Published in final edited form as:

Anal Chem. 2008 October 1; 80(19): 7549–7553. doi:10.1021/ac801116u.

Self-consistent metabolic mixture analysis by heteronuclear NMR - Application to a human cancer cell line

Fengli Zhang¹, Lei Bruschweiler-Li^{1,2}, Steven. L. Robinette¹, and Rafael Brüschweiler^{1,2,*}

¹National High Magnet Field Laboratory, Florida State University, Florida, FL 32306

²Chemical Sciences Laboratory, Department of Chemistry and Biochemistry, Florida State University, Florida, FL 32306

Abstract

Elucidation of the chemical composition of biological samples is a main focus of systems biology and metabolomics. In order to comprehensively study these complex mixtures, reliable, efficient, and automatable methods are needed to identify and quantify the underlying metabolites and natural products. Because of its rich information content, nuclear magnetic resonance (NMR) spectroscopy has a unique potential for this task. Here we present a generalization of the recently introduced homonuclear TOCSY-based DemixC method to heteronuclear HSQC-TOCSY NMR spectroscopy. The approach takes advantage of the high resolution afforded along the ¹³C-dimension due to the narrow ¹³C linewidths for the identification of spin systems and compounds. The method combines information from both 1D ¹³C and ¹H traces by querying them against an NMR spectral database using our COLMAR query web server. The complementarity of ¹³C and ¹H spectral information improves the robustness of compound identification. The method is demonstrated for a metabolic model mixture and is then applied to an extract from DU145 human prostate cancer cells.

Keywords

Complex mixture analysis; metabolomics/metabonomics; covariance NMR spectroscopy; DemixC; HSQC-TOCSY; DU145 human prostate cancer cell line

Introduction

Identification of individual chemical components of biological systems and monitoring of their changes in response to a multitude of factors such as genetics, age, pathology, development, environment, stress, and treatment are key aspects of Metabolomics and Metabonomics. The comprehensive, systems biological approach to the study of metabolic mixtures thereby promises a better understanding of complex biochemical processes in living systems.^{1–3} Efficient and reliable analysis of these complex mixtures in terms of the underlying metabolites is an important prerequisite for achieving this goal. Different approaches are being developed for this task. Mass spectrometry-based approaches that are coupled with chromatography (e.g. GC-MS and UPLC-MS) provide high sensitivity for targeted compound analysis.⁴ Nuclear magnetic resonance (NMR) spectroscopy, on the other hand, has a unique potential as it does not *a priori* require potentially labor-intensive and costly physical separation of the components.⁵ This characteristic is utilized in a number

Correspondence to be addressed to: Rafael Brüschweiler, Chemical Sciences Laboratory, Department of Chemistry and Biochemistry, National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL 32306, Tel. 850-644-1768, Fax: 850-644-8281, bruschweiler@magnet.fsu.edu.

of NMR methods for complex mixture identification, including diffusion-ordered spectroscopy (DOSY),⁶ differential analysis of COSY spectra,⁷ HSQC spectra,^{8–10} selective 1D TOCSY^{11,12} and 2D TOCSY,¹³ STOCYSY,¹⁴ and DemixC.¹⁵

DemixC identifies individual components in a mixture based on a homonuclear ^1H - ^1H TOCSY NMR spectrum,¹⁶ which monitors spin-spin connectivity information across each molecule. The TOCSY spectrum is then covariance processed^{17–19} and deconvoluted by the clustering of its rows to identify traces that belong to individual spin systems or molecules.¹⁵ The chemical components are identified by screening these subspectra against a NMR spectral database, such as the BMRB²⁰ and the HMDB.²¹ This protocol has been recently implemented in our suite of public web servers at <http://spinportal.magnet.fsu.edu/>, termed COLMAR (for Complex Mixture Analysis by NMR), which includes a separate covariance NMR server, a DemixC server, and a database query server.

DemixC has been demonstrated recently for the chemical elucidation of an insect venom.²² The algorithms underlying COLMAR query for consensus matching have been described and demonstrated.²³ For a biological mixture that contains metabolites whose ^1H NMR spectra do strongly overlap, the difference of the COLMAR query score between the top database hits can be small and may require independent cross-validation of the query results.

For this purpose, an approach is introduced here that uses heteronuclear ^1H - ^{13}C HSQC-TOCSY NMR spectroscopy at natural abundance. While less sensitive than the homonuclear TOCSY, the ^1H - ^{13}C HSQC-TOCSY possesses narrow ^{13}C line widths, which makes peak overlaps unlikely. This allows application of DemixC to cross sections both along the ^1H and the ^{13}C dimensions to cross validate the results. This heteronuclear generalization of DemixC is first demonstrated for a metabolic test mixture and then applied to the analysis of the DU145 human prostate cancer cell line mixtures extracts.²⁴

Methods

Sample Preparation

A metabolic model mixture was prepared by mixing carnitine, glucose, lysine, myo-inositol and shikimate at final concentrations of 1.0 mM in D_2O .

An extract from human prostate cancer cell line DU145 (obtained from the American Type Culture Collection (ATCC), www.atcc.org) was obtained as follows. The cells were cultured at 37°C in 5% CO_2 incubator in DMEM medium supplemented with 10% fetal bovine serum and penicillin/streptomycin. About 7×10^7 cells were lysed by sequentially adding 3 ml each of methanol, chloroform and water.⁸ The sample was vortexed vigorously after the addition of each solvent, and the final mixture was stored at -20°C overnight for phase separation. The aqueous phase, which is completely separated from the organic phase by centrifugation at 10,000 g for 40 minutes, is lyophilized and dissolved in D_2O for NMR experiments.

NMR data collection and analysis

2D heteronuclear ^1H - ^{13}C HSQC-TOCSY NMR data were collected at 800 MHz using a cryogenic probe and 5-mm NMR tubes. The sample temperature was maintained at 298 K. MLEV-17 mixing sequence with 300 ms mixing time was used. The spectral widths were 8012.8 Hz (^1H), 32206.1 Hz (^{13}C) for the test mixtures and 10000.0 Hz (^1H), 32206.1 Hz (^{13}C) for the DU145 sample. Data were collected using 2048 t_2 and 1024 t_1 (complex) data points with 16 to 48 scans per t_1 increment. The NMR data were Fourier transformed and phase and baseline corrected by NMRPipe.²⁵ The HSQC-TOCSY spectrum of the model

mixture was collected in 22 hours for and the one for the DU145 cell line extract in 70 hours.

For the heteronuclear HSQC-TOCSY 2D Fourier transform NMR spectrum (**F**), the direct ^1H covariance NMR spectrum $\mathbf{C}_\text{H}^2$ and indirect ^{13}C covariance NMR spectrum $\mathbf{C}_\text{C}^2$ can be obtained by the following matrix operations

$$\mathbf{C}_\text{H}^2 = \mathbf{F}^T \mathbf{F} \quad \text{and} \quad \mathbf{C}_\text{C}^2 = \mathbf{F} \mathbf{F}^T \quad (1)$$

These spectra were used to compute the importance index of each row. (Note that in contrast to standard applications of direct and indirect covariance NMR no matrix square root operation is applied.)²⁶ The ^1H importance index belonging to a given row of $\mathbf{C}_\text{H}^2$, which is defined as the sum of all elements of the corresponding row of $\mathbf{C}_\text{H}^2$, is a measure of the cumulative overlap of this trace with all other traces of $\mathbf{C}_\text{H}^2$.¹⁵ The ^1H importance index profile, which resembles a 1D ^1H spectrum with altered peak amplitudes, is then peak picked and traces of **F** are selected at the picked ^1H positions along the vertical ^{13}C (ω_1) dimension. After each of these traces has been normalized, i.e. $\mathbf{v}' = \mathbf{v}/(\mathbf{v} \cdot \mathbf{v})^{1/2}$, the traces are clustered using their inner products as the similarity metric. For each cluster, the trace with the lowest importance index is picked as the cluster representative, because of its low probability to be contaminated by other traces. All normalized traces whose inner products with the cluster representative are 0.4 or larger are assigned to this cluster. In this way, the likelihood is optimized that each selected ^{13}C trace corresponds to the 1D ^{13}C spectrum of an individual component (or spin system) free of spurious contributions from other spin systems.

The same procedure is then applied to $\mathbf{C}_\text{C}^2$ with the roles of ^1H and ^{13}C resonances interchanged. This yields a set of unique 1D ^1H NMR traces corresponding to individual components (or spin systems).

The most straightforward treatment of the ^{13}C and ^1H traces, which is pursued here, consists of separately submitting the traces for screening against a database of NMR spectra. The return of the same metabolite as a top match for both a ^{13}C and ^1H trace indicates a high probability for the metabolite to be present in the mixture.

It is noted that critical processing and analysis steps described above can be performed using the COLMAR suite of web servers (<http://spinportal.magnet.fsu.edu/>). This concerns direct and indirect covariance processing of the HSQC-TOCSY spectrum by the COLMAR *covariance* web server and querying of the traces against a NMR spectral database using COLMAR *query*.²³

Results

Analysis of model mixture

The heteronuclear DemixC method was first tested on the model mixture containing the five common metabolites carnitine, glucose, lysine, myo-inositol, and shikimate. The ^1H - ^{13}C HSQC-TOCSY NMR spectrum of this mixture is shown in Figure 1.

From the HSQC-TOCSY spectrum, the importance index for the traces of $\mathbf{C}_\text{H}^2$ and $\mathbf{C}_\text{C}^2$ are computed and plotted in Figures 2A,B. The HSQC-TOCSY traces belonging to the peaks picked in the importance index profile are then subjected to the clustering analysis described in the Methods section.¹⁵ For each cluster, the trace with the lowest importance index above a predefined threshold of 4% of the maximal ^1H peak and 2.5% of the largest ^{13}C peak is selected. As an example, Figure 2C shows the lysine ^{13}C trace of the HSQC-TOCSY

spectrum associated with a peak in the ^1H importance index determined from C_H^2 in Figure 2A at 1.88 ppm (vertical arrow), whereas the ^1H trace lysine trace (Figure 2D) belongs to the ^{13}C importance index peak at 24.04 ppm (vertical arrow) of Figure 2B. Peaks seen in the 1D traces are also found in the importance index profile of the same spin species as indicated by diagonal arrows.

The selected ^1H and ^{13}C traces are plotted in Figure 3. Pairs of ^1H and ^{13}C traces are assigned to each other if they belong to the same compound according to COLMAR query. Examples of COLMAR query outputs are shown in the Supporting Information for selected ^{13}C and ^1H traces. The top five spectra in Figure 3 are the 1D ^{13}C (Panel A) and ^1H (Panel B) reference spectra of the mixture components taken from the BMRB database.²⁰ The six traces at the bottom are sorted and numbered according to their importance index and labeled with the first character of the compound name. The slowly interconverting α - and β -forms of glucose are represented by separate traces (bottom traces 3 and 5), respectively, whereas the BMRB reference spectrum represents their superposition.²² The agreement between peak positions of the HSQC-TOCSY traces and the corresponding reference spectra of the pure components is very good, which clearly demonstrates the suitability of this method. Differences in the peak amplitudes are observed, which are caused by incomplete magnetization transfer among spins during TOCSY mixing at a single mixing time (300 ms), but they do not affect spin system identification. Submission of the traces to the COLMAR query server returns for each of the ^1H and ^{13}C DemixC traces the correct compound as top hits (Figure 3).

DU145 human prostate cancer cell line analysis

The ^1H - ^{13}C HSQC-TOCSY NMR spectrum of the mixture of DU145 human prostate cancer cell line is shown in Figure 4. The importance index for C_H^2 and C_C^2 are computed and plotted in Figure 5A and Figure 5B. The complete sets of picked ^{13}C and ^1H traces are plotted in Figure 5C and 5D, respectively. The individual ^1H traces are searched against the BMRB database using the COLMAR query web portal at <http://spinportal.magnet.fsu.edu/webquery/webquery.html> with the identified compound names labeled in Figure 5. The following components are identified both based on their ^1H and ^{13}C spectra: reduced glutathione, taurine, glutamate, lactate, and myo-inositol. Although two compounds labeled as compound1 and compound2 are not contained in the BMRB database, their ^{13}C and ^1H traces can be unambiguously assigned to each other. Carnitine only shows the $\text{N}-(\text{CH}_3)_3$ sub-spin system, which involves 3 equivalent carbons and 9 protons. The other carnitine spin system is not detected, presumably because it is too weak. Note that reduced glutathione appears in several traces because it contains multiple spin systems, which do not interchange significant amounts of magnetization during TOCSY mixing. Glycerol and glycine show up only in ^1H traces, not in ^{13}C traces. This is because the ^1H line widths (Figure 5A) are broader and have multiplet splittings due to $J(^1\text{H}, ^1\text{H})$ -couplings whereas those of the ^{13}C lines (Figure 5B) are narrow singlets. As a consequence, the ^1H importance index (Figure 5A) exhibits more overlap and therefore allows fewer unique traces to be extracted along the ^{13}C dimension in the HSQC-TOCSY spectrum. Closer inspection of the HSQC-TOCSY spectrum confirms the presence of ^{13}C signals of glycerol and glycine.

Discussion

The results demonstrate the effectiveness of the heteronuclear DemixC method to analyze the composition of a complex metabolic mixture of practical importance, in this case a DU145 human prostate cancer cell extract. The method allows the easy extraction of unique 1D ^{13}C and ^1H cross sections that can be directly submitted to a database query server, such as COLMAR query, for compound identification. The redundancy of the information from

both queries enhances the reliability and robustness of the compound identification procedure and allows one to establish in a self-consistent way the composition of the mixture. Those compounds that score well in both queries have a high probability to be present in the mixture.

The heteronuclear HSQC-TOCSY DemixC method has two advantages over the homonuclear TOCSY DemixC approach: 1) the ^{13}C dimension has much narrower line widths than the ^1H dimension, which results in a larger number of unique ^1H traces and 2) the components identified based on ^1H spectra can be cross-validated by independent identification from ^{13}C 1D spectra and *vice versa*. The main disadvantage of the heteronuclear HSQC-TOCSY DemixC method is that at natural ^{13}C abundance the spectrum has lower signal-to-noise than the corresponding homonuclear ^1H - ^1H TOCSY and hence the HSQC-TOCSY requires either higher sample concentration or longer measurement time or both. In the case of the cancer cell line extract the signal-to-noise ratios of the two experiments differ by a factor of 68.

Recently, metabolite components were directly identified from 2D ^1H - ^{13}C HSQC spectra.⁸⁻¹⁰ The ^1H - ^{13}C HSQC-TOCSY experiment can be considered as an extension of the ^1H - ^{13}C HSQC experiment to which it reduces in the limit of vanishing mixing time, $\tau_m = 0$. For finite τ_m it produces for a ^{13}C resonance a cross peak to the directly attached proton as well as to other protons that belong to the same spin system. The cross peak amplitudes vary as a function of the spin-coupling network and the TOCSY mixing time. For typical mixing times, many of the stronger HSQC-TOCSY peaks have amplitudes that are of the order of 50% of the HSQC peak amplitudes. This sensitivity loss is compensated for by access to long-range spin-connectivity information displayed by the HSQC-TOCSY spectrum that is not available via the HSQC spectrum. The redundant long-range correlation information is utilized in DemixC to reconstruct 1D ^1H and ^{13}C spectra of individual compounds for the 1D screening against metabolite databases.

Over the past decade, cancer cell lines have served as model systems in metabolic studies of cancer²⁷ and NMR-based specific metabolites studies have contributed important information to the understanding of the underlying biochemical events. However, it is still challenging to apply NMR in comprehensive metabolomic studies partly because of strong peak overlaps in NMR spectra of complex mixtures. The heteronuclear DemixC method is particularly promising for the initial study of a complex metabolic sample of largely unknown composition, for instance cancer cell extracts. The retrieved compound information serves then as a starting point for a subsequent quantitative metabolomics study of a series of related samples whose metabolite concentrations may quantitatively differ due to genetic, behavioral, and environmental factors using, for example, the more sensitive and therefore faster homonuclear ^1H - ^1H TOCSY-based DemixC approach. Due to its general nature, the heteronuclear spectral deconvolution strategy introduced here is applicable in a wide range of contexts, including metabolic biomarker discovery, drug research, impurity profiling, food sciences, and environmental sample analysis.

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 to R.B.). The NMR experiments were conducted at the National High Magnetic Field Laboratory (NHMFL) supported by cooperative agreement DMR 0654118 between the NSF and the State of Florida.

References

1. Nicholson JK, Wilson ID. *Nature Reviews Drug Discovery*. 2003; 2:668–676.
2. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. *Trends in Biotechnology*. 2004; 22:245–252. [PubMed: 15109811]
3. Fiehn O. *Plant Molecular Biology*. 2002; 48:155–171. [PubMed: 11860207]
4. Nordstrom A, O'Maille G, Qin C, Siuzdak G. *Anal. Chem.* 2006; 78:3289–3295. [PubMed: 16689529]
5. Lenz EM, Wilson ID. *J. Proteome Research*. 2007; 6:443–458. [PubMed: 17269702]
6. Johnson CS. *Progress in Nucl. Magn. Reson. Spectrosc.* 1999; 34:203–256.
7. Schroeder FC, Gibson DM, Churchill AC, Sojikul P, Wursthorn EJ, Krasnoff SB, Clardy J. *Angew. Chem. Int. Ed.* 2007; 46:901–904.
8. Hyberts SG, Heffron GJ, Tarragona NG, Solanky K, Edmonds KA, Luithardt H, Fejzo J, Chorev M, Aktas H, Colson K, Falchuk KH, Halperin JA, Wagner G. *J. Am. Chem. Soc.* 2007; 129:5108–5116. [PubMed: 17388596]
9. Lewis IA, Schommer SC, Hodis B, Robb KA, Tonelli M, Westler WM, Suissman MR, Markley JL. *Anal. Chem.* 2007; 79:9385–9390. [PubMed: 17985927]
10. Xi YX, de Ropp JS, Viant MR, Woodruff DL, Yu P. *Analytica Chimica Acta*. 2008; 614:127–133. [PubMed: 18420042]
11. Kessler H, Mrona S, Gemmecker G. *Magn. Reson. Chem.* 1991; 29:527–557.
12. Sandusky P, Raftery D. *Anal. Chem.* 2005; 77:7717–7723. [PubMed: 16316181]
13. Zhang F, Brüschweiler R. *Chemphyschem*. 2004; 5:794–796. [PubMed: 15253306]
14. Cloarec O, Dumas ME, Craig A, Barton RH, Trygg J, Hudson J, Blancher C, Gauguier D, Lindon JC, Holmes E, Nicholson J. *Anal. Chem.* 2005; 77:1282–1289. [PubMed: 15732908]
15. Zhang F, Brüschweiler R. *Angew. Chem. Int. Ed.* 2007; 46:2639–2642.
16. Braunschweiler L, Ernst RR. *J. Magn. Reson.* 1983; 53:521–528.
17. Brüschweiler R. *J. Chem. Phys.* 2004; 121:409–414. [PubMed: 15260561]
18. Brüschweiler R, Zhang F. *J. Chem. Phys.* 2004; 120:5253–5260. [PubMed: 15267396]
19. Trbovic N, Smirnov S, Zhang F, Brüschweiler R. *J. Magn. Reson.* 2004; 171:277–283. [PubMed: 15546754]
20. Seavey BR, Farr EA, Westler WM, Markley JL. *J. Biomol. NMR*. 1991; 1:217–236.
21. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly MA, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, MacInnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L. *Nucleic Acids Research*. 2007; 35:D521–D526. [PubMed: 17202168]
22. Zhang F, Dossey AT, Zachariah C, Edison AS, Brüschweiler R. *Anal. Chem.* 2007; 79:7748–7752. [PubMed: 17822309]
23. Robinette SL, Zhang F, Bruschweiler-Li L, Brüschweiler R. *Anal. Chem.* 2008; 80:3606–3611. [PubMed: 18422338]
24. Stone KR, Mickey DD, Wunderli H, Mickey GH, Paulson DF. *Int. J. Cancer*. 1978; 21:274–281. [PubMed: 631930]
25. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. *J. Biomol. NMR*. 1995; 6:277–293. [PubMed: 8520220]
26. Zhang F, Brüschweiler R. *J. Am. Chem. Soc.* 2004; 126:13180–13181. [PubMed: 15479045]
27. Griffin JL, Shockcor JP. *Nature Reviews Cancer*. 2004; 4:551–561.

Figure 1.
 ^1H - ^{13}C HSQC-TOCSY NMR spectrum of mixture containing carnitine, glucose, lysine, myo-inositol and shikimate in aqueous solution.

Figure 2.

Importance index profile calculated from HSQC-TOCSY spectrum of model mixture (A) along ^1H dimension and (B) along ^{13}C dimension. At the peak positions picked in Panels A and B (indicated by 'x' symbols) a subset of traces is extracted in the HSQC-TOCSY spectrum, such as the Lysine carbon trace obtained from the ^1H importance index (Panel C) and the Lysine proton trace obtained from ^{13}C importance index (Panel D). The vertical arrows indicate the importance index peak positions where the traces were extracted in the HSQC-TOCSY. The diagonal arrows indicate the connections between the importance index and the traces picked linking the ^{13}C (C) and ^1H (D) traces to the same compound (Lysine).

Figure 3.

(A) ^{13}C and (B) ^1H proton traces obtained by the heteronuclear DemixC method applied to the HSQC-TOCSY spectrum of Figure 1. Top five spectra (labeled with full compound names): standard reference spectra taken from the BMRB databank Six bottom spectra: DemixC picked traces sorted according to the importance index (with lower index at the bottom). The traces are labeled with the first character of compound names. Bottom traces 3 and 5 correspond to α - and β -Glucose, respectively.

Figure 4.
 ^1H - ^{13}C HSQC-TOCSY NMR spectrum of extracts of human prostate cancer cell line DU145.

Figure 5.

(A) ^{13}C and (B) ^1H traces obtained by the heteronuclear DemixC method applied to the HSQC-TOCSY of Figure 4. Peaks in Panels A and B that are picked to represent traces are indicated by 'x' symbols. (C) ^{13}C and (D) ^1H traces obtained from proton and carbon importance index profiles of Panels A and B. All compound names are indicated except for two compounds that are not in the BMRB database (labeled as compound1 and compound2).