Published in final edited form as:
J Am Chem Soc. 2010 December 1; 132(47): 16922–16927. doi:10.1021/ja106781r.

Simultaneous De Novo Identification of Molecules in Chemical Mixtures by Doubly Indirect Covariance NMR Spectroscopy

Fengli Zhang[1], **Lei Bruschweiler-Li**[1,2], and **Rafael Brüschweiler**[1,2,*]

[1] National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL 32310, U.S.A

[2] Chemical Sciences Laboratory, Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306, U.S.A

Abstract

The detailed characterization of complex molecular mixtures plays a key role in many areas of modern Chemistry. Here we report a novel NMR spectroscopic method that deconvolutes a complex mixture of organic molecules simultaneously into individual components and depicts their chemical structure without requiring physical separation of the components. Doubly indirect covariance spectroscopy is introduced and applied to 2D ^{13}C-^{1}H HSQC and 2D ^{1}H-^{1}H COSY spectra, which results in a ^{13}C-^{13}C 2D spectrum with unprecedented high resolution. This reconstituted spectrum is indeed a carbon-connectivity map that can be directly analyzed with basic graph theory to obtain the skeletal structures of individual mixture components or their fragments. The method is demonstrated for a model mixture and a natural product mixture extracted from cancer cells. Its suitability for automation makes this approach attractive for the analysis of a broad range of mixtures of natural or synthetic products.

Keywords

Complex mixture analysis; natural and synthetic products; indirect covariance NMR; skeletal molecular structure determination

Introduction

Many natural and synthetic products are components of complex chemical mixtures whose analysis can be a daunting task. NMR spectroscopy has proven to be a particularly promising tool, since it does not require the time-consuming and potentially costly physical separation of these complex mixtures into individual components. Identification of the molecules in a sample can be facilitated considerably and speeded up by screening the NMR spectra against an NMR spectral database.[1–3] Although NMR spectral databases are steadily growing, a large number of organic compounds are not indexed (yet), a situation that will continue into the foreseeable future. *Ab initio* analysis of the components in a complex mixture requires a different strategy. While a broad range of homo- and heteronuclear NMR methods exist that report on the intramolecular nuclear spin-spin connectivities,[4] the

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

Supporting Information Available: Complete Refs. [1] and [2]. This material is available free of charge via the Internet at http://pubs.acs.org.

simultaneous analysis of ^{1}H and ^{13}C correlation spectra is not straightforward, because it usually requires labor-intensive interactive analysis. By contrast, the ^{13}C-only 2D INADEQUATE spectrum[5] permits the tracing of the molecular carbon skeleton using a minimal number of NMR cross-peaks. Unfortunately, the intrinsically low sensitivity of this experiment precludes its application to mixtures with components in the mM and sub-mM range at ^{13}C natural abundance.

We present a novel strategy for the *ab initio* characterization of molecules in complex mixtures, which is based on a combination of homo- and heteronuclear NMR experiments. The method produces, like 2D INADEQUATE, a ^{13}C-only correlation spectrum, but at the much higher sensitivity of the ^{13}C-^{1}H HSQC experiment. The method uses the principles of covariance NMR spectroscopy[6–8] and its generalizations.[9–11] Covariance NMR spectroscopy exploits the favorable spectroscopic properties, such as increased spectral resolution or spin-spin correlation information of certain projections of higher dimensional spectra, which are hard to obtain from direct measurements. In particular, indirect covariance NMR produces a ^{13}C-^{13}C total correlation spectrum (TOCSY) (represented by matrix **C**) from a ^{13}C-^{1}H HSQC-TOCSY (matrix **F**) by standard linear algebraic matrix operations: $\mathbf{C} = (\mathbf{F}\cdot\mathbf{F}^{T})^{1/2}$.[9] Recently, indirect covariance NMR has been extended to the concatenation of two different spectra, such as 2D HSQC, TOCSY, COSY, or NOESY spectra.[12,13]

Materials and Methods

NMR Samples

An isoleucine sample was prepared consisting of 60 mM solution of isoleucine in D_2O. A metabolite model mixture was prepared consisting of 1.0 mM final concentrations of carnitine, isoleucine, lysine, and shikimate dissolved in D_2O. An extract from human prostate cancer cell line DU145 (from the American Type Culture Collection (ATCC), www.atcc.org) was obtained as follows:[14,15] 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. 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 experiments and processing

2D ^{1}H-^{1}H 2QF-COSY[16] datasets were collected with $N_1 = 1024$ and $N_2 = 1024$ complex data points and 2D ^{13}C-^{1}H HSQC spectra[17] with $N_1 = 1024$ and $N_2 = 1024$ complex data points. For the isoleucine sample, the 2QF-COSY was collected in 10.9 h with spectral widths of 4000.0 Hz in both dimensions and the HSQC spectrum in 5.3 h with spectral widths of 4000.0 Hz (^{1}H) and 20122.8 Hz (^{13}C). For the model mixture, the 2QF-COSY was collected in 10.2 h with spectral widths of 6393.8 Hz in both dimensions and the HSQC spectrum in 5.0 h with spectral widths of 6393.8 Hz (^{1}H) and 20122.8 Hz (^{13}C). The DOSY spectrum was collected in 64.6 h with a ^{13}C spectral width of 20161.2 Hz with 48 different gradient field strengths from 0.68 to 32.35 G/cm. The DOSY spectrum was processed using Bruker Topspin 2.1 software. Peak maxima and the errors in the diffusion coefficients, determined as the standard deviation of the cross section along the diffusion dimension, are indicated in Figure 2D. For the cancer cell sample, the 2QF-COSY was collected in 40 h with spectral widths of 10,000.0 Hz in both dimensions and the HSQC spectrum in 38.8 h with spectral widths of 10,000.0 Hz (^{1}H) and 32,197.2 Hz (^{13}C). All NMR spectra were collected at 800 MHz proton frequency at 298 K. The NMR data was zero-filled, Fourier

transformed, phase and baseline corrected using NMRPipe,[18] and converted to a Matlab-compatible format for further processing and analysis. The real part of the 2D FT COSY spectrum F was covariance processed using regularization,[19] $\mathbf{Y} = \mathrm{abs}((\mathbf{F}_\alpha{}^{\mathrm{T}}\mathbf{F}_\alpha)^{1/2} - \alpha\mathbf{1})$ where "abs" indicates the absolute value of each matrix element, $\mathbf{F}_\alpha = \mathbf{F} + \alpha\mathbf{1}$ where **1** is the unity matrix and the regularization factor α was set to 100 times the negative of the sum of the real parts of the eigenvalues of **F**.

Doubly indirect ^{13}C-^{13}C covariance spectroscopy

This method first constructs from a ^{13}C-^{1}H 2D HSQC spectrum, represented by the $N_1 \times N_2$ matrix **H**, and a ^{1}H-^{1}H regularized covariance 2QF-COSY spectrum, represented by the $N_2 \times N_2$ matrix **Y**, the doubly indirect covariance spectrum $\mathbf{C}(\omega_1^C, \omega_2^C)$ according to:

$$C_{ij} = \sum_{k,l=1}^{N_2} H_{ik} Y_{kl} H_{jl} \tag{1}$$

where $i, j = 1,\ldots,N_1; k,l = 1,\ldots,N_2$. The double sum amounts to a projection that removes the two ^{1}H dimensions leaving only the two ^{13}C dimensions. Spectrum C can be compactly expressed in matrix notation:

$$\mathbf{C} = \mathbf{H} \cdot \boldsymbol{Y} \cdot \mathbf{H}^{\mathbf{T}} \tag{2}$$

where superscript T denotes the matrix transpose. The ^{13}C-^{13}C correlation spectrum **C** represents a doubly indirect covariance spectrum that correlates two carbon resonances whose directly attached protons display a cross-peak in the 2QF-COSY spectrum (due to a direct scalar J-coupling). Each cross-peak of **C** corresponds to a carbon-carbon chemical bond. Spectrum **C** is symmetric and has the appearance of a ^{13}C-^{13}C COSY-type spectrum or a covariance-processed INADEQUATE.[20] However, since spectrum **C** does not require that correlated ^{13}C spins physically reside on the same molecule (which has a very low probability ($p \approx 10^{-4}$) for natural abundance ^{13}C samples), it is not subject to the sensitivity limitations of the INADEQUATE experiment. Spectrum **C** is also more sensitive than the ADEQUATE experiment,[21] which detects protons but also requires that correlated ^{13}C spins reside on the same molecule. Assessment of the absolute sensitivity of spectrum C is not straightforward due to thresholding applied to the input spectra in Eq. (2) and due to the non-linear propagation of signals and noise during matrix multiplication. Finally, indirect covariance HSQC-TOCSY spectra exhibit relay peaks, making it hard to establish a connectivity matrix with connectivities only between directly bonded carbons.

Since each cross-peak consists along both ^{13}C-frequency dimensions of a single line free of multiplet splittings, the resulting spectrum has very high resolution. This renders it ideally suited for downstream automated analysis of the carbon-carbon bond network of the mixture components. Using basic graph theory, matrix **C** can be converted into a connectivity matrix where each ^{13}C resonance corresponds to a node and each cross-peak to an edge. The display of the connectivity matrix as a graph yields the skeletal formula of the chemical compound(s) of the sample in a straightforward manner. Figure 1 illustrates the method for isoleucine serving as a test system. It shows spectrum **C** of isoleucine together with its HSQC and COSY parent spectra. The graph theoretical treatment of **C** yields the characteristic branched carbon skeletal structure of isoleucine (Panel E) where nodes 1, 2, and 3 represent C_α, C_β, and C_γ carbons, respectively, and nodes 4, 5 represent the two methyl groups.

In both the COSY spectra **Y** and the HSQC spectra **H**, the residual water line along ω_1 was removed. In the COSY spectrum **Y**, 17 points right and left of the diagonal are set to zero, which correspond to 33 Hz (isoleucine sample), 53 Hz (metabolite model mixture), and 83 Hz (DU145 extract), respectively. The results are rather insensitive to this parameter as long as no cross-peaks are removed in the COSY spectrum. The removal prevents partially overlapping proton resonances along the diagonal from causing false ^{13}C-^{13}C cross-peaks in **C**. The points that lie exactly on the diagonal are retained as they are responsible for the diagonal peaks in **C**. For each row $\mathbf{C} = \mathbf{H}\cdot\mathbf{Y}\cdot\mathbf{H}^T$ (Eq. (2)) a 1D "node index" is calculated as the sum of the elements of the row. The ^{13}C nodes of the connectivity graph are obtained by peak picking the node index with 0.1% of strongest peak as lower threshold. A connectivity matrix is then obtained for these nodes where an edge is introduced between two nodes if C_{ij} is larger than a user-defined threshold.

In case of spectral overlap between 2 protons, matrices **H** and **Y** need to be modified to prevent false correlations. This is achieved by considering the multiplet fine structures of overlapping protons, which in general differ sufficiently to permit discrimination of false correlations. This leads to moment-based filtering[22] as is explained in the following:

To identify overlapping protons the indirect covariance matrix **S** of the HSQC spectrum is calculated according to $\mathbf{S} = (\mathbf{H}\cdot\mathbf{H}^T)^{1/2} = \mathbf{U}\cdot\mathbf{D}\cdot\mathbf{U}^T$ where **U** is an orthogonal matrix from the singular value decomposition $\mathbf{H} = \mathbf{U}\cdot\mathbf{D}\cdot\mathbf{V}^T$. Off-diagonal peaks of **S** connect the carbon spins whose protons overlap (note that due to the high resolution of the ^{13}C-spectrum without multiplet structure the chance of accidental ^{13}C overlaps is very low and in fact much lower than for $^1\mathbf{H}$ overlap). For every column in **H** that shows a cross-peak in S, the following procedure is applied:

1. All peaks in the column of **H** and the same column of **Y** are picked. To identify which peaks of **Y** belong to which of the two peaks in **H**, the first and second moments μ and σ^2 of the peak shape along the proton dimension are determined for each peak and used in the difference metric $\Delta = [(\mu_a - \mu_b)^2 + 0.2\cdot(\sigma_a - \sigma_b)^2]$. Peaks that belong to the same proton show similar μ and σ and therefore have a low Δ that permits identification of false correlations in matrix **C** as follows.
2. Matrices **H** and **Y** are modified to **H′** and **Y′** by sequentially going through the list of ^{13}C resonances attached to a proton X that overlaps with one (or several) protons. All non-overlapping peaks are retained and all cross peaks in **H** and **Y** that belong to proton(s) that overlap(s) with proton X are removed. The doubly-indirect covariance spectrum is then determined based on $\mathbf{C}' = \mathbf{H}'\cdot\mathbf{Y}'\cdot\mathbf{H}'^T$.
3. Step 2 is repeated until the ^{13}C resonance list is depleted. Molecular connectivity graphs of all matrices **C′** are generated and a minimal set of disconnected graphs is displayed corresponding to individual molecules or their fragments.

Note that geminal protons (belonging to the same ^{13}C node) are automatically and correctly handled as they do not impact the carbon topological graph.

Results

Application to a metabolite mixture

The high spectral resolution of **C** permits application of the method in 'parallel mode' for the simultaneous analysis of a complex chemical mixture containing multiple chemical components. This is illustrated for a metabolite model mixture consisting of carnitine, isoleucine, lysine, and shikimate dissolved in D_2O. Figure 2 shows the corresponding doubly indirect covariance spectrum **C** (Panel C) together with its 2QF-COSY (Panel A) and ^{13}C-^1H HSQC (Panel B) parent spectra. Graph theoretical analysis yields the carbon

traces of the two molecules carnitine and shikimate (compounds a,c in Panel F). Since the proton resonances $H_{\gamma 2}$ of lysine and $H_{\gamma 12}$ of isoleucine overlap, as indicated by arrows in Panels A and B, these two molecules emerge as a single graph. Spectral overlap can be addressed in two different ways: i) by diffusion-ordered spectroscopy (DOSY)[23] that separates spectra according to differential translational diffusion or ii) by our filtering method based on differences in the cross-peak fine structure of the parent spectra.

The ^{13}C-detected INEPT DOSY experiment[24] reveals differential translational diffusion rates of individual resonances as shown in Figure 2 (Panel D). The DOSY information permits partial isolation of the HSQC subspectra for each of the mixture components, which can then be applied separately together with the COSY spectrum of the mixture in Eq. (2). The resulting covariance spectra **C** are then subjected to graph theoretical analysis yielding the individual skeletal structures shown in Figure 2G. It should be noted that the uncertainties in the diffusion coefficients of individual resonances introduce some ambiguities and the DOSY spectrum on its own would not allow the inference that there are 4 components. However, taking the data in conjunction with the (unfiltered) ^{13}C-^{13}C covariance spectrum clearly simplifies the analysis, because the latter helps define the peak grouping by identifying those peaks that do not come from the same species.

The spectral moment-based filtering method, which has the advantage that it does not require the collection of an additional DOSY spectrum, suppresses false correlations based on deviations in peak positions and peak fine structure. This leads to a ^{13}C-^{13}C graph connectivity matrix **C** that is free of spurious connectivities yielding the skeletal formulae of all the mixture components (Figures 2F and 2G). Because Eq. (2) uses ^{1}H-^{1}H COSY cross-peaks to connect adjacent C-H moieties with each other, the covariance matrix **C** correlates carbons that possess directly attached protons. Hence, the carboxyl carbon of the Lys and Ile amino acids and any other proton-free carbons, such as carbonyl or quarternary carbons, are not displayed by **C**.

Application to a natural product mixture

Figure 3 shows the doubly indirect covariance method after moment filtering applied to the extract of prostate cancer cell line DU145. It readily reveals the skeletal structure of lactate, taurine, myo-inositol, glutamate, glutathione reduced, and ethanol. With the exception of ethanol, which was added during sample preparation, these compounds are in agreement with previous results that used DemixC clustering followed by searching of the BMRB using the COLMAR query web server.[3,25] By contrast, in the present study the mixture components are characterized *de novo* in terms of their carbon skeletons, which does not require a database search.

Discussion

The doubly indirect covariance NMR approach introduced here is particularly suitable for the simultaneous analysis of complex molecular mixtures of unknown composition. With the commonly used and rather sensitive 2D NMR COSY and HSQC spectra as input, it produces very high-resolution carbon-carbon connectivity maps, which can be directly translated into carbon skeletal structures and substructures of the molecules present in the mixture. The high sensitivity of the approach presented here afforded by proton detection, however, precludes the direct analysis of connectivities to non-protonated carbons. Such connectivities could be established by replacing one of the flanking HSQC experiments by an HMBC experiment[26] that probes long-range ^{1}H-^{13}C correlations also to non-protonated carbons. Furthermore, it can be helpful to map the full spin system via medium and long-range spin-spin correlations by replacing in Eq. (2) the 2D COSY by a 2D TOCSY.[27]

The topological similarity of different graphs does not pose a problem as long as they can be disambiguated based on their associated chemical shifts (first moments) or linewidths (second moments). In addition, information about carbon multiplicity that is contained in the HSQC spectrum can be added to the carbon skeleton. The ^{13}C chemical shift information is directly useful to obtain information about functional groups and the presence of chemical bonds to atoms, such as nitrogen and oxygen, which are not displayed in the COSY and HSQC spectra. Moreover, ^{1}H chemical shifts can be readily extracted by tracing the ^{13}C-^{13}C graph in the HSQC spectrum,[28] which will further benefit the assignment of the chemical components. If multiple graphs belong to the same compound they can be connected by the use of long-range spin correlations as provided by the HMBC experiment. The same applies to the establishment of connectivities to non-protonated carbons, provided that the associated sensitivity loss can be afforded. Together with other NMR-derived parameters, such as NOE distances and scalar coupling-derived dihedral angles, this information can serve as a starting point for 3D structure elucidation of the molecules of interest.

The doubly indirect covariance HSQC-COSY-HSQC spectrum possesses ultra-high spectral resolution along both ^{13}C frequency dimensions, which greatly facilitates the subsequent analysis by peak picking and graph theory. Because complete or partial overlap of the underlying ^{1}H resonances could lead to false correlations, a filtering step has been introduced here that compares the first and second moments of associated COSY and HSQC cross-peaks. It suppresses resulting peaks in the doubly indirect covariance spectrum, if the weighted squared difference of their average positions μ and peak widths σ exceeds a predefined threshold. While the 4^{th} moments (kurtosis) can be included in an analogous manner, they turned out not to be required for the systems studied here. In case that two overlapping proton resonances have within a given tolerance the same μ and σ, the filtering method will not be able to unambiguously identify the false correlation in **C** between the ^{13}C spins belonging to these protons. Although this situation was not encountered for the samples presented here, its likelihood should increase for larger molecules with broader lines and unresolved J-splittings as well as for chemically closely related species.

The doubly indirect covariance HSQC-COSY-HSQC spectrum constitutes a new general building block for the analysis of molecules in complex mixtures by NMR. It establishes correlations between carbon spins that share a chemical bond. The need for the accurate analysis of mixtures of unknown composition with components that are not necessarily contained in NMR spectral databases is prevalent in many areas of modern chemistry and biochemistry. This includes natural and chemical synthesis, metabolomics, cellular biochemistry, nutritional sciences, toxicology, and chemical ecology to name only a few. The doubly indirect covariance NMR approach introduced here will facilitate and speed-up this pivotal process.

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 GM066041). 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. Ulrich EL, et al. Nucleic Acids Res. 2008; 36:D402–D408. [PubMed: 17984079]
2. Wishart DS, et al. Nucleic Acids Res. 2007; 35:D521–D526. [PubMed: 17202168]

3. Zhang F, Robinette SL, Bruschweiler-Li L, Brüschweiler R. Magn Reson Chem. 2009; 47:S118–S122. [PubMed: 19634130]
4. Friebolin, H. Basic One- and Two-Dimensional NMR Spectroscopy. Wiley-VCH; Weinheim: 2005.
5. Bax A, Freeman R. J Magn Reson. 1981; 44:542–561.
6. Brüschweiler R, Zhang F. J Chem Phys. 2004; 120:5253–5260. [PubMed: 15267396]
7. Brüschweiler R. J Chem Phys. 2004; 121:409–414. [PubMed: 15260561]
8. Trbovic N, Smirnov S, Zhang F, Brüschweiler R. J Magn Reson. 2004; 171:277–283. [PubMed: 15546754]
9. Zhang F, Brüschweiler R. J Am Chem Soc. 2004; 126:13180–13181. [PubMed: 15479045]
10. Blinov KA, Larin NI, Williams AJ, Zell M, Martin GE. Magn Reson Chem. 2006; 44:107–9. [PubMed: 16392113]
11. Kupce E, Freeman R. J Am Chem Soc. 2006; 128:6020–6021. [PubMed: 16669655]
12. Blinov KA, Larin NI, Williams AJ, Mills KA, Martin GE. J Heterocyclic Chemistry. 2006; 43:163–166.
13. Blinov KA, Williams AJ, Hilton BD, Irish PA, Martin GE. Magn Reson Chem. 2007; 45:544–546. [PubMed: 17437315]
14. 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]
15. Zhang F, Bruschweiler-Li L, Robinette SL, Brüschweiler R. Anal Chem. 2008; 80:7549–7553. [PubMed: 18771235]
16. Rance M, Sorensen OW, Bodenhausen G, Wagner G, Ernst RR, Wüthrich K. Biochem Biophys Res Comm. 1983; 117:479–485. [PubMed: 6661238]
17. Bodenhausen G, Ruben DG. Chem Phys Lett. 1980; 69:185–189.
18. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. J Biomol NMR. 1995; 6:277–93. [PubMed: 8520220]
19. Chen Y, Zhang F, Snyder DA, Gan Z, Brüschweiler-Li L, Brüschweiler R. J Biomol NMR. 2007; 38:73–77. [PubMed: 17351818]
20. Zhang F, Trbovic N, Wang JB, Brüschweiler R. J Magn Reson. 2005; 174:219–222. [PubMed: 15862237]
21. Reif B, Kock M, Kerssebaum R, Kang H, Fenical W, Griesinger C. J Magn Reson A. 1996; 118:282–285.
22. Bingol K, Salinas RK, Brüschweiler R. J Phys Chem Lett. 2010; 1:1086–1089. [PubMed: 20436937]
23. Johnson CS. Progress in Nucl Magn Reson Spectrosc. 1999; 34:203–256.
24. Wu DH, Chen AD, Johnson CS. J Magn Reson A. 1996; 123:215–218.
25. Robinette SL, Zhang F, Bruschweiler-Li L, Brüschweiler R. Anal Chem. 2008; 80:3606–3611. [PubMed: 18422338]
26. Bax A, Summers MF. J Am Chem Soc. 1986; 108:2093–2094.
27. Braunschweiler L, Ernst RR. J Magn Reson. 1983; 53:521–528.
28. Kupce E, Freeman R. J Am Chem Soc. 2008; 130:10788–92. [PubMed: 18642908]

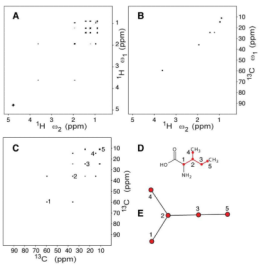

Figure 1.
Carbon skeleton of isoleucine derived from doubly indirect covariance spectrum C) constructed from A) 2QF-COSY and B) ^{13}C-^{1}H HSQC (see Eq. (2)). D) Chemical structure of isoleucine. E) Graph theoretical representation of skeletal formula obtained from covariance spectrum. The carboxyl carbon is not included since it does not have a proton attached. Nodes 1, 2, and 3 represent C_α, C_β, and $C_{\gamma 1}$, respectively, and nodes 4, 5 represent the CH_3 groups of $C_{\gamma 2}$ and $C_{\delta 1}$, respectively. All spectra were collected at 800 MHz proton frequency at 298 K. The chemical structures in this and the following figures were drawn using ChemSketch (Advanced Chemistry Development Inc, Toronto, Canada).

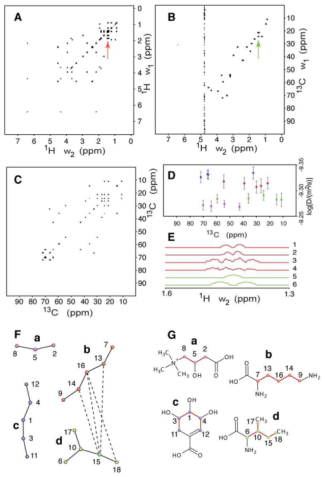

Figure 2.

Doubly indirect covariance spectroscopy of model mixture containing carnitine, lysine, isoleucine, and shikimate. A) 2QF-COSY spectrum, B) ^{13}C-^{1}H HSQC spectrum, C) doubly indirect covariance spectrum (see Eq. (2)), D) INEPT DOSY spectrum, E) absolute value cross sections along ^{1}H dimension of COSY and HSQC for the overlap ω_2 = 1.45 ppm (traces 1-4 belong to COSY and traces 5,6 to HSQC). F) Carbon-carbon connectivity graphs of mixture components. Dashed lines indicate extraneous connectivities to nodes 15 and 16 due to ^{1}H overlap at 1.45 ppm as indicated by arrows in Panels A and B. They are identified either based on DOSY (Panel D), or by filtering according to differential peak positions and shapes (Panel E). The labels a, b, c, d in Panels F and G belong to carnitine, lysine, shikimate and isoleucine, respectively. The computational time on a workstation with a single-core processor is 19 s.

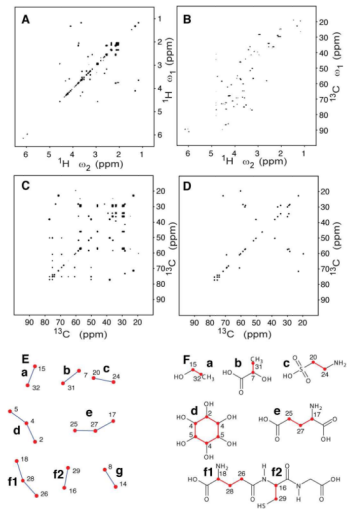

Figure 3.

Doubly indirect covariance spectroscopy of cancer cell line DU145 extract. A) DQF-COSY spectrum, B) ^{13}C-^{1}H HSQC spectrum, C,D) doubly indirect ^{13}C-^{13}C covariance spectrum constructed from parent spectra A, B (see Eq. (2)) before (C) and after (D) removing effects due to ^{1}H overlap, E) carbon-carbon connectivity graphs of mixture components determined from spectrum D. F) Labels a, b, c, d, e, f1 and f2 in Panels E, F indicate ethanol, lactate, taurine, myo-inositol, glutamate, and glutathione reduced, respectively. Graph g in panel E does not match any of the compounds in the database. Myo-inositol displays only 3 nodes (graph d in Panel E) because of ^{13}C chemical shift degeneracy induced by its molecular symmetry.