HHS Public Access

Author manuscript

Anal Chem. Author manuscript; available in PMC 2017 December 20.

Published in final edited form as:

Anal Chem. 2016 December 20; 88(24): 12411–12418. doi:10.1021/acs.analchem.6b03724.

Comprehensive Metabolite Identification Strategy using Multiple 2D NMR Spectra of a Complex Mixture Implemented in the COLMARm Web Server

Kerem Bingol[1,#], **Da-Wei Li**[2,#], **Bo Zhang**[3], and **Rafael Brüschweiler**[2,3,4,*]

[1]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99354, United States

[2]Campus Chemical Instrument Center, The Ohio State University, Columbus, Ohio 43210, United States

[3]Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States

[4]Department of Biological Chemistry and Pharmacology, The Ohio State University, Columbus, Ohio 43210, United States

Abstract

Identification of metabolites in complex mixtures represents a key step in metabolomics. A new strategy is introduced, which is implemented in a new public web server, COLMARm, that permits the co-analysis of up to three 2D NMR spectra, namely ^{13}C-^{1}H HSQC, ^{1}H-^{1}H TOCSY, and ^{13}C-^{1}H HSQC-TOCSY for the comprehensive, accurate, and efficient performance of this task. The highly versatile and interactive nature of COLMARm permits its application to a wide range of metabolomics samples independent of the magnetic field. Database query is performed using the HSQC spectrum and the top metabolite hits are then validated against the TOCSY-type experiment(s) by superimposing the expected cross-peaks on the mixture spectrum. In this way the user can directly accept or reject candidate metabolites by taking advantage of the complementary spectral information offered by these experiments and their different sensitivities. The power of COLMARm is demonstrated for a human serum sample uncovering the existence of 14 metabolites that hitherto were not identified by NMR.

TOC image

*To whom correspondence should be addressed: Rafael Brüschweiler, Ph.D., Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, bruschweiler.1@osu.edu.

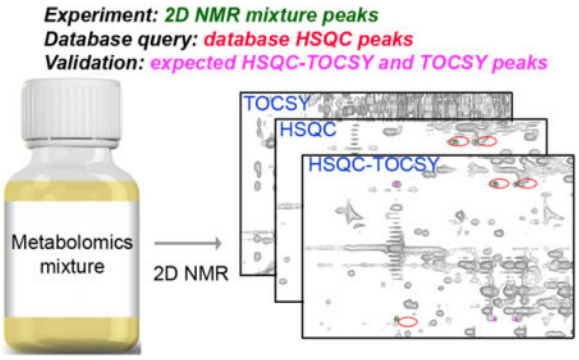
#These authors contributed equally to this work

CONFLICTS OF INTEREST

There are no conflicts of interest.

Supporting Information

2D HSQC spectrum of human serum, demonstration for two metabolites how they can be validated by COLMARm. This information is available free of charge via the Internet at http://pubs.acs.org/.



1. INTRODUCTION

Multidimensional nuclear magnetic resonance (NMR) spectroscopy offers excellent resolution, accuracy, and reproducibility for metabolite identification in metabolomics studies.[1234] In 2D NMR experiments, where pairs of spins (e.g. two ^{1}H spins or a ^{1}H spin and a ^{13}C spin) are physically correlated with each other leading to the appearance of cross-peaks in the resulting 2D NMR spectrum, peak overlap is greatly diminished when compared with the corresponding 1D NMR spectrum.[56] Two of the most commonly used 2D NMR experiments in metabolomics are the 2D ^{13}C-^{1}H heteronuclear single quantum coherence spectroscopy (HSQC) experiment,[7] which provides correlations between chemical shifts of ^{1}H spins with their directly attached ^{13}C spins, and the 2D ^{1}H-^{1}H total correlation spectroscopy (TOCSY),[8] which monitors correlations between all ^{1}H spins within the same spin system or molecule. The TOCSY experiment contains valuable connectivity information about entire spin systems within a molecule, which is not available from HSQC. The combination of the two experiments into one, the 2D HSQC-TOCSY,[9] combines some of the benefits of both experiments, but has a lower sensitivity than either one of them. The short- and long-range connectivity information provided by TOCSY-type spectra makes them uniquely suitable for the validation of metabolites that were identified from HSQC experiments alone. Moreover, such connectivity information allows deconvolution[10] as well as structure elucidation of unknown metabolites in complex mixtures.[1112]

2D NMR-based metabolite identification has made significant progress in recent years.[131415161718192021] In our research, customized metabolomics databases for the querying HSQC and TOCSY experiments have proven instrumental to significantly increase the accuracy of metabolite identification.[22] The ^{13}C-TOCCATA customized database[23] is optimized for the querying of ^{13}C-^{13}C TOCSY spectra of uniformly ^{13}C-labeled metabolomics samples, whereas the ^{1}H(^{13}C)-TOCCATA customized database[24] permits the querying of ^{1}H-^{1}H TOCSY and ^{13}C-^{1}H HSQC-TOCSY spectra of complex metabolite mixtures at natural ^{13}C abundance. A key feature of the TOCCATA databases is that they sort the spectral information of each metabolite into individual spin systems and slowly interchanging isomers matching the information directly obtainable from these spectra. This substantially increases the accuracy (reliability) and confidence of metabolite identification based on these types of experiments. Following a similar philosophy, the COLMAR HSQC

metabolomics database[25] was developed for the querying of ^{13}C-^{1}H HSQC spectra by sorting the HSQC spectra of metabolites into individual slowly interchanging isomeric states. This improves the query result, since it is insensitive to the relative populations of different isomers, which sometimes can be quite uneven. The unambiguous detection of one isomer per compound is then sufficient for a successful query when the other isomer(s) are below the detection limit. COLMAR HSQC and TOCCATA databases were generated by unifying chemical shift information of two of the largest public metabolomics databases, namely the Biological Magnetic Resonance Data Bank[13] and The Human Metabolome Database.[17]

Although HSQC and TOCSY-type spectra each provide powerful information on their own, their combined use could further improve the reliability of metabolite identification. However, the complementary nature of these experiments has limited their combined use. At present, they are used primarily for separate and independent querying against HSQC and TOCSY metabolomics databases, respectively, without taking advantage of potential synergies.[262728]

Here, we introduce an integrated metabolite identification and validation approach, COLMARm, which combines the unique strengths of 2D HSQC and TOCSY-type experimental data. The approach allows the simultaneous analysis of multiple NMR spectra, namely HSQC, TOCSY, and HSQC-TOCSY, thereby improving the accuracy and scope of NMR-based metabolite identification of complex mixtures. The approach is implemented in the new COLMARm web server, which is public and which provides a large number of interactive capabilities that enable the comprehensive identification and analysis of a large number of metabolites from a single sample in a way that is intuitive and efficient. COLMARm is the most advanced and most interactive web server within the COLMAR suite of web servers. The COLMARm approach is demonstrated for a human serum sample allowing the detection of a significant number of metabolites that previously could not be identified by NMR methods alone.

2. EXPERIMENTAL SECTION

2.1. Sample Preparation

10 mL of pooled normal human serum, which is a combination of serum samples from multiple non-diseased (normal) individuals between the age of 18 and 65, was purchased from Innovative Research Inc. (Novi, MI). The serum was treated using nanoparticle-assisted protein removal combined with ultrafiltration as described previously.[29] Briefly, 2 mL Bindzil 2040 silica nanoparticles (AkzoNobel and Eka Chemicals) were used to remove the majority of serum protein. The remaining proteins were removed with an omega membrane filter unit with 1 kD molecular weight cut-off (Pall Life Sciences, Port Washington, NY). The flow-through was lyophilized and dissolved in 200 μL D_2O with 50 mM phosphate buffer and 0.5 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Next, the sample was transferred to a 3-mm tube for all NMR experiments. The spike-in experiment was performed by spiking 1 mM isoleucine, 2 mM triglyceric acid, 2 mM trigonelline, and 2 mM 2-aminobutyric acid into a fresh serum sample followed by the collection of the same 2D NMR data sets for comparison with the original serum sample.

Only the spiked metabolites were found to have increased peak intensities whereas the other metabolites remained unchanged, which confirmed the presence of these metabolites in the original serum sample.

2.2 NMR Experiments and Processing

2D ^{13}C-^{1}H HSQC and 2D ^{13}C-^{1}H HSQC-TOCSY spectra of human serum were collected with N_1=512 and N_2=1024 complex points along the two dimensions. The spectral widths along the indirect and the direct dimensions were 34211.2 Hz and 10204.1 Hz, respectively. The number of scans per t_1 increment was 64 for the ^{13}C-^{1}H HSQC and 96 for the ^{13}C-^{1}H HSQC-TOCSY experiment. The transmitter frequency offsets were 75 ppm in the ^{13}C dimension and 4.726 ppm in the ^{1}H dimension. 2D ^{1}H-^{1}H TOCSY spectrum of human serum was collected with N_1=512 and N_2=2048 complex points. The spectral widths along the indirect and the direct dimensions were 10202.0 Hz and 10204.1 Hz, respectively. The number of scans per t_1 increment was 32. The transmitter frequency offsets were 4.726 ppm in both ^{1}H dimension. The isotropic mixing times for ^{13}C-^{1}H HSQC-TOCSY and ^{1}H-^{1}H TOCSY were 80 ms and 117 ms, respectively. The relaxation delay between consecutive scans (d1) was 1.5 seconds. The NMR spectrum was collected using a cryogenically cooled TCI probe at 850 MHz proton frequency on a Bruker Avance III spectrometer. All NMR spectra, which were collected at 298 K, were zero-filled, Fourier transformed, and phase and baseline corrected using NMRPipe.[30] The collection of all three 2D NMR experiments takes typically 2 – 3 days. Cosine-squared and cosine bell window functions were used to for apodization of the HSQC and HSQC-TOCSY spectra along their direct and indirect dimensions, respectively. Cosine bell functions were used for TOCSY spectra along the direct and indirect dimensions.

3. RESULTS AND DISCUSSION

3.1. General strategy for simultaneous analysis of 2D HSQC and TOCSY for metabolite identification

The spectral information contained in HSQC and TOCSY spectra is highly complementary. The HSQC spectrum is simpler than TOCSY and at the same time its interpretation is also more ambiguous, since it does not contain information about medium- or long-range molecular connectivities, i.e. it is not a priori known whether two cross-peaks belong to the same molecule or not. At ^{13}C natural abundance, HSQC experiment has an inherently reduced sensitivity because only 1 in 99 carbons is a ^{13}C isotope. Among the two TOCSY-type experiments, HSQC-TOCSY has better resolution along the indirect ω_1 (^{13}C) dimension, whereas ^{1}H-^{1}H TOCSY has better sensitivity because it does not depend on the low abundance ^{13}C isotopes. Hence, among the 3 2D experiments, ^{1}H-^{1}H TOCSY has the highest sensitivity, which is followed by ^{13}C-^{1}H HSQC with medium sensitivity. The ^{13}C-^{1}H HSQC-TOCSY is the least sensitive experiment, since it does not only suffer from ^{13}C natural abundance, but also from the distribution of the initial magnetization over multiple spins during TOCSY mixing.

The new COLMARm metabolite identification strategy consists of two main steps, which are depicted in Figure 1: the first step consists of querying against the HSQC spectrum and

the second step concerns validation using the TOCSY and/or HSQC-TOCSY spectra. In the first step, all resolved cross-peaks in the ^{13}C-^{1}H HSQC are queried against the COLMAR HSQC metabolomics database.[25] Next, for each positive metabolite hit that fulfills the well-defined quality criteria, the corresponding ^{1}H-^{1}H TOCSY and/or ^{13}C-^{1}H HSQC-TOCSY spectrum is reconstructed from the COLMAR $^{1}H(^{13}C)$-TOCCATA database[24] and superimposed on the experimental TOCSY and/or HSQC-TOCSY spectrum. Depending on whether the reconstructed TOCSY and HSQC-TOCSY cross-peaks superimpose well on the corresponding cross-peaks observed in the mixture spectra, the metabolite hit is either manually confirmed or rejected. In this way, a list of compounds is assembled whose spectral signatures are consistent with the mixture cross-peaks observed in HSQC, TOCSY and/or HSQC-TOCSY. This provides a level of confidence and accuracy in the identified compounds, which is not obtainable when using any one of these spectra alone.

This approach was implemented in the new COLMARm web server, which is the latest and most versatile addition to our suite of public COLMAR web servers: http://spin.ccic.ohio-state.edu/index.php/colmarm. COLMARm is equipped with many interactive features to provide a flexible and easy-to-use platform for a broad range of scientists with different NMR backgrounds in the analysis of complex metabolite mixtures. The relationship between the workflow of COLMARm and the existing COLMAR HSQC web server is shown in Figure 2. The COLMARm web server uses a 2D ^{13}C-^{1}H HSQC spectrum and one or both 2D TOCSY-type spectra (^{1}H-^{1}H TOCSY and/or ^{13}C-^{1}H HSQC-TOCSY) of the same mixture sample as input. Currently accepted file formats for the 2D spectra are the widely used NMRPipe (.ft2) and Sparky (.ucsf) formats. An example input data set is available on the web server for demonstration and training purposes. After uploading the experimental mixture spectra, the web server displays the experimental 2D HSQC spectrum on the browser window with an automatically adjusted contour level, which is followed by automated peak picking. A useful web server feature is the adjustment of contour levels by the user: one can decrease the lowest contour level to pick more peaks especially the ones belonging to metabolites at low concentration or increase the contour level to pick fewer peaks to remove thermal noise, t_1-noise, and other artifacts. Upon change of the contour levels, the web server redraws the spectrum, and performs automated peak picking to update the peak list for the new contour levels. Furthermore, the user can add and remove cross-peaks to manually curate the final peak list. Since chemical shift reference errors can have a large adverse effect on database query, COLMARm provides a convenient semi-automatic referencing option, which is based on pattern matching of either the standard referencing compound DSS or commonly occurring metabolites alanine, leucine, glucose, and lactic acid. Users can visually inspect the matching results of any of these metabolites when they are present in the sample and perform automated or manual referencing. Once the HSQC peak list has been finalized, it is queried against the database. The querying algorithm used for the matching of compounds was adopted from the COLMAR HSQC web server.[25] For each database compound, or isomer, the average of ^{1}H and ^{13}C chemical shift differences (1^{st} and 2^{nd} output parameters) are computed to the closest cross-peaks of the mixture. If the database cross-peak is within a given frequency cutoff, it is considered a “matched peak”. The “matching ratio” is then defined as the ratio of the matched peaks to the total number of peaks of the metabolite (3^{rd} output parameter). In contrast to COLMAR HSQC,

COLMARm does not use the uniqueness parameter to filter false positives. Instead, the TOCSY-based connectivity information is used to remove false positives. On the new COLMARm web server, the default HSQC cutoff parameters for the average ^{1}H and ^{13}C chemical shift differences are set at 0.03 and 0.3 ppm, respectively, and the lowest accepted matching ratio is set to 0.6. Users have the option to interactively change these three parameters on the web server.

Next, for each positive metabolite hit obtained using the above criteria, COLMARm automatically computes the expected TOCSY and/or HSQC-TOCSY spectra for this compound allowing the user to directly compare the reconstructed cross-peaks with the experimental TOCSY and/or HSQC-TOCSY spectra. Further down on the COLMARm web server page, experimental HSQC, TOCSY, and HSQC-TOCSY spectra can be plotted side-by-side or top-to-bottom. The interactive user interface written in Java allows the superposition of HSQC, TOCSY, and HSQC-TOCSY peaks of individual compounds in the database, displayed as circles or ellipses, with the experimental peaks upon selecting the metabolite from the “matched compound” list. Using its automated zoom-in feature, the web server selects and reassembles only the relevant regions of the HSQC, TOCSY and HSQC-TOCSY spectra of each metabolite of interest for visual inspection at very high spectral resolution. This allows the simultaneous and user-friendly inspection of cross-peaks in matched compounds in up to three experimental 2D spectra. This is the first metabolomics database that allows co-analysis of HSQC, TOCSY and HSQC-TOCSY spectra to maximize the confidence in metabolite identification. The matched compounds are returned in the format “Metabolite Name_Number_Number”, where the integer in the middle is used to label the different isomeric states of a given metabolite, whereas the integer at the end labels the different spin systems of the metabolite. Metabolites with only one isomeric state appear as “Metabolite_1_1”, such as “Acetic_acid_1_1”, whereas metabolites with more than one isomeric state and/or spin system appear as “Metabolite Name_n_m” where n=1,2,3, … are different isomers of the same metabolite (e.g., Glucose_1_1, Glucose_2_1 for α-glucose and β-glucose, respectively) and m=1,2,3, … are different spin systems of the metabolite (e.g., L-Carnitine_1_1, L-Carnitine_1_2). Only NMR data of compounds that were measured in aqueous solution (H_2O/D_2O) at pH 7.0 or 7.4 were included in COLMARm. A complete list of database compounds with their number of isomeric states and spin systems can be found on the web server. Since the previous publication of COLMAR,[25] the number of metabolites in the database has increased from 555 to 701 metabolites offering improved coverage of common metabolites. The web server includes a number of user-friendly interactive features. For example, users can adjust the zoom level and the figure size with built-in sliders, download matched compounds information and produce high-quality figures at any stage of their work. Minimal effort is required if users want to start over, e.g., to explore other choices for contour levels or database query parameters. Each job is saved on the server with a unique file name only known to the user, which permits users to resume work at a later time without the need to upload the spectra again.

3.2. Application to human serum

COLMARm is demonstrated on a human serum sample collected from a pool of healthy individuals. A concentrated serum sample (10 mL), which was treated with the recently

introduced nanoparticle-assisted ultrafiltration protein removal approach,[29] provided information-rich TOCSY and HSQC-type spectra (Figure S1). After uploading of all three spectra to COLMARm, the contour levels of the HSQC spectrum were adjusted first using the slider above the spectrum. Next, automated peak picking was performed and the resulting peak list was queried against COLMARm. The returned metabolite hits were analyzed visually one-by-one. Figure 3 illustrates the protocol for serum for the identification of isoleucine. The experimental HSQC peaks of isoleucine (green circles) matched reasonably well the database chemical shifts of isoleucine (red circles) (Figure 3A). A small shift was observed because of differences in pH and ionic strength between the serum sample and the sample of the isolated reference compound despite pH adjustment before the measurement. Observation of the cross-peak connectivity patterns of isoleucine in serum TOCSY (Figure 3B) as well as HSQC-TOCSY (Figure 3C), which were consistent with the expected cross-peak patterns from the database, validated the presence of isoleucine in serum. By using the same protocol, we identified 62 distinct metabolites in serum. A list of the identified metabolites can be found in Table 1, where those serum metabolites that were identified from the NMR spectrum of the entire mixture for the first time are marked with “N” (new). One of the new metabolites identified in serum was glyceric acid (Figure 4). Glyceric acid peaks appeared in the crowded region of the HSQC (Figure 4A). Its validation based on ^{1}H-^{1}H TOCSY was difficult because of heavy peak overlap in the 3.7 – 3.9 ppm proton region (Figure 4B). Peak overlaps could be resolved in the ^{13}C-^{1}H HSQC-TOCSY due to the better peak dispersion along the carbon dimension, which provided a direct validation of the HSQC hit (Figure 4C). This example illustrates the benefit of using the complementarity information of TOCSY and HSQC-TOCSY spectra in COLMARm. Another instructive example concerns the identification of 2-aminobutyric acid, which has three HSQC cross-peaks (Figure S2A) that all reside in an overlapped ^{1}H region and made validation by TOCSY alone difficult (Figure S2B). Peak overlaps were resolved in the HSQC-TOCSY along the carbon dimension, which validated this identification (Figure S2C).

It is well-known that HSQC-TOCSY and HSQC spectra have higher resolution than TOCSY, but at the same time lower sensitivity. This makes TOCSY particularly useful for the identification of low abundance metabolites. One of the low abundance metabolites identified in serum was trigonelline. Trigonelline has five HSQC peaks, but only two had a sufficiently high signal-to-noise in the HSQC spectrum of serum (Figure S3A), which made the identification of trigonelline by using HSQC alone ambiguous. On the other hand, many of the trigonelline signals were detected in the serum TOCSY spectrum (Figure S3B), which again illustrates the beneficial complementarity of TOCSY.

As mentioned above, an important difference between COLMAR HSQC and COLMARm query is that the latter does not use the uniqueness parameter to filter out false positives, since TOCSY-based connectivity information is more effective to filter out false positives in the HSQC spectrum. In this study, several of the true positive metabolites such as 1-methlyhistidine and 2-aminobutyric acid had a uniqueness value 0, which would have been missed if the uniqueness filter had been applied in the HSQC-based query. It demonstrates the advantage of this type of co-analysis of multiple 2D NMR spectra to maximize the number of identified metabolites in serum.

Human serum is a well-studied biological sample. Out of the 62 serum metabolites identified in this study, 48 of them were previously identified in human serum by NMR.[3132] However, the other 14 metabolites (1-methyluric acid, 3-hydroxypropionic acid, alpha-hydroxyisobutyric acid, glyceric acid, glycerophosphocholine, ethanolamine, citrulline, hypotaurine, malate, trigonelline, taurine, valeric acid, beta-alanine, trans-4-hydroxyproline) have been identified here for the first time by using NMR alone. Taurine, citrulline and trans-4-hydroxyproline were previously identified in human blood plasma by NMR.[3334] The Metabolomics Standards Initiative (MSI) categorizes the confidence of metabolite identification according to four different levels, where level 1 (highest confidence level) is achieved by spiking of authentic standards into the original sample.[35] In this study only the presence of isoleucine, 2-aminobutyrate, glyceric acid, and trigonelline were confirmed by spiking experiments (level 1 identification). According to MSI, which defines the use of databases for metabolite identification as level 2 identification,[35] the remaining 58 metabolites have been identified at this level by COLMARm. Some of the newly identified metabolites were previously detected and identified in human serum by other methods. For instance, glyceric acid and alpha-hydroxyisobutyric acid were previously identified in serum by GC-MS,[31] glycerophosphocholine was previously identified by LC-MS,[36] and taurine and citrulline were previously identified by ion-exchange chromatography.[37] Literature references for the other identified metabolites in this study can be found in the human metabolome database (HMDB).[17] The total number of unique HSQC peaks detected in serum is 480, but only 225 of them belong to the 63 compounds (including DSS). Assuming a similar number of HSQC peaks per compound, another ~71 metabolites are estimated to be present that show up in the HSQC spectrum, but their identity remains unknown. In a recent study, 1012 unique HSQC peaks were reported for human urine;[38] therefore, with the sensitivity of our current TCI cryoprobe, concentrated human serum with its 480 unique HSQC peaks is not as metabolite rich as urine. Notably, some of the newly identified serum metabolites such as trigonelline, taurine and beta-alanine were previously detected in human urine.[394041]

Currently, peak-picking and querying of HSQC spectra with the COLMARm web server is fully automated, while the validation step with TOCSY and HSQC-TOCSY involves human intervention, since the trained human eye is very capable at distinguishing true cross-peaks, including overlapping cross-peaks, from spectral noise and other artifacts. The graphical web server interface helps users to efficiently make these decisions. A future goal is to further automate the entire COLMARm workflow.

4. CONCLUSIONS

The accurate and comprehensive identification of metabolites is one of the most important steps in metabolomics. Here, we presented the first metabolomics web server that allows simultaneous analysis of multiple commonly used 2D NMR spectra, namely HSQC, TOCSY and HSQC-TOCSY, of metabolite mixtures on the same web platform by using the customized 2D NMR metabolomics databases COLMAR HSQC and TOCCATA. By combining the high-resolution nature of HSQC with spin-connectivity information from TOCSY and HSQC-TOCSY, we achieved metabolite identification with high accuracy, which would not be possible using 1D NMR spectra or any of the 2D NMR spectra alone.

The approach allowed identification of 14 new metabolites in serum, thereby expanding the list of metabolites that can be routinely studied in serum by NMR spectroscopy. Still, a large number of HSQC signals in human serum remains unidentified, which can be attributed to the incompleteness of the current NMR metabolomics databases. In order to identify unknown metabolites, one can use physical separation methods[4243] or database-independent hybrid strategies, such as the recently developed SUMMIT MS/NMR approach.[4445] Expansion of the COLMARm web server with newly discovered metabolites enables routine identification of an ever larger number of metabolites in metabolomics mixtures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health (grant R01 GM 066041 and SECIM (Southeast Center for Integrated Metabolomics) grant U24 DK097209-01A1).

References

1. Bingol K, Brüschweiler R. Curr Opin Biotechnol. 2017; 43:17–24. [PubMed: 27552705]
2. Larive CK, Barding GA Jr, Dinges MM. Anal Chem. 2014; 87:133–146. [PubMed: 25375201]
3. Fan TWM, Lane AN. Prog Nucl Magn Reson Spectrosc. 2016; 92:18–53. [PubMed: 26952191]
4. Markley JL, Brüschweiler R, Edison AS, Eghbalnia HR, Powers R, Raftery D, Wishart DS. Curr Opin Biotechnol. 2017; 43:34–40. [PubMed: 27580257]
5. Bingol K, Brüschweiler R. Anal Chem. 2014; 86:47–57. [PubMed: 24195689]
6. Bingol K, Zhang F, Bruschweiler-Li L, Brüschweiler R. Anal Chem. 2013; 85:6414–6420. [PubMed: 23773204]
7. Bodenhausen G, Ruben DJ. Chem Phys Lett. 1980; 69:185–189.
8. Braunschweiler L, Ernst RR. J Magn Reson. 1983; 53:521–528.
9. Lerner L, Bax A. J Magn Reson. 1986; 69:375–380.
10. Bingol K, Brüschweiler R. Anal Chem. 2011; 83:7412–7417. [PubMed: 21848333]
11. Bingol K, Zhang F, Bruschweiler-Li L, Brüschweiler R. J Am Chem Soc. 2012; 134:9006–9011. [PubMed: 22540339]
12. Komatsu T, Ohishi R, Shino A, Kikuchi J. Angew Chem Int Ed. 2016; 55:6000–6003.
13. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL. Nucleic Acids Res. 2008; 36:D402–408. [PubMed: 17984079]
14. Cui Q, Lewis IA, Hegeman AD, Anderson ME, Li J, Schulte CF, Westler WM, Eghbalnia HR, Sussman MR, Markley JL. Nat Biotechnol. 2008; 26:162–164. [PubMed: 18259166]
15. Chikayama E, Sekiyama Y, Okamoto M, Nakanishi Y, Tsuboi Y, Akiyama K, Saito K, Shinozaki K, Kikuchi J. Anal Chem. 2010; 82:1653–1658. [PubMed: 20128615]
16. Ludwig C, Easton JM, Lodi A, Tiziani S, Manzoor SE, Southam AD, Byrne JJ, Bishop LM, He S, Arvanitis TN, Günther UL, Viant MR. Metabolomics. 2012; 8:8–18.
17. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, Arndt D, Xia J, Liu P, Yallou F, Bjorndahl T, Perez-Pineiro R, Eisner R, Allen F, Neveu V, Greiner R, Scalbert A. Nucleic Acids Res. 2013; 41:D801–807. [PubMed: 23161693]
18. Qiu F, McAlpine JB, Lankin DC, Burton I, Karakach T, Chen S-N, Pauli GF. Anal Chem. 2014; 86:3964–3972. [PubMed: 24673652]

19. Dubey A, Rangarajan A, Pal D, Atreya HS. Anal Chem. 2015; 87:7148–7155. [PubMed: 26101967]
20. Clendinen CS, Pasquel C, Ajredini R, Edison AS. Anal Chem. 2015; 87:5698–5706. [PubMed: 25932900]
21. Kikuchi J, Tsuboi Y, Komatsu K, Gomi M, Chikayama E, Date Y. Anal Chem. 2015; 88:659–665. [PubMed: 26624790]
22. Bingol K, Bruschweiler-Li L, Li D, Zhang B, Xie M, Brüschweiler R. Bioanalysis. 2016; 8:557–573. [PubMed: 26915807]
23. Bingol K, Zhang F, Bruschweiler-Li L, Brüschweiler R. Anal Chem. 2012; 84:9395–9401. [PubMed: 23016498]
24. Bingol K, Bruschweiler-Li L, Li DW, Brüschweiler R. Anal Chem. 2014; 86:5494–5501. [PubMed: 24773139]
25. Bingol K, Li DW, Bruschweiler-Li L, Cabrera OA, Megraw T, Zhang F, Brüschweiler R. ACS Chem Biol. 2015; 10:452–459. [PubMed: 25333826]
26. Sadykov MR, Zhang B, Halouska S, Nelson JL, Kreimer LW, Zhu Y, Powers R, Somerville GA. J Biol Chem. 2010; 285:36616–36624. [PubMed: 20861019]
27. Garcia-Manteiga JM, Mari S, Godejohann M, Spraul M, Napoli C, Cenci S, Musco G, Sitia R. J Proteome Res. 2011; 10:4165–4176. [PubMed: 21744784]
28. Da Silva L, Godejohann M, Martin FP, Collino S, Burkle A, Moreno-Villanueva M, Bernhardt J, Toussaint O, Grubeck-Loebenstein B, Gonos ES, Sikora E, Grune T, Breusing N, Franceschi C, Hervonen A, Spraul M, Moco S. Anal Chem. 2013; 85:5801–5809. [PubMed: 23718684]
29. Zhang B, Xie M, Bruschweiler-Li L, Brüschweiler R. Anal Chem. 2015; 88:1003–1007. [PubMed: 26605638]
30. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax AD. J Biomol NMR. 1995; 6:277–293. [PubMed: 8520220]
31. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, Sinelnikov I, Krishnamurthy R, Eisner R, Gautam B. PloS one. 2011; 6:e16957. [PubMed: 21359215]
32. Nagana Gowda GA, Gowda YN, Raftery D. Anal Chem. 2014; 87:706–715. [PubMed: 25485990]
33. Nicholson JK, Foxall PJD, Spraul M, Farrant RD, Lindon JC. Anal Chem. 1995; 67:793–811. [PubMed: 7762816]
34. Gowda GAN, Tayyari F, Ye T, Suryani Y, Wei S, Shanaiah N, Raftery D. Anal Chem. 2010; 82:8983–8990. [PubMed: 20879716]
35. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TWM, Fiehn O, Goodacre R, Griffin JL. Metabolomics. 2007; 3:211–221. [PubMed: 24039616]
36. Lin L, Huang Z, Gao Y, Chen Y, Hang W, Xing J, Yan X. Proteomics. 2012; 12:2238–2246. [PubMed: 22685041]
37. Le Boucher J, Charret C, Coudray-Lucas C, Giboudeau J, Cynober L. Clin Chem. 1997; 43:1421–1428. [PubMed: 9267323]
38. Bingol K, Brüschweiler R. J Proteome Res. 2015; 14:2642–2648. [PubMed: 25881480]
39. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, Bjorndahl TC, Krishnamurthy R, Saleem F, Liu P. PloS one. 2013; 8:e73076. [PubMed: 24023812]
40. Zhang B, Xie M, Bruschweiler-Li L, Bingol K, Brüschweiler R. Anal Chem. 2015; 87:7211–7217. [PubMed: 26087125]
41. Walker LR, Hoyt DW, Walker SM II, Ward JK, Nicora CD, Bingol K. Magn Reson Chem. 2016; 54:998–1003. [PubMed: 27539910]
42. van der Hooft JJJ, de Vos RCH, Ridder L, Vervoort J, Bino RJ. Metabolomics. 2013; 9:1009–1018.
43. Sumner LW, Lei Z, Nikolau BJ, Saito K. Nat Prod Rep. 2015; 32:212–229. [PubMed: 25342293]
44. Bingol K, Bruschweiler-Li L, Yu C, Somogyi A, Zhang F, Brüschweiler R. Anal Chem. 2015; 87:3864–3870. [PubMed: 25674812]
45. Bingol K, Brüschweiler R. Curr Opin Clin Nutr Metab Care. 2015; 18:471–477. [PubMed: 26154280]

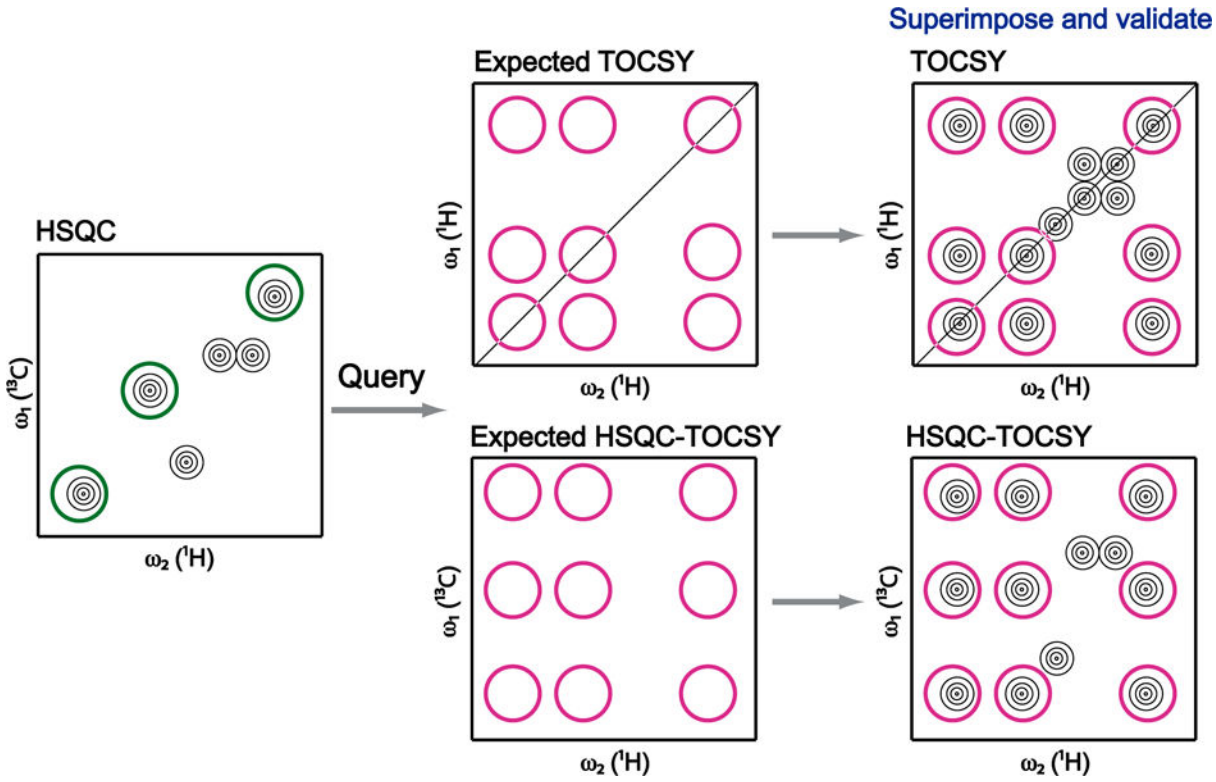


Figure 1.

The workflow of the proposed metabolite identification strategy implemented in the new interactive COLMARm web server for the combined analysis of 2D ^{13}C-^{1}H HSQC with 2D ^{1}H-^{1}H TOCSY and/or 2D ^{13}C-^{1}H HSQC-TOCSY. In the first step, all HSQC cross-peaks are queried against the COLMAR HSQC metabolomics database. Next, for each positive match, the corresponding TOCSY (and/or HSQC-TOCSY) spectrum is reconstructed from the TOCCATA metabolomics database and the expected diagonal and cross-peaks are superimposed on the experimental TOCSY and/or HSQC-TOCSY spectra for direct validation.

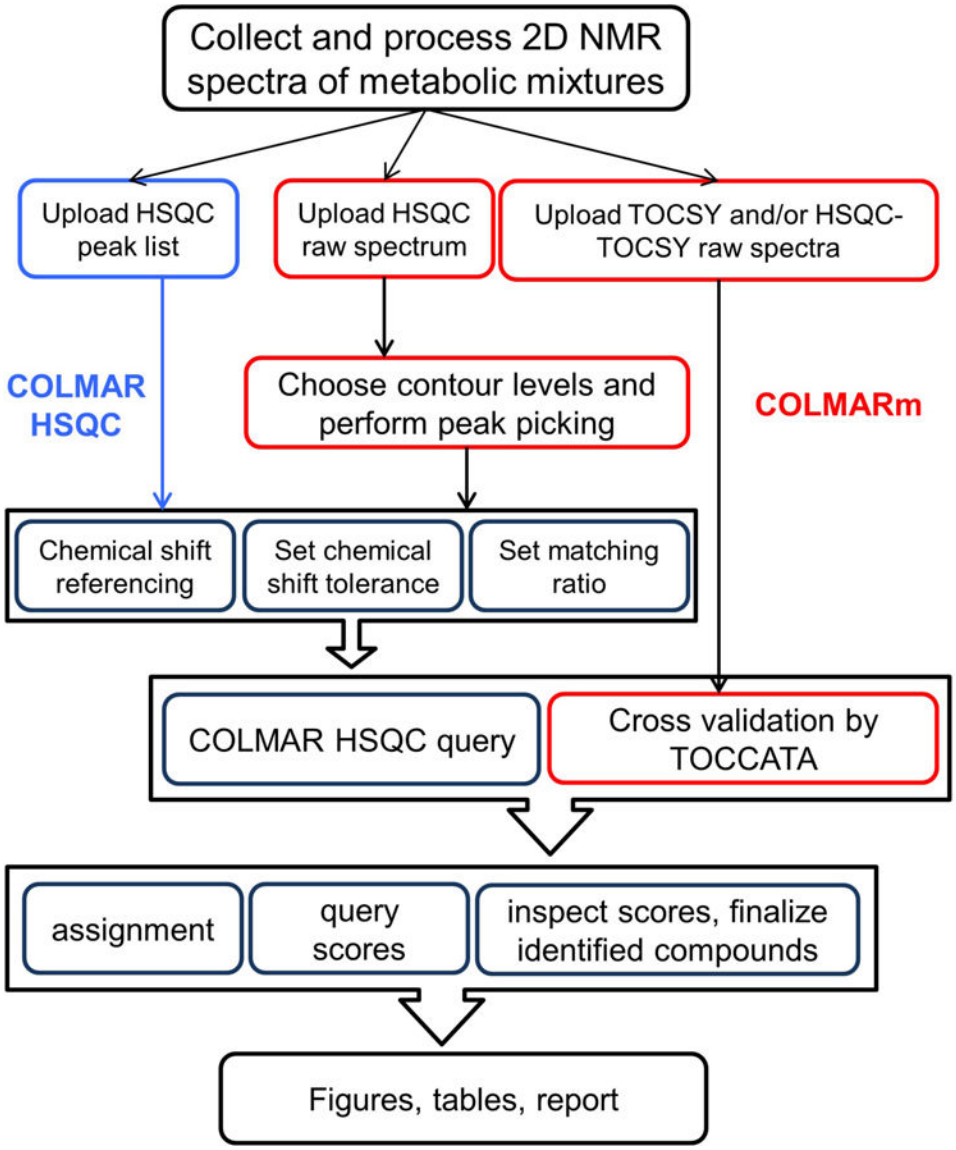


Figure 2.

Workflow of the new COLMARm web server in comparison with the existing COLMAR HSQC web server. Uploading of a single HSQC spectrum is sufficient for COLMAR HSQC, whereas COLMARm uses in addition: at least one TOCSY-type spectrum as input, i.e. ^{1}H-^{1}H TOCSY or ^{13}C-^{1}H HSQC-TOCSY or both, for the validation of the metabolite identities derived from HSQC.

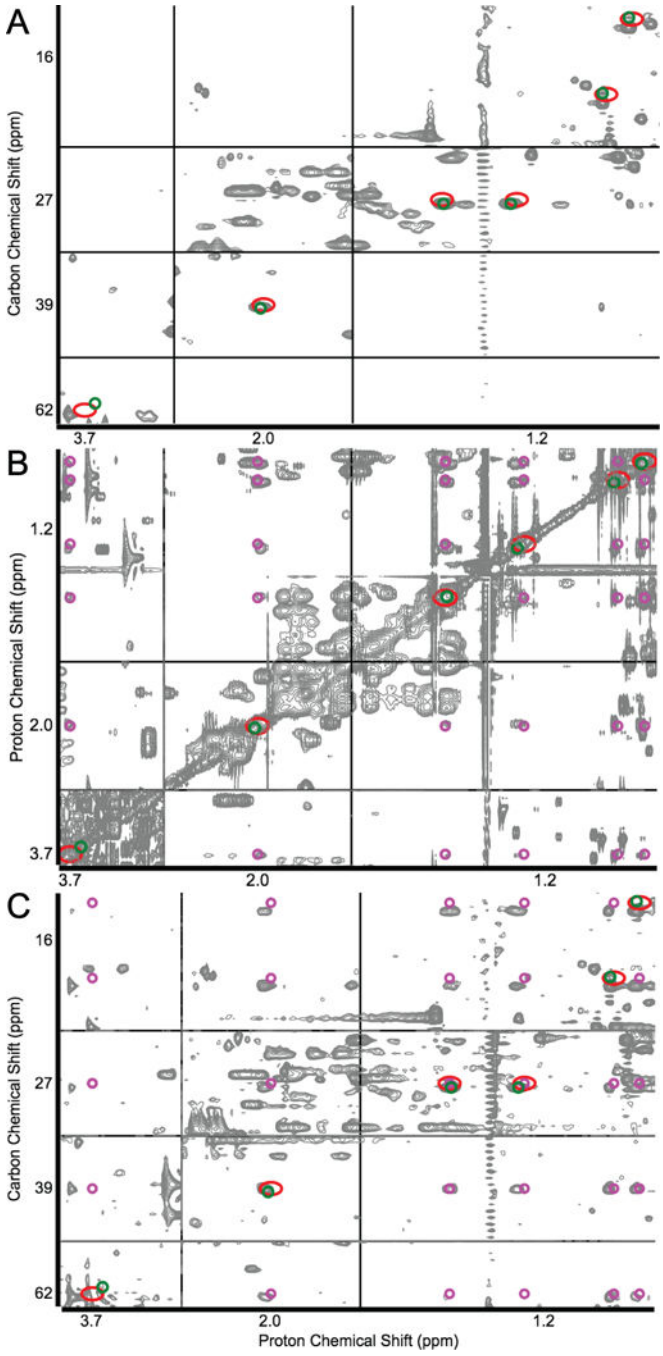


Figure 3.

Illustration of COLMARm with the identification of isoleucine in human serum by the co-analysis of (A) HSQC, (B) TOCSY, and (C) HSQC-TOCSY. Green and red circles represent experimental and database cross-peaks of isoleucine, respectively, whereas magenta circles represent the expected isoleucine peaks according to the TOCCATA database. The close match between green and red circles identifies isoleucine as a strong candidate. This was validated by the good agreement found for the expected magenta peaks with the experimentally observed TOCSY and HSQC-TOCSY cross-peaks.

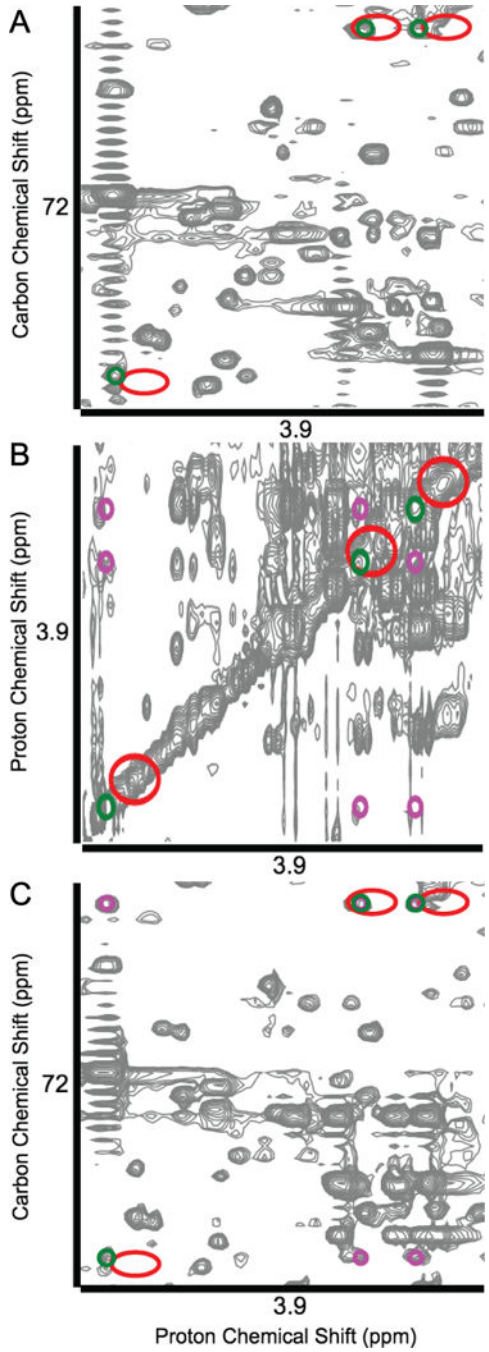


Figure 4.

Identification of glyceric acid with the COLMARm web server by uploading (A) HSQC, (B) TOCSY and (C) HSQC-TOCSY of human serum as input. Green and red circles represent experimental and database peaks of glyceric acid, respectively, whereas magenta circles represent the expected glyceric acid peaks from the TOCCATA database. The close match between green and red circles indicates that glyceric acid is a strong candidate, which was

validated by the good agreement between the magenta peaks and the TOCSY and HSQC-TOCSY peaks.

Table 1

Metabolites identified in human serum spectrum by the COLMARm web server. The letter "N" behind metabolite names (in parentheses) denote new metabolites identified in human serum by NMR spectroscopy alone. Chemical shift difference cutoff parameters for the query were 0.4 matching ratio and 0.6 ppm and 0.06 ppm for ^{13}C and ^{1}H chemical shift differences, respectively.

	M[a]	^{13}C[b]	^{1}H[c]		M	^{13}C	^{1}H
1. 1-Methylhistidine	1	0.34	0.030	**31.** Glyceric acid (N)	1	0.09	0.030
2. 1-Methyluric acid (N)	1	0.15	0.031	**32.** Glycerophosphocholine (N)	1	0.12	0.022
3. 2-Aminobutyric acid	1	0.38	0.032	**33.** Histidine	1	0.30	0.033
4. 2-Hydroxy-3-methyl butyric acid	1	0.29	0.024	**34.** Hypotaurine (N)	1	0.51	0.041
5. 2-Hydroxbutyric acid	1	0.30	0.016	**35.** Isobutyric acid	1	0.08	0.015
6. 3-Hydroxypropionic acid (N)	1	0.15	0.016	**36.** Isoleucine	1	0.19	0.015
7. 3-Hydroxyisovaleric acid	1	0.29	0.032	**37.** Isovaleric acid	1	0.14	0.012
8. 3-Hydroxyisobutyric acid	1	0.23	0.033	**38.** Lactate	1	0.07	0.040
9. 1,2-Propanediol	1	0.25	0.028	**39.** Leucine	1	0.21	0.019
10. Acetic acid	1	0.17	0.005	**40.** Lysine	1	0.26	0.026
11. Acetylcarnitine	1	0.15	0.025	**41.** N,N-Dimethylglycine	1	0.37	0.021
12. Alanine	1	0.22	0.046	**42.** Ornithine	1	0.20	0.026
13. Alpha-hydroxyisobutyric acid (N)	1	0.12	0.007	**43.** Proline	1	0.20	0.031
14. Arginine	1	0.14	0.023	**44.** Sarcosine	1	0.42	0.040
15. Betaine	1	0.21	0.030	**45.** Serine	1	0.26	0.038
16. Beta-alanine (N)	1	0.39	0.032	**46.** Succinic acid	1	0.16	0.008
17. Carnitine	1	0.18	0.021	**47.** Trans-4-hydroxyproline (N)	1	0.24	0.032
18. Choline	1	0.13	0.029	**48.** Trimethylamine-N-oxide	1	0.32	0.033
19. Citrate	1	0.03	0.025	**49.** Tryptophan	1	0.23	0.032
20. Citrulline (N)	1	0.25	0.029	**50.** Tyrosine	1	0.27	0.016
21. Creatine	1	0.18	0.023	**51.** Valeric acid (N)	1	0.32	0.023
22. Creatinine	1	0.13	0.021	**52.** Valine	1	0.14	0.007
23. Ethanolamine (N)	1	0.38	0.035	**53.** Phenylalanine	0.83	0.20	0.029
24. Fructose isomer 1	1	0.13	0.024	**54.** Uridine	0.88	0.15	0.035

	M[a]	^{13}C[b]	^{1}H[c]		M	^{13}C	^{1}H
24. Fructose isomer 2	1	0.21	0.027	**55.** 2-Pyrrolidinone-5-carboxylate	0.75	0.34	0.022
25. Fumaric acid	1	0.22	0.017	**56.** Hippuric acid	0.75	0.26	0.047
26. Glucose isomer 1	1	0.18	0.013	**57.** Methionine	0.75	0.35	0.024
26. Glucose isomer 2	1	0.16	0.017	**58.** Aspartate	0.67	0.42	0.027
27. Glutamate	1	0.19	0.024	**59.** Malate (N)	0.67	0.12	0.005
28. Glutamine	1	0.28	0.019	**60.** Threonine	0.67	0.20	0.026
29. Glycine	1	0.50	0.051	**61.** Taurine (N)	0.50	0.39	0.020
30. Glycerol	1	0.23	0.016	**62.** Trigonelline (N)	0.40	0.34	0.024

[a]**M** = matching ratio, which is the ratio between the number of matched peaks in the experimental HSQC spectrum to the number of peaks of the metabolite in the database. 1 (= 100%) corresponds to a perfect match and the matching quality decreases as it goes toward 0.

[b]Average ^{13}C chemical shift difference (in units of ppm) between the matched input and the database chemical shifts.

[c]Average ^{1}H chemical shift difference (in units of ppm) between the matched input and the database chemical shifts.