, Zhang R, Song Y, Zhan Q, et al. Liquid chromatography–tandem mass spectrometry-based plasma metabonomics delineate the effect of metabolites' stability on reliability of potential biomarkers. Anal Chem. 2013;85(5):2606–10.

29. Eriksson L, Johansson E, Wikström C. Mixture design-design generation, PLS analysis, and model usage. Chemom Intell Lab Syst. 1998;43(1–2):1–24.

30. Lin CY, Wu H, Tjeerdema RS, Viant MR. Evaluation of metabolite extraction strategies from tissue samples using NMR metabolomics. Metabolomics. 2007;3(1):55–67.

31. Robert O, Sabatier J, Desoubzdanne D, Lalande J, Balayssac S. Gilard V, et al. pH optimization for a reliable quantification of brain tumor cell and tissue extracts with ^{1}H NMR: focus on choline-containing compounds and taurine. Anal Bioanal Chem. 2011;399(2):987–99.

32. Anwar MA, Vorkas PA, Li JV, Shalhoub J, Want EJ, Davies AH, et al. Optimization of metabolite extraction of human vein tissue for ultra performance liquid chromatography-mass spectrometry and nuclear magnetic resonance-based untargeted metabolic profiling. Analyst. 2015;140(22):7586–97.