analytical chemistry

Article

Development of High-Performance Chemical Isotope Labeling LC-MS for Profiling the Carbonyl Submetabolome

Shuang Zhao, Margot Dawe, Kevin Guo, and Liang Li

Anal. Chem., **Just Accepted Manuscript** • Publication Date (Web): 15 May 2017

ACS Publications

Analytical Chemistry is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

(Manuscript submitted to *Anal Chem* on March 24, 2017; Revised on May 8, 2017)

Development of High-Performance Chemical Isotope Labeling LC-MS for Profiling the Carbonyl Submetabolome

Shuang Zhao[†], Margot Dawe[†], Kevin Guo and Liang Li*

Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada

[†]Equal contribution. *Corresponding author. E-mail: Liang.Li@ualberta.ca.

Abstract

Metabolites containing a carbonyl group represent several important classes of molecules including various forms of ketones and aldehydes such as steroids and sugars. We report a high-performance chemical isotope labeling (CIL) LC-MS method for profiling the carbonyl submetabolome with high coverage and high accuracy and precision of relative quantification. This method is based on the use of dansylhydrazine (DnsHz) labeling of carbonyl metabolites to change their chemical and physical properties to such an extent that the labeled metabolites can be efficiently separated by reversed phase LC and ionized by electrospray ionization MS. In the analysis of six standards representing different carbonyl classes, acetaldehyde could be ionized only after labeling and MS signals were significantly increased for other 5 standards with an enhancement factor ranging from ~15-fold for androsterone to ~940-fold for 2-butanone. Differential ^{12}C- and ^{13}C-DnsHz labeling was developed for quantifying metabolic differences in comparative samples where individual samples were separately labeled with ^{12}C-labeling and spiked with a ^{13}C-labeled pooled sample, followed by LC-MS analysis, peak pair picking and peak intensity ratio measurement. In the replicate analysis of a 1:1 ^{12}C-/^{13}C-labeled human urine mixture (n=6), an average of 2030±39 pairs per run were detected with 1737 pairs in common, indicating the possibility of detecting a large number of carbonyl metabolites as well as high reproducibility of peak pair detection. The average RSD of the peak pair ratios was 7.6% and 95.6% of the pairs had a RSD value of less than 20%, demonstrating high precision for peak ratio measurement. In addition, the ratios of most peak pairs were close to the expected value of 1.0 (e.g., 95.5% of them had ratios of between 0.67 and 1.5), showing the high accuracy of the method. For metabolite identification, a library of DnsHz-labeled standards was constructed, including 78 carbonyl metabolites with each containing MS, retention time (RT) and MS/MS information. This library and an online search program for labeled carbonyl metabolite identification based on MS, RT and MS/MS matches have been implemented in a freely available website, www.mycompoundid.org. Using this library, out of the 1737 peak pairs detected in urine, 33 metabolites were positively identified. In addition, 1333 peak pairs could be matched to the metabolome databases with most of them belonging to the carbonyl metabolites. These results show that ^{12}C-/^{13}C-DnsHz labeling LC-MS is a useful tool for profiling the carbonyl submetabolome of complex samples with high coverage.

Introduction

Metabolomics has been proved to be very useful in a number of research areas ranging from disease biomarker discovery to systems biology studies.[1] In metabolomics, comparative metabolomic analysis is carried out to determine the metabolic differences between two or more groups of samples. To do this, high-coverage profiling with accurate relative quantification of individual metabolites is crucial. However, profiling complex samples (e.g., bio-fluids, cells and tissues) is a huge analytical challenge, as biological samples contain thousands of metabolites with diverse structures and physical/chemical properties and with a very wide range of concentrations (e.g., pM to mM).[2]

To increase coverage and/or improve quantification accuracy and precision, many enabling methods and strategies, mainly based on NMR and mass spectrometry (MS), have been reported.[3-5] A common approach is to group the metabolites according to their physical properties (e.g., solubility and hydrophobicity) and then analyze them using conditions optimized for each group (e.g., reversed phase (RP) LC for separating hydrophobic metabolites and hydrophilic interaction LC for separating more polar compounds).[5] An alternative approach is to combine high-performance chemical isotope labeling (CIL) of metabolites with liquid chromatography-mass spectrometry (LC-MS) for metabolomic analysis.[6] In this approach, the whole metabolome is divided into several chemical-group-based submetabolomes, followed by CIL LC-MS analysis of each submetabolome. Judging from the structures of more than 8000 endogenous human metabolites in the Human Metabolome Database (HMDB),[7] more than 95% of these metabolites contain one or more of the four functional groups: amine, carboxyl, hydroxyl and carbonyl. Thus, if these four groups of submetabolomes can be separately analyzed with complete coverage, the combined result of the four submetabolomic profiles would represent over 95% of the whole metabolome.

There are a number of labeling chemistries that have been reported for targeted and untargeted analysis of different groups of metabolites with varying degrees of performance.[8-25] To detect as many metabolites as possible, a rationally designed labeling reagent can be used to react with the common functional group of a submetabolome (e.g., the amine group in all amine-containing metabolites) to alter the chemical and physical properties of the metabolites to an extent that the labeled metabolites can be efficiently separated by RPLC and effectively ionized in electrospray ionization (ESI) in a positive ion mode.[6, 26, 27] In this way, the LC-MS setup does not need to be changed in the metabolomic analysis workflow. However, this chemical-group-based submetabolome profiling approach requires the use of robust and effective labeling chemistries targeting the individual chemical groups.

In previous studies, three high-performance labeling methods have been reported for submetabolome profiling: ^{12}C-/^{13}C-dansyl chloride (DnsCl) labeling for the amine submetabolome,[6] ^{12}C-/^{13}C-dimethylaminophenacyl (DmPA) bromide labeling for the carboxyl submetabolome,[28] and ^{12}C-/^{13}C-DnsCl with base-activation for the hydroxyl submetabolome.[29] Each labeling allows the profiling of thousands of metabolites within a group with high accuracy and precision for relative quantification of metabolomic changes in bio-samples. However, to our knowledge, there is no available method capable of profiling the carbonyl submetabolome with very high coverage (i.e., detecting thousands of metabolites). Metabolites containing carbonyl group (ketones and aldehydes) include several important classes of molecules that often have low ionization efficiency (e.g., sugars) and/or are present at very low concentrations (e.g., steroids in plasma).[2] GC-MS methods have been used for analyzing ketones and aldehydes,[30-32] while a number of derivatization chemistries[33-40] including dansylhydrazine labeling[41-46] have been reported for labeling ketones/aldehydes to improve MS detectability; however, these methods were often used for the analysis of a limited number of analytes. In this report, we present a method using ^{12}C-/^{13}C-dansylhydrazine labeling for analyzing the carbonyl submetabolome with high coverage and high accuracy and precision of relative quantification. Optimal labeling conditions tailored for high-coverage profiling work are described, followed by the demonstration of the analytical performance of the workflow.

Experimental Section

Principle and Workflow. The method presented herein complements to the three previously reported labeling methods for relative quantification of chemical-group-based-submetabolomes. In order to generate accurate and precise relative quantification results, differential chemical isotope labeling (CIL) is used for untargeted profiling of the carbonyl submetabolome. Figure 1 shows the schematic of the workflow, including the labeling reaction scheme using the ^{12}C- or ^{13}C-dansylhydrazine (DnsHz) reagent. In this workflow tailored to perform comparative metabolomics of two or more groups of samples, the individual samples are separately labeled by ^{12}C-DnsHz (light labeling), while a pooled sample prepared from mixing aliquots of individual samples is labeled by ^{13}C-DnsHz (heavy labeling). The ^{13}C-labeled pool serves as an internal standard or control for all the ^{12}C-labeled individual samples; the same amount of the ^{13}C-labeled pool is spiked into each ^{12}C-labeled individual sample to form a ^{12}C-/^{13}C-mixture for LC-MS analysis. All the labeled metabolites are detected as peak pairs in the mass spectra, making them easily distinguished from the singlet peaks originated from noise or unlabeled chemical backgrounds introduced in LC-MS. Using IsoMS software (see below), peak pairs are picked and their intensity ratios are calculated. The intensity of a ^{12}C-labeled peak from a ^{12}C-labeled sample is measured with reference to that of the ^{13}C-labeled peak of the same metabolite from a ^{13}C-labeled pool. Since the same pool is added to all individual samples, the intensity ratio values determined from LC-MS analyses of the prepared ^{12}C-/^{13}C-mixtures are related to the concentrations of a given metabolite in these individual samples. Thus, the peak intensity ratios can be used for relative quantification of metabolites in different samples. While it is not shown in this work, absolute quantification of a metabolite can also be performed by adding a known concentration of ^{13}C-labeled standard to a ^{12}C-labeled sample. To identify the metabolites, a carbonyl standard library has been constructed (see below).

Experimental conditions used for some steps are described below and Supplemental Note S1 provides information on chemicals and reagents, urine sample preparation and LC-MS setups.

Synthesis of Dansylhydrazine. Dansylhydrazine (DnsHz) was prepared by treating dansyl chloride with hydrazine following a literature procedure with some modification (see Supplemental Figure S1A and Note S1).[47, 48]

Solution Preparation. The 78 standards of carbonyl metabolites (Supplemental Table S1) were obtained from the Human Metabolome Database (HMDB) compound library at the University of Alberta. Each compound was dissolved in methanol/water mixture to a concentration of 5 mM. If the solubility of a compound was too low, the supernatant of a saturated solution was used. All the prepared solutions were stored in -20 ºC. The labeling reagent, DnsHz, was dissolved in methanol. Hydrochloric acid (HCl) was prepared by diluting a commercial 12 M HCl solution to 144 mM with methanol.

Dansylhydrazine Labeling Reaction. For standard labeling, 20 µL of a standard solution were mixed with 20 µL of 144 mM HCl solution and 20 µL of 20 mM ^{12}C-DnsHz. After vortexing and spinning down, the mixture was incubated at 40 ºC for 60 min. Then the mixture was cooled down in -80 ºC freezer for 10 min to stop the labeling reaction. After that, the cooled solution was dried down using SpeedVac to remove the acid catalyst, followed by re-dissolving in 80 µL of acetonitrile/water mixture (ACN/H_2O, 50:50, v/v). The solution was diluted and centrifuged at 15294 *g* for 10 min before injecting into LC-MS for analysis.

To prepare the 1:1 ^{12}C-/^{13}C-labeled urine sample for method development and analytical performance evaluation, 40 µL of human urine were divided into two parts: 20 µL was labeled by ^{12}C-DnsHz and the other 20 µL was labeled by ^{13}C-DnsHz following the same procedure described above for standard labeling. After labeling and re-dissolving, the ^{12}C-labeled urine sample was mixed with the same volume of the ^{13}C-labeled urine. The mixture was injected into LC-MS for analysis.

LC-MS. The LC-QTOF-MS setup and running conditions used are provided in Supplemental Note S1.

Data Processing and Analysis. A set of software programs developed in house as reported previously and also freely available from www.mycompoundid.org were used to process the CIL LC-MS data. In the 1st step, IsoMS[49] was used to extract peak pairs from mass spectra, including filtering out the redundant pairs (e.g., those of adduct ions and dimers). In this way, only a protonated ion of a peak pair was retained for a ^{12}C-/^{13}C-labeled metabolite. IsoMS also calculated the intensity ratio of each peak pair, and aligned multiple pairs of the same individual metabolites from different runs according to retention time and accurate mass. In the 2nd step, the missing values of intensity ratios in the aligned file were filled by the Zerofill program.[50] Finally, the chromatographic peak ratio of each peak pair was determined by IsoMS-Quant[51] and used to produce the final metabolite-intensity table for relative quantification.

Positive identification of the labeled metabolites was done using the newly constructed DnsHz standard library (see below). For putative metabolite identification, accurate masses of peak pairs were searched against those of the metabolite entries in metabolome databases by using MyCompoundID (MCID) at www.mycompoundid.org with zero-reaction and one-reaction libraries.[52]

Results and Discussion

Standard-Mixture Labeling. In order to optimize the DnsHz labeling conditions, six aldehyde/ketone standards with very different structures (Supplemental Figure S1B) were mixed and labeled together under different conditions with ^{12}C-DnsHz. These standards were chosen to represent several important classes of carbonyl metabolites: acetaldehyde and 2-butanone for low-molecular-weight aldehydes and ketones, cyclohexanone for cyclic ketones, benzaldehyde for aromatic carbonyl compounds, androsterone for steroids and galactose for carbohydrates or sugars. The ^{12}C-labeled standard-mixture was mixed with a ^{13}C-labeled standard-mixture that served as an internal standard, followed by LC-MS analysis. Six experimental replicate analyses were performed for each condition. The peak area ratio of a ^{12}C-/^{13}C-labeled standard-mixture was used to evaluate the efficiencies of different labeling conditions (Figure 2A-E). The relative standard deviation (RSD) of the peak area ratio (n=6) was used as the indicator of reproducibility (Figure 2F-J). Using this approach, the optimal conditions for reaction solvents, acids, time and temperature were screened.

Figure 2A shows the effect of solvent type on labeling efficiency. For all six standards, the labeling efficiency was similar in the tested solvents, namely acetonitrile, methanol, ethanol and methanol/water (1:1, v/v). In theory, in the formation of hydrazones, decreasing water (also a product of the labeling reaction) would promote the forward reaction. However, our results indicated that the presence of water in the reaction solution did not adversely affect the labeling efficiency. This is good for practical applications as water is often present in biological samples and direct labeling of water-containing samples can simplify the experimental procedure (i.e., no need to remove water). As Figure 2F-J shows, the RSD of peak ratios calculated from repeated labeling experiments was generally less than 10% when acetonitrile, methanol and methanol/water were used, showing good labeling reproducibility. In the case of ethanol, the RSD was up to 12%.

In addition to the selection of a solvent, acid was added to serve as a catalyst which also plays an important role in the labeling reaction. Acid or proton makes the carbonyl group more electrophilic to aid the attack by the nucleophilic $-NH_2$ group in dansylhydrazine and also helps the elimination of water from the intermediate to form the final product. The results from acid concentration screening confirmed that the reaction needs protons to catalyze the reaction; 0.5% trifluoroacetic acid (TFA) was found to be optimal in terms of both efficiency (Figure 2B) and reproducibility (Figure 2G). Within the similar pH range, HCl and TFA shared similar labeling efficiency (Figure 2C) and reproducibility (Figure 2H). For several standards in the mixture, formic acid (FA) was not a good acid for the labeling and the RSD was relatively larger than that of HCl and TFA. The other two parameters affecting the reaction process are reaction temperature and time. The optimal conditions were found to be 40 ºC for 1 h (Figure 2D and 2E).

The benefit of using DnsHz labeling for MS detection is shown in Supplemental Table S2 where the signals (i.e., the areas of chromatographic peaks) of unlabeled and labeled standards from injecting the same amount of each compound are compared. For acetaldehyde, the unlabeled compound could not be detected, but the labeled compound could be readily analyzed.

For the other 5 standards, signals were significantly increased by DnsHz labeling. The signal enhancement factor ranged from ~15-fold increase for androsterone to ~940-fold for 2-butanone.

In this work, we did not determine the absolute efficiency of DnsHz labeling for individual compounds. Derivatization efficiency is very much compound-dependent. It is also dependent on the compound concentration and sample matrix. Fortunately, the information on derivatization efficiency of a given metabolite is not needed for relative quantification of the metabolite in comparative samples. For relative quantification using a pooled sample as the control, derivatization efficiency should be very similar for labeling a metabolite in an individual sample vs. labeling the same metabolite in the pooled sample as this metabolite has similar concentrations and matrices in these samples. As long as the derivatization efficiency is constant, the peak intensity ratio of the ^{12}C-labeled metabolite in a sample vs. the ^{13}C-labeled metabolite in the control should reflect the concentration difference and thus accurate relative quantification can be obtained (see below for urine sample labeling results).

Urine Sample Labeling. While the six standards used for optimizing the experimental conditions represent different classes of carbonyl metabolites, a real biological sample contains many more different compounds with different concentrations. To determine whether the optimized condition is applicable to analyze complex samples, human urine was chosen to test the labeling method. Urine was aliquoted to two parts. One was labeled by ^{12}C-DnsHz and the other aliquot was labeled by ^{13}C-DnsHz using the same condition. Two labeled samples were then mixed by equal volume to form a 1:1 ^{12}C-/^{13}C-labeled urine sample for LC-MS analysis. The number of peak pairs detected was used to gauge the performance of a tested experimental condition. For each condition, six LC-MS runs were performed from experimental triplicates of sample processing and injection duplicate of each processed sample.

Figure 3A shows the study of the effect of water in a urine sample on labeling. Two methods were applied to remove water from urine before labeling. One was to dry down the urine and then re-dissolve it with methanol, followed by labeling. The other way was to use ethyl acetate to extract the urine metabolites, followed by drying, re-dissolving and labeling the extract. Figure 3A clearly shows that direct labeling of urine generated a much higher number of peak pairs. The differences can be attributed to the metabolite loss in the two drying methods (e.g., loss of volatile compounds during solvent evaporation and loss of low-abundance metabolites from adsorption to the vials). The results shown in Figure 3A also confirmed that the existence of water in the labeling condition was acceptable.

Figure 3B shows the effect of TFA concentration in the urine labeling solution on the number of peak pairs detected. As in the case of the analysis of the standard mixture, 0.5% TFA gave the optimal result, i.e., the highest peak pair number in this case. There were no significant differences between labeling with TFA (0.5% v/v) and HCl (48 mM) at the same pH value (Figure 3C). However, if TFA was used as the acid catalyst, a large peak from the labeled TFA products was observed in the ion chromatogram. For targeted analysis of a few metabolites, this peak may not interfere with the analysis of the metabolites of interest. However, for a real sample metabolomic profiling work, due to the presence of so many metabolites, the labeled TFA peak can interfere with the detection of co-eluting metabolites at the retention window of the labeled TFA (from 12.6 to 13.3 min). Thus, 48 mM HCl was decided to be the optimal catalyst for labeling complex samples such as urine.

The results of reaction time and temperature screening were also matched with those of labeling the standard mixture (Figure 3D and 3E). Thus, incubating at 40 ºC for 1 h was chosen as the optimal condition for labeling complex samples. For labeling a complex sample, using a sufficient amount of reagent is also important to drive the labeling reaction to yield the maximum amounts of products for all the metabolites. To determine how the reagent amounts affect the labeling, different amounts of DnsHz reagent were added to a urine sample to contain 5 mM, 10 mM, or 20 mM dansylhydrazine. The results shown in Figure 3F indicate that the use of 10 or 20 mM DnsHz reached the optimal number of peak pairs detected.

Taken together from the results obtained from the analyses of the standard mixture and human urine, we determined that the optimal conditions for DnsHz labeling for profiling the carbonyl submetabolome should include the use of 48 mM HCl as a catalyst and incubation of labeling solution at 40 ºC for 1 h. Water in a sample is acceptable. For urine sample analysis, 20 mM dansylhydrazine in the labeling solution is recommended. For other type of biological samples, the optimal reagent amount to be used for mixing with a sample should be optimized to ensure that the labeling reaction is driven to a maximal extent of completion.

Accuracy and Precision. As Figure 1 shows, the entire workflow of DnsHz labeling LC-MS involves several steps with each step potentially affecting the accuracy and precision of the method. We evaluated the performance of our method using the ^{12}C-/^{13}C-labeled urine samples. In the intra-day comparison experiments, urine was labeled three times and analyzed by LC-MS with duplicate for each sample in one day. Figure 4A shows the Venn diagram of the peak pair numbers obtained from triplicate labeling. An average of 2030±39 pairs per sample (n=6) were detected with 1737 pairs in common, indicating high reproducibility of peak pair detection. Figure 4B-C shows the distribution of peak pair numbers as a function of the averaged peak area ratio and the RSD of the ratio (n=6). The ratios of most peak pairs were close to the expected value of 1.0. For example, 95.5% of the pairs had values of between 0.67 and 1.5, i.e., within ±50% of the expected value, indicating that high accuracy could be achieved. The average RSD of the entire data set was 7.6% and 95.6% of the peak pairs had a RSD value of less than 20%, demonstrating that high precision could also be achieved.

Similar evaluation was also conducted with inter-day experiments (Figure 4D-F). Urine was labeled three times in different days. The Venn diagram of the peak pair numbers obtained from the experimental triplicate with injection duplicate also shows high overlap of the peak pairs detected (Figure 4D). A total of 1800 peak pairs were detected in all runs with an average of 2121±16 peak pairs detected per run (n=6). The results of peak pair number distributions (Figure 4E,F) were in agreement with those of the intra-day experiments.

In this work, we did not determine the linear range for relative quantification; we will do this in future work for real world applications where a wide range of metabolite concentration changes among comparative samples are expected.

Product Stability. The stability of labeled samples was also tested since analyzing a large batch or multiple batches of samples may take days or a month to complete in metabolomics. In this work, a ^{12}C-/^{13}C-labeled urine sample mixture was aliquoted and stored under different conditions for different periods of time, including 4 ºC for overnight, -20 ºC for a week, -20 ºC for a month, and -80 ºC for a month. The results of the analyses of these samples are shown in Supplemental Figure S2. As Figure S2 shows, there is no significant difference of the peak pair numbers detected among these storage conditions. Thus, the labeled samples are very stable and can meet the need of month-long storage for completing a batch analysis of many samples in a large-scale metabolomic profiling project (e.g., with the current setup, a monthly throughput is 30 days × 32 samples/day = 960 samples per month).

Dansylhydrazine-labeled Standard Library. The final step of the workflow shown in Figure 1 is metabolite identification. Positive or definitive identification requires the use of an authentic standard as a reference for comparing the measured parameters with those of an unknown metabolite. To this end, we have constructed a DnsHz-labeled standard library that currently consists of 78 endogenous human metabolites (Supplemental Table S1). This library is expandable in the future by adding more standards when they become available. To construct the library, each standard was individually labeled and analyzed using LC-MS and MS/MS to generate "triple-parameters" information: accurate molecular mass, retention time (RT) and MS/MS fragmentation spectrum. This library, along with a database search program (DnsHz-ID), is freely available from a publicly accessible website at www.mycompoundid.org. A tutorial for the use of this resource for metabolite identification based on MS, RT and MS/MS matches is shown in Supplemental Note S2. An example of metabolite identification is given in Supplemental Note S3. These notes are also provided on the website.

It should be noted that retention time is not often used as a match parameter, as its value can be easily affected by a number of experimental conditions used for LC-MS, including minor variations in LC equipment used, connecting tube length from LC to MS, solvent composition for running LC, temperature, etc. To facilitate the comparison of user-generated RT values with those of labeled standards in the library, a RT normalization method was used to correct the RT shift due to differences in instrument setups and running conditions between a user's system and our library system.[53] In this approach, all RT values in the library were normalized to a set of RT calibrants containing 17 DnsHz-labeled standards (Supplemental Table S3). The ion chromatogram of this mixture is shown in Figure 5A. The RT difference between two adjacent peaks is less than 2 min. A user needs to run the same mixture of RT calibrants and then upload the LC-MS data to the DnsHz-ID program on the website for RT correction. A built-in program in DnsHz-ID compares the RT values of the calibrants in the user's file with those of the library calibrants and determines the correction parameters. These parameters are then applied, automatically, to the real sample LC-MS data to produce a corrected RT value for each unknown metabolite detected in the sample. These corrected RT values are compared to the library RT values of the standards for matching. The RT calibrants are available upon request from the authors and can also be made by the user.

Urine Carbonyl Submetabolome Profiling. To demonstrate the applicability and performance of the developed method for profiling complex samples, the urinary carbonyl submetabolome was analyzed. Figure 5B shows a base-peak ion chromatogram of ^{12}C-/^{13}C-labeled urine. Many peaks can be detected, indicating the presence of various carbonyl compounds in urine. To detect as many peak pairs as possible, sample injection amount for LC-MS should be optimized. In our work, we normally generate a plot of the peak pair number as a function of injection volume (see Figure 5C). A plateau was achieved when the injection volume was 10 µL for urine carbonyl submetabolome profiling. For analyzing other type of samples, a similar approach can be used to find the optimal injection amount for LC-MS.

As discussed earlier, replicate analyses of the ^{12}C-/^{13}C-labeled urine resulted in the detection of 1737 common peak pairs in all 6 runs. These peak pairs were identified using three steps at different levels of confidence. In the first step, highly confident and positive metabolite identification was done using the DnsHz labeled standard library. 33 peak pairs could be identified based on both accurate mass and retention time matches with those of the library standards (Supplemental Table S4). It is clear that only a very small fraction of the detected peak pairs were positively identified. Increasing the number of standards in the library is needed. However, the availability of standards is currently limited. In the future, we plan to use chemical synthesis or biologically derived compounds to expand the library.

In the 2nd step of metabolite identification, the remaining peak pairs were matched to some metabolites by searching accurate masses of peak pairs against the MyCompoundID (MCID) zero-reaction library containing 8,021 known human endogenous metabolites (i.e., zero-reaction library). A total of 406 peak pairs were matched to the zero-reaction library (Supplemental Table S5). The structures of matched metabolites were manually checked. Over 96.3% of these peak pairs were matched to metabolites containing C=O moiety. In the last step, the remaining unmatched peak pairs were searched against the MCID one-reaction library that contains 375,809 predicted compounds after applying one round of metabolic reactions to the 8,021 human endogenous metabolites. A total of 927 peak pairs were matched (Supplemental Table S6). Thus, out of the 1737 peak pairs detected, 78.6% (1366 pairs) could be either identified or matched.

Conclusions

We have developed a high-performance chemical isotope labeling LC-MS method for profiling the carbonyl submetabolome with high coverage and accuracy and precision. This method, used alone, should be useful for profiling important groups of metabolites such as hormones and sugars in complex biological samples. This method, in combination with previously reported three labeling methods for amine, carboxyl and hydroxyl submetabolomes, can potentially be used to profile over 95% of the entire metabolome, based on the grouping of metabolites using the four functional groups. The real world applications of this method for quantitative metabolomics will be reported in the future.

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada, the Canada Research Chairs program, Genome Canada and Alberta Innovates. We thank Dr. Taibo Luo, Mr. Hao Li and Professor Guohui Lin for their assistance on MCID website update for the DnsHz library.

Disclosure

The authors declare no competing financial interest.

Supporting Information

The Supporting Information available:

- Supplemental Note S1 for experimental, Notes S2-S3 for tutorial and search example, Figures S1-S2 for reaction, structures, and stability study and Tables S1-S6 listing metabolite standards and metabolite identification results.

This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Newgard, C. B. *Cell Metab* **2017**, *25*, 43-56.
(2) *Normal Laboratory Values for Blood, Urine, CSF, Stool, and Other Fluids, Merck Manual Homepage.* http://www.merckmanuals.com *(accessed March 20, 2017).*
(3) Gowda, G. A. N.; Raftery, D. *Anal Chem* **2017**, *89*, 490-510.
(4) Theodoridis, G. A.; Gika, H. G.; Want, E. J.; Wilson, I. D. *Anal Chim Acta* **2012**, *711*, 7-16.
(5) Ortmayr, K.; Causon, T. J.; Hann, S.; Koellensperger, G. *Trac-Trends Anal. Chem.* **2016**, *82*, 358-366.
(6) Guo, K.; Li, L. *Anal Chem* **2009**, *81*, 3919-3932.
(7) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S., et al. *Nucleic Acids Res* **2007**, *35*, D521-D526.
(8) Tang, Z. M.; Guengerich, F. P. *Anal Chem* **2010**, *82*, 7706-7712.
(9) Yang, W. C.; Regnier, F. E.; Jiang, Q.; Adamec, J. *J Chromatogr A* **2010**, *1217*, 667-675.
(10) Yuan, W.; Zhang, J. X.; Li, S. W.; Edwards, J. L. *J Proteome Res* **2011**, *10*, 5242-5250.
(11) Dai, W. D.; Huang, Q.; Yin, P. Y.; Li, J.; Zhou, J.; Kong, H. W.; Zhao, C. X.; Lu, X.; Xu, G. W. *Anal Chem* **2012**, *84*, 10245-10251.

(12) Leng, J. P.; Wang, H. Y.; Zhang, L.; Zhang, J.; Wang, H.; Guo, Y. L. *Anal Chim Acta* **2013**, *758*, 114-121.
(13) Tayyari, F.; Gowda, G. A. N.; Gu, H. W.; Raftery, D. *Anal Chem* **2013**, *85*, 8715-8721.
(14) Yuan, W.; Edwards, J. L.; Li, S. W. *Chem Commun* **2013**, *49*, 11080-11082.
(15) Mochizuki, T.; Todoroki, K.; Inoue, K.; Min, J. Z.; Toyo'oka, T. *Anal Chim Acta* **2014**, *811*, 51-59.
(16) Hao, L.; Zhong, X. F.; Greer, T.; Ye, H.; Li, L. J. *Analyst* **2015**, *140*, 467-475.
(17) Wagner, M.; Ohlund, L. B.; Shiao, T. C.; Vezina, A.; Annabi, B.; Roy, R.; Sleno, L. *Rapid Commun Mass Spectrom* **2015**, *29*, 1632-1640.
(18) Chu, J. M.; Qi, C. B.; Huang, Y. Q.; Jiang, H. P.; Hao, Y. H.; Yuan, B. F.; Feng, Y. Q. *Anal Chem* **2015**, *87*, 7364-7372.
(19) Arrivault, S.; Guenther, M.; Fry, S. C.; Fuenfgeld, M.; Veyel, D.; Mettler-Altmann, T.; Stitt, M.; Lunn, J. E. *Anal Chem* **2015**, *87*, 6896-6904.
(20) Liu, P.; Qi, C. B.; Zhu, Q. F.; Yuan, B. F.; Feng, Y. Q. *Sci Rep* **2016**, *6*, 21433; doi:21410.21038/srep21433.
(21) Wong, J. M. T.; Malec, P. A.; Mabrouk, O. S.; Ro, J.; Dus, M.; Kennedy, R. T. *J Chromatogr A* **2016**, *1446*, 78-90.
(22) Yu, L.; Ding, J.; Wang, Y. L.; Liu, P.; Feng, Y. Q. *Anal Chem* **2016**, *88*, 1286-1293.
(23) Hao, L.; Johnson, J.; Lietz, C. B.; Buchberger, A.; Frost, D.; Kao, W. J.; Li, L. J. *Anal Chem* **2017**, *89*, 1138-1146.
(24) Higashi, T.; Ogawa, S. *J Pharm Biomed Anal* **2016**, *130*, 181-193.
(25) Fan, R. J.; Guan, Q.; Zhang, F.; Leng, J. P.; Sun, T. Q.; Guo, Y. L. *Anal Chim Acta* **2016**, *908*, 132-140.
(26) Zhou, R. K.; Guo, K.; Li, L. *Anal Chem* **2013**, *85*, 11532-11539.
(27) Zhou, R.; Huan, T.; Li, L. *Anal Chim Acta* **2015**, *881*, 107-116.
(28) Guo, K.; Li, L. *Anal Chem* **2010**, *82*, 8789-8793.
(29) Zhao, S.; Luo, X.; Li, L. *Anal Chem* **2016**, *88*, 10617-10623.
(30) Dong, J. Z.; Moldoveanu, S. C. *J Chromatogr A* **2004**, *1027*, 25-35.
(31) Moreira, N.; Meireles, S.; Brandao, T.; de Pinho, P. G. *Talanta* **2013**, *117*, 523-531.
(32) Biladeau, S. K.; Richmond, W. N.; Laulhe, S.; Nantz, M. H. *Analytical Methods* **2016**, *8*, 3704-3710.
(33) Richardson, S. D.; Caughran, T. V.; Poiger, T.; Guo, Y. B.; Crumley, F. G. *Ozone-Sci Eng* **2000**, *22*, 653-675.
(34) Jakober, C. A.; Charles, M. J.; Kleeman, M. J.; Green, P. G. *Anal Chem* **2006**, *78*, 5086-5093.
(35) Yuan, W.; Li, S. W.; Edwards, J. L. *Anal Chem* **2015**, *87*, 7660-7666.
(36) Bereman, M. S.; Comins, D. L.; Muddiman, D. C. *Chem Commun* **2010**, *46*, 237-239.
(37) Eggink, M.; Wijtmans, M.; Ekkebus, R.; Lingeman, H.; de Esch, I. J. P.; Kool, J.; Niessen, W. M. A.; Irth, H. *Anal Chem* **2008**, *80*, 9042-9051.
(38) Mattingly, S. J.; Xu, T.; Nantz, M. H.; Higashi, R. M.; Fan, T. W. M. *Metabolomics* **2012**, *8*, 989-996.
(39) Siegel, D.; Meinema, A. C.; Permentier, H.; Hopfgartner, G.; Bischoff, R. *Anal Chem* **2014**, *86*, 5089-5100.
(40) Flinders, B.; Morrell, J.; Marshall, P. S.; Ranshaw, L. E.; Clench, M. R. *Anal Bioanal Chem* **2015**, *407*, 2085-2094.

(41) Appelblad, P.; Ponten, E.; Jaegfeldt, L. H.; Backstrom, T.; Irgum, K. *Anal Chem* **1997**, *69*, 4905-4911.
(42) Hidalgo, F. J.; Navarro, J. L.; Delgado, R. M.; Zamora, R. *Food Chem* **2013**, *140*, 183-188.
(43) Al-Dirbashi, O. Y.; Rashed, M. S.; Jacob, M.; Al-Ahaideb, L. Y.; Al-Amoudi, M.; Rahbeeni, Z.; Al-Sayed, M. M.; Al-Hassnan, Z.; Al-Owain, M.; Al-Zeidan, H. *Biomed Chromatogr* **2008**, *22*, 1181-1185.
(44) Appelblad, P.; Ahmed, A.; Ponten, E.; Backstrom, T.; Irgum, K. *Anal Chem* **2001**, *73*, 3701-3708.
(45) Herrington, J.; Zhang, L.; Whitaker, D.; Sheldon, L.; Zhang, J. *J Environ Monit* **2005**, *7*, 969-976.
(46) Tomono, S.; Miyoshi, N.; Ohshima, H. *J. Chromatogr. B* **2015**, *988*, 149-156.
(47) Backes, G. L.; Neumann, D. M.; Jursic, B. S. *Bioorg Med Chem* **2014**, *22*, 4629-4636.
(48) Lee, H. J.; Cho, M. J.; Chang, S. K. *Inorg Chem* **2015**, *54*, 8644-8649.
(49) Zhou, R.; Tseng, C. L.; Huan, T.; Li, L. *Anal Chem* **2014**, *86*, 4675-4679.
(50) Huan, T.; Li, L. *Anal Chem* **2015**, *87*, 1306-1313.
(51) Huan, T.; Li, L. *Anal Chem* **2015**, *87*, 7011-7016.
(52) Li, L.; Li, R. H.; Zhou, J. J.; Zuniga, A.; Stanislaus, A. E.; Wu, Y. M.; Huan, T.; Zheng, J. M.; Shi, Y.; Wishart, D. S., et al. *Anal Chem* **2013**, *85*, 3401-3408.
(53) Huan, T.; Wu, Y. M.; Tang, C. Q.; Lin, G. H.; Li, L. *Anal Chem* **2015**, *87*, 9838-9845.

Figure Captions

Figure 1. Workflow of ^{12}C-/^{13}C-dansylhydrazine labeling for relative quantification of the carbonyl submetabolome differences in comparative samples. The dansylhydrazine labeling reaction scheme is also shown.

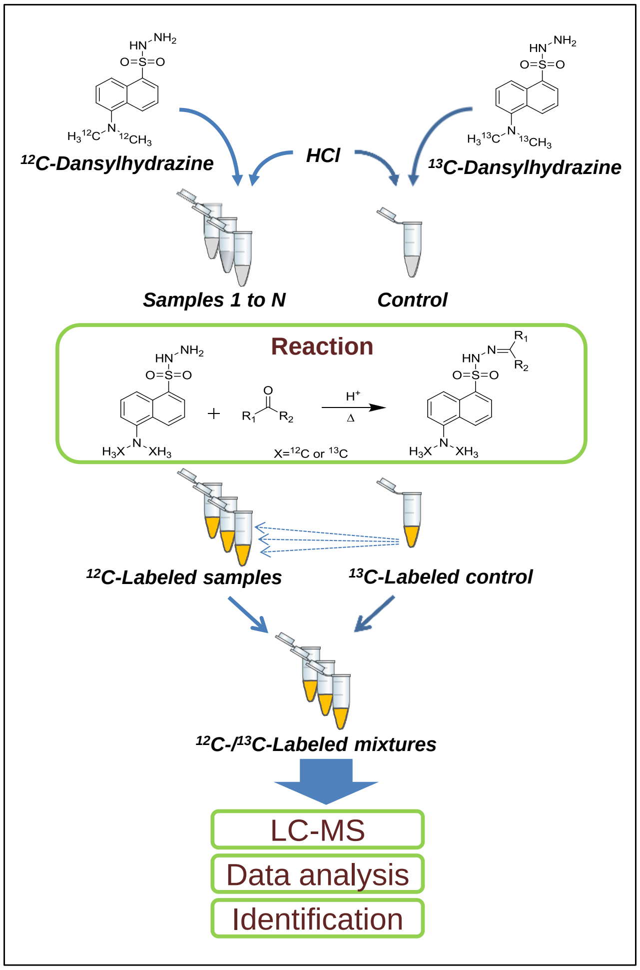
Figure 2. Comparison of efficiency (A-E) and reproducibility (F-J) for labeling a standard-mixture under different reaction conditions: (A and F) effect of solvent type, (B and G) effect of acid concentration, (C and H) effect of acid type, (D and I) effect of reaction temperature, and (E and J) effect of reaction time. Data are presented as the mean of six replicates (n=6).

Figure 3. Peak pair numbers detected from ^{12}C-/^{13}C-dansylhydrazine labeled urine samples prepared under different labeling conditions: (A) effect of sample preparation, (B) effect of acid concentration, (C) effect of acid type, (D) effect of temperature, (E) effect of time, and (F) effect of dansylhydrazine amount. Data are presented as mean ± S.D. from experimental triplicate and injection duplicate (n=6).

Figure 4. Venn diagrams of the peak pair numbers detected in triplicate labeled urine from (A) the intra-day experiments and (D) the inter-day experiments. Distributions of peak pair numbers as a function of averaged peak ratio of (B) intra-day results and (E) inter-day results and RSD of (C) intra-day results and (F) inter-day results. Data are presented as mean ± S.D. from experimental triplicate and injection duplicate (n=6).

Figure 5. (A) LC-MS ion chromatogram of RT calibrants consisting of 17 labeled standards. (B) LC-MS ion chromatogram of ^{12}C-/^{13}C-dansylhydrazine labeled urine. (C) Peak pair number detected as a function of injection amount of ^{12}C-/^{13}C-dansylhydrazine labeled urine. Data are presented as mean ± S.D. (n=3).

Figure 1.



12C-Dansylhydrazine
HCl
13C-Dansylhydrazine
Samples 1 to N
Control
Reaction
X=12C or 13C
12C-Labeled samples
13C-Labeled control
12C-/13C-Labeled mixtures
LC-MS
Data analysis
Identification

Figure 2.

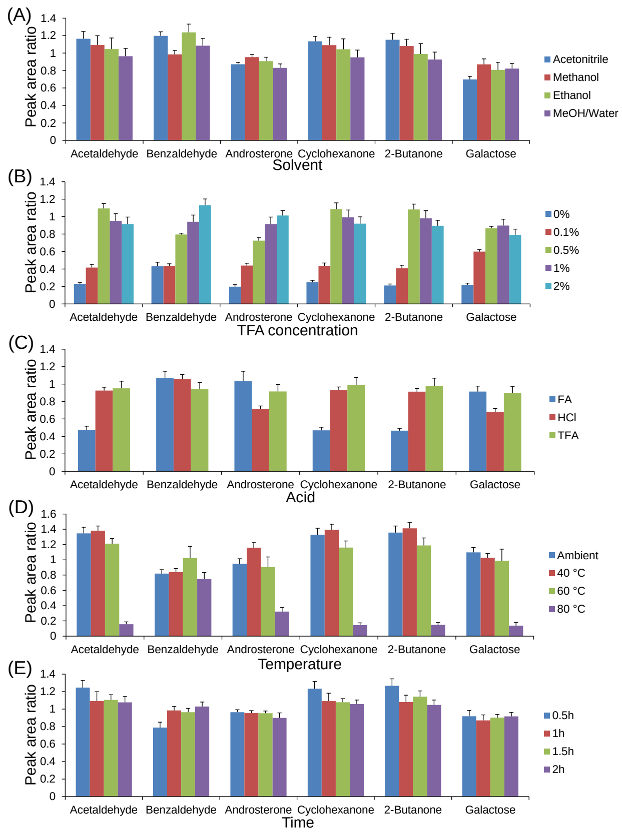

(A)
Peak area ratio
Acetaldehyde Benzaldehyde Androsterone Cyclohexanone 2-Butanone Galactose
Solvent
Acetonitrile
Methanol
Ethanol
MeOH/Water
(B)
Peak area ratio
TFA concentration
0%
0.1%
0.5%
1%
2%
(C)
Peak area ratio
Acid
FA
HCl
TFA
(D)
Peak area ratio
Temperature
Ambient
40 °C
60 °C
80 °C
(E)
Peak area ratio
Time
0.5h
1h
1.5h
2h

Figure 3.

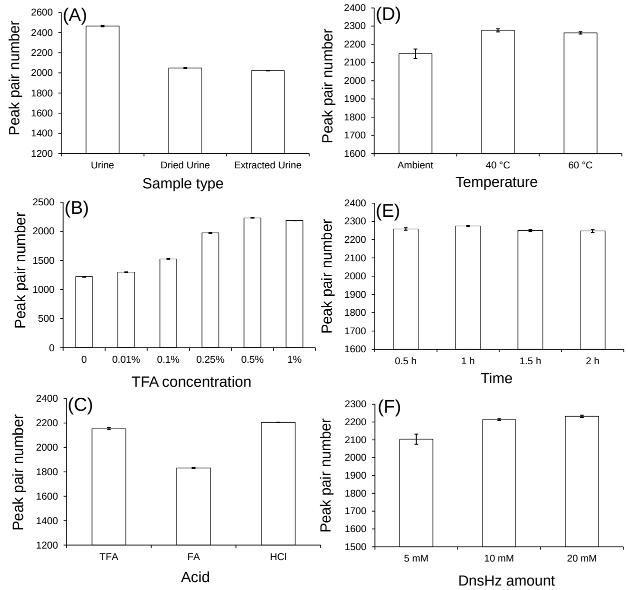

(A)
Peak pair number
Urine
Dried Urine
Extracted Urine
Sample type
(B)
Peak pair number
0
0.01%
0.1%
0.25%
0.5%
1%
TFA concentration
(C)
Peak pair number
TFA
FA
HCl
Acid
(D)
Peak pair number
Ambient
40 °C
60 °C
Temperature
(E)
Peak pair number
0.5 h
1 h
1.5 h
2 h
Time
(F)
Peak pair number
5 mM
10 mM
20 mM
DnsHz amount

Figure 4.

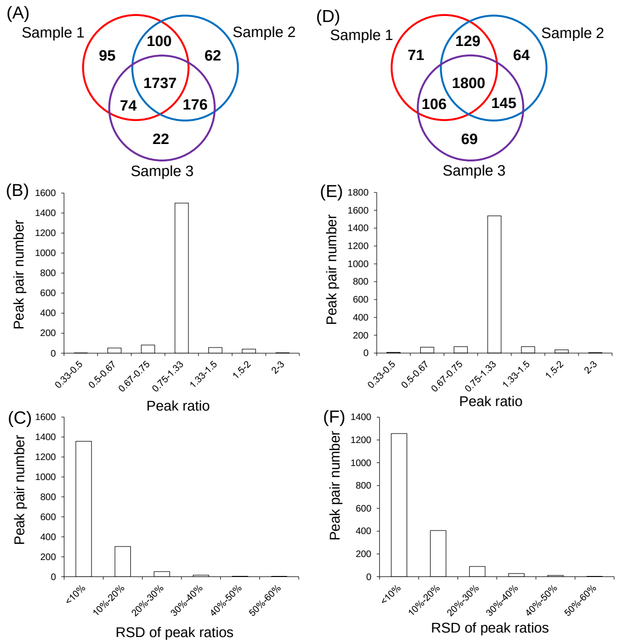
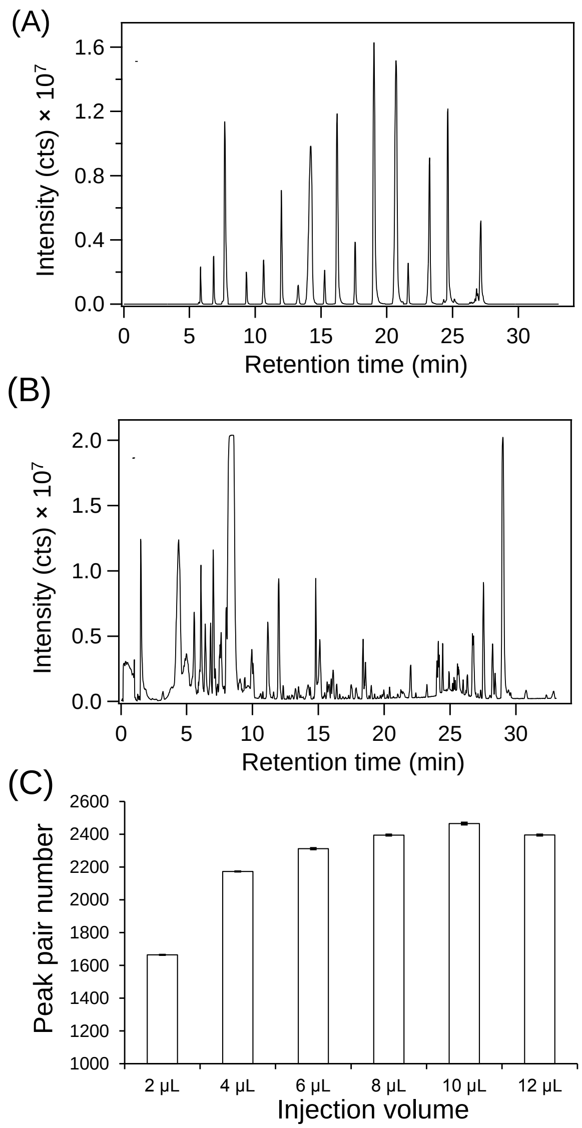

(A)
Sample 1
Sample 2
Sample 3
95
100
62
1737
74
176
22
(D)
Sample 1
Sample 2
Sample 3
71
129
64
1800
106
145
69
(B)
Peak pair number
Peak ratio
0.33-0.5
0.5-0.67
0.67-0.75
0.75-1.33
1.33-1.5
1.5-2
2-3
(E)
Peak pair number
Peak ratio
(C)
Peak pair number
RSD of peak ratios
<10%
10%-20%
20%-30%
30%-40%
40%-50%
50%-60%
(F)
Peak pair number
RSD of peak ratios

Figure 5.


(A)
Intensity (cts) × 10^7
1.6
1.2
0.8
0.4
0.0
0
5
10
15
20
25
30
Retention time (min)
(B)
Intensity (cts) × 10^7
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
30
Retention time (min)
(C)
Peak pair number
2600
2400
2200
2000
1800
1600
1400
1200
1000
2 μL
4 μL
6 μL
8 μL
10 μL
12 μL
Injection volume

For Table of Contents: