Published in final edited form as:

Anal Chem. 2014 January 7; 86(1): 47–57. doi:10.1021/ac403520j.

Multidimensional approaches to NMR-based metabolomics

Kerem Bingol and Rafael Brüschweiler*

Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210;

Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32306;

National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32310

Introduction

The field of metabolomics, which is also referred to as metabonomics, has gained significant attention over the recent past as it is developing rapidly as a powerful way to comprehensively study complex biological systems from a small molecule perspective. According to the Web of Science, since 2010 over 5,000 papers have been published with the key words “metabolomics”, “metabonomics” or “metabolite profiling”. Small biological molecules (or metabolites) with molecular weight <1500 Da are involved in many critical functions in biological systems, such as energetics, signaling, and as building blocks of more complex biopolymers, which makes the understanding of their composition, chemical structure, and reaction pathways important. Two main objectives of metabolic analysis are the discovery of modified and new natural products and the detection of biologically meaningful changes in metabolite concentration and fluxes.¹

Mass spectrometry (MS) and NMR spectroscopy are by far the most powerful methods in metabolomics.² This is because of the excellent resolution power of both of these methods to uniquely detect individual molecular species. MS and NMR can be considered as “universal” techniques as essentially every conceivable metabolite can be measured by these techniques, which is in contrast to other methods, such as optical spectroscopy methods, which provide information only about the subset of optically active metabolites. This review focuses on recent progress in metabolomics by NMR spectroscopy.

NMR provides information at atomic resolution of NMR-active nuclei. These include ^1H , ^{13}C and ^{31}P nuclei at natural abundance as well as ^{13}C and ^{15}N nuclei in isotopically enriched metabolites. The chemical shifts of these nuclei define the resonance positions in the spectrum reporting about their chemical environment and the scalar J-couplings define the fine structure of the NMR peaks reporting about through-bond spin-spin connectivities. Other parameters are the T_1 and T_2 relaxation times and the nuclear Overhauser effect reflecting the inter-spin distances and overall reorientational diffusion rates, which are often directly related to the molecular size. In addition, translational diffusion rates can be extracted by pulsed-field gradient based methods, which also report on the size and shape of a metabolite, e.g., via the Stokes-Einstein relationship.

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

An inherent drawback of NMR is its limited sensitivity, which restricts its application to metabolite concentrations of the order of μM^3 . On the other hand, NMR offers a number of unique advantages. The atomic-resolution information permits the characterization of the chemical structure of a metabolite, e.g., through spin-spin connectivity information. NMR also allows one to unambiguously identify different slowly interconverting isomers, which are present for many carbohydrates (e.g. α - vs. β -glucose). Since the NMR peak integrals are directly proportional to the molecular concentration, NMR is a highly quantitative method when it comes to the determination of metabolite concentrations and their changes. The preparation of NMR samples is often straightforward and may involve as little as dissolving the lyophilized metabolite sample in a buffered solution. In addition to liquid samples, semi-solid samples, such as tissues, can also be analyzed by NMR spectroscopy. Since NMR spectroscopy is non-destructive, the same sample can be analyzed for an extended period of time. Finally, because for the same sample essentially identical results can be obtained by different users on different spectrometers operating at the same magnetic field strength, the reproducibility of NMR data is very high. The simplest and fastest NMR techniques are based on 1D Fourier-transform (FT) NMR as shown in Figure 1, which allow the measurements of hundreds or even thousands of samples in a relatively short period of time, such as urine samples of sizeable populations⁴ or cell extracts⁵. This is because the experimental time required for a single sample is between a few seconds and a few minutes. Such high-throughput applications are significantly facilitated by the use of automatic NMR sample changers. More detailed and better resolved information can be obtained from higher-dimensional NMR experiments, in particular 2D NMR, at the expense of somewhat longer measurement times.⁶⁷ At some stage, nearly all metabolomics studies revolve around identification of metabolites and their quantification.

Identification

Metabolite identification is typically performed in two steps. In the first step, the metabolite mixture is deconvoluted into its individual components and in the second step, each metabolite is identified by querying one or several metabolite databanks. Due to the limited number of metabolites compiled in metabolite databases, a key challenge in metabolomics is the identification of ‘unknown’ signals, which are signals that belong to compounds not found in databanks. Positive database identification requires correct matches of typically several NMR parameters of an unknown metabolite with the ones collected of the pure compound under identical or near identical conditions. These independent NMR parameters can be chemical shifts and peak splittings due to scalar spin-spin couplings (J-couplings), which can be obtained from 1D ^1H , 1D ^{13}C or 2D NMR experiments.⁸ Metabolite identification is sometimes based only on a subset of the independent parameters, for example a single peak, which obviously carries the risk of false identification.

There exist several public NMR metabolomics databanks, such as the BMRB⁹ and the HMDB¹⁰, containing experimental data of pure metabolite standards. Experimental information can be queried against these databanks in order to identify metabolites. Querying algorithms provided for these and other databases can vary significantly in their accuracy. Still, the main limitation of database querying lies in the fact that of the total “metabolite universe” only a small fraction of compounds, currently of the order of ~ 1000 –

2000 molecules, have their NMR spectra stored in databanks.⁸ As a result, many signals observed in metabolomics studies cannot be assigned by database query alone. Traditionally, identification of uncatalogued compounds requires their isolation through labor-intensive purification from the complex mixtures by using separation techniques, such as chromatography, followed by extensive characterization by spectroscopy and other methods. For obvious reasons, such an approach is not suitable for high-throughput applications. Below we will discuss a 2D NMR-based approach that permits *de novo* carbon-backbone structure determination without the need for purification.

Quantification

The other key challenge of metabolomics is quantification. Although NMR peak areas or peak heights have been used as semi-quantitative measures in comparative studies, such information does not provide accurate and absolute metabolite concentrations. In principle, analysis of a standard 1D NMR spectrum provides one of the most convenient and accurate concentration determinations methods among all techniques. Since NMR quantification methods require complete relaxation of spins between scans to prevent differential saturation effects, the required interscan delay should be at least 5 times as long as the longitudinal relaxation time, T_1 , of the slowest relaxing nucleus of interest in the sample.¹¹ In 1D NMR, a reference compound (internal standard such as DSS) with known concentration is simply added to the sample.¹² Absolute concentrations are then obtained by relating the measured peak volumes of the compounds with unknown concentrations to the peak volume of the internal standard. This method works well for molecules with peaks that are well isolated. However, the method is hampered by peak overlaps, which poses a challenge for the analysis of complex mixtures where the entire spectrum or a spectral region is resolution limited (see Methods section). Peaks that overlap in 1D spectra can often be resolved in 2D NMR spectra. However, quantification of metabolites from 2D NMR spectra is less straightforward and is an active area of research as discussed below.

In metabolomics, changes in concentrations of individual metabolites are either monitored as a function of time, e.g. by taking blood samples of an individual at different time points, or by studying differences in metabolic concentrations between different individuals, e.g., healthy vs. diseased. The first type of analysis provides insight into the biological activity of individual metabolites while the second type can reveal potential biomarkers, which are metabolites whose presence or absence, possibly in combination with other metabolites, significantly correlate with, e.g., a specific disease.

Data Analysis

Biomarkers can be discovered through NMR-based metabolomics by detecting characteristic spectral patterns in relationship to the state of a biological system, for example, healthy vs. diseased, old vs. young, or mutant vs. wild-type. Operationally, this boils down to the identification of correlations between the changes in the concentrations of specific metabolites and the state of the system, which can be accomplished by multivariate analysis, such as principal component analysis (PCA) and partial least squares regression (PLS). In its basic form, a set of 1D NMR spectra of the biological system in various states is represented

by a $n \times q$ matrix $\mathbf{X}$, where each row $k = 1 \dots n$ corresponds to a different 1D spectrum consisting of q data points. The $q \times q$ covariance matrix $\mathbf{C}$ has elements

$$C_{ij} = \frac{1}{n} \sum_{k=1}^n (X_{ki} - X_i)(X_{kj} - X_j) \quad (1)$$

$$\text{where } X_i = \frac{1}{n} \sum_{k=1}^n X_{ki} \quad (2)$$

Diagonalization of $\mathbf{C}$ produces eigenvectors $\mathbf{v}_m$ to eigenvalues λ_m

$$\mathbf{C}\mathbf{v}_m = \lambda_m \mathbf{v}_m \quad (3)$$

where the first principal component $\mathbf{v}_1$ is the eigenvector to the largest eigenvalue λ_1 corresponding to the linear combination of NMR resonances that accounts for the largest amount of variability (variance) across the set of spectra. The eigenvector to the second-largest eigenvalue is the 2nd principal component, etc. The projections, $\mathbf{p}_m$, of the offset-free original 1D spectra $\mathbf{X}$ onto a 2D subspace usually spanned by the first and second principal components $\mathbf{v}_m$ ($m = 1, 2$)

$$\mathbf{p}_m = \mathbf{X}\mathbf{v}_m \quad (4)$$

can reveal characteristic clustering properties with respect to the system's state when plotting $\mathbf{p}_2$ vs. $\mathbf{p}_1$. Hence, the projection of the 1D spectrum of a biological system whose state is unknown onto the principal components permits a prognosis about the system's state. The advantage of PCA is that it can be performed in an unsupervised manner, since the principal components do not depend on the state of the system. By contrast, partial least squares regression (PLS) is a 'supervised' linear regression model that optimizes the predictability of the state of a system based on the 1D spectra of systems by taking into account the knowledge of a system's state.¹³ Generally, this allows a more accurate prediction of a system's state based on its 1D spectrum as compared to PCA. Metabolites whose concentration changes correlate (or anticorrelate) strongest with a system's state are potential biomarkers. In Figure 2, both PCA and PLS methods applied to 1D ^1H NMR spectra of extracts from rat brains successfully differentiates between animals that have been infected by *F. hepatica* (flatworms) and ones that are uninfected.¹⁴

An important pre-requisite of multivariate data analysis is that resonances belonging to the same metabolites are properly aligned across different spectra. Furthermore, the linewidth of corresponding resonances should be similar across multiple samples. Misalignment and linewidth changes can be caused by variability between samples in terms of pH, temperature, spectral calibration or shim. For the alignment of NMR peaks a large number of algorithms are available.¹⁵ Alignment and linewidth problems can also be minimized by binning procedures, where the spectrum is divided into small sections (bins) whereby the

total spectral area for each bin is used for multivariate statistical analysis. Improvements in spectral alignment software continue to be made substituting in part binning procedures.³

In addition to PCA and PLS, there are other useful tools available that extract differences in NMR data. 1D NMR spectra of multiple samples can be analyzed by statistical correlation to identify individual compounds or a number of compounds that are implicated in the same metabolic pathways. In statistical total correlation spectroscopy (STOCSY) method, the correlation matrix **C** is calculated from 1D NMR spectra of multiple samples by using the following equation $\mathbf{C} = \mathbf{X}^T \mathbf{X} / (n-1)$, where **X** is a $n \times q$ matrix where n is the number of 1D spectra (one per sample) and q is the number of data points per spectrum. This concept can be also generalized to different spectral sets **X**₁ and **X**₂.¹⁶ In Figure 3, a 2D STOCSY plot of a urinary NMR data set is illustrated.¹⁷ The **C** matrix in Figure 3B illustrates negative and positive correlations of metabolites across a set of samples, where each cross section (Figure 3A) shows the correlation of a particular metabolite with the other metabolites. STOCSY is formally related to covariance NMR, which is described below.

If only few metabolic samples are available, high-resolution 2D NMR spectroscopy can be directly used to extract differences between samples as has been shown in a spectral difference approach, DANS¹⁸, and a pattern recognition approach, HATS-PR¹⁹. These approaches benefit from the higher resolution of 2D NMR spectra and potentially provide higher accuracy. In the presence of larger sample size, they are however more time-consuming and computationally more demanding.

Methods: Identification

Spectral deconvolution of complex mixtures

Identification of metabolites in metabolomics samples requires the deconvolution of the NMR spectrum into discrete sets of signals where each set belongs to an individual metabolite. This can be done by physical separation of the mixture into its different components, for example by HPLC, followed by the acquisition of the NMR data of each component. This process is time-consuming and potentially labor-intensive and some metabolites or natural products may even lose their activity during extraction.

Molecular properties reflected in the NMR spectra can be used as additional dimensions to deconvolute mixtures. For instance, mixture components can be deconvoluted based on their characteristic translational diffusion constants obtained by diffusion-ordered spectroscopy (DOSY) or NMR relaxation rates, which act as a pseudo-second dimension.²⁰²¹²² These approaches are, however, restricted to mixtures that contain relatively few components, typically below 10.

A more comprehensive as well as more time-consuming approach is based on 2D NMR.⁶⁷ 2D NMR helps to address the overlap issue in 1D spectra by taking advantage of spin-spin interactions allowing the transfer of spin magnetization across a molecule. 2D NMR approaches that use a cross-relaxation mechanism for magnetization transfer are NOESY²³, which has found application for the analysis of metabolites,²⁴²⁵ and ROESY²⁶²⁷. 2D NMR approaches that use scalar J-couplings for magnetization transfer are more versatile for small

molecules, in particular ^1H - ^1H COSY²⁸ ^1H - ^1H TOCSY²⁹ and ^{13}C - ^1H HSQC³⁰. COSY and TOCSY experiments permit identification of individual ^1H spin systems that can be assigned to the various mixture components. ^1H - ^1H TOCSY is particularly well suitable for computational analysis, since each cross-section of compounds represents the 1D spectrum of the whole spin system.

In order to obtain sufficiently high resolution along the indirect frequency domain by 2D Fourier-transform processing, a relatively large number of time increments along the indirect dimension is required, which leads to prolonged measurement times. As a solution, covariance processing can be applied, which endows the indirect dimension with the same high resolution as the direct dimension, independent of the number of increments.³¹ Similar to STOCSY, covariance NMR spectroscopy extracts correlations across multiple 1D NMR spectra. However rather than physically different samples, it uses different spin evolution times as the determining factor of variation. Hence, in covariance NMR, different rows do not correspond to different samples, but they correspond to different t_1 increments of the same sample. A covariance spectrum $\mathbf{C}$ can be most easily computed using the following expression:

$$\mathbf{C}=(\mathbf{F}^T\mathbf{F})^{1/2} \quad (5)$$

where, $\mathbf{F}$ is the standard 2D FT NMR spectrum. The advantage of covariance NMR processing as compared to Fourier transformation is that a substantial increase in resolution can be achieved when the number of t_1 increments needs to be limited as in the context of high-throughput applications. In Figure 4A, a high resolution ($t_1=512$) 2D FT ^1H - ^1H TOCSY spectrum of a model mixture containing 4 model compounds is shown. When the number of t_1 increments is reduced to 96, in order to save a significant amount of NMR time, it leads to poor spectral resolution along the indirect frequency domain ω_1 of the 2D FT spectrum (Figure 4B). On the other hand, covariance processing of the low-resolution spectrum ($t_1=96$) generates the high-resolution spectrum in Figure 4C. Although the spectra in Figure 4A and 4C have both high resolution, the experimental time of the covariance-processed spectrum is only about 20% of the full 2D FT spectrum.

^{13}C NMR spectra generally display a larger chemical shift dispersion than ^1H spectra and they display characteristically narrow lines, which reduces the chances of peak overlaps. A significant downside of ^{13}C NMR at natural abundance ($\sim 1.1\%$) with direct ^{13}C detection is the inherently low sensitivity. Sensitivity of ^{13}C NMR is increased by using 2D ^{13}C - ^1H HSQC³⁰, where the initial polarization and detection is performed through ^1H spins. A downside of HSQC-type spectra, as compared to TOCSY, is the lack of complete spin system information, because each cross-peak is independent of all others. On the other hand, HSQC spectra of individual compounds represent useful fingerprints, providing the number of C-H spin pairs of the molecule together with the ^{13}C and ^1H chemical shifts, which report on the nature of the chemical groups.³²³³³⁴

The low natural abundance problem of ^{13}C spins can be overcome by using uniform ^{13}C labeling.³⁵ The large one-bond J-couplings ($^1J(^{13}\text{C}, ^{13}\text{C}) > 30 \text{ Hz}$) make the efficient transfer

of spin magnetization during ^{13}C - ^{13}C TOCSY mixing possible. The same $^1\text{J}(^{13}\text{C}, ^{13}\text{C})$ -couplings, however, lead to broad multiplet structures resulting in increased peak overlap, which are mitigated along the indirect ω_1 dimension by ^{13}C - ^{13}C constant-time (CT) TOCSY spectroscopy as originally demonstrated for side-chain assignments of proteins.³⁶ Additionally, the multiplet pattern along the direct dimension can be decoupled by indirect covariance processing,³⁷ which yields a homonuclear decoupled spectrum along both dimensions.³⁸ ^{13}C - ^{13}C CT-TOCSY is a powerful experiment for the simultaneous characterization of a large number of metabolites (>100) in a single sample and can be directly applied to various organisms that can be uniformly ^{13}C -labeled (*E. coli*, yeast, *C. elegans*, plants, etc.).³⁵

2D NMR spectra can be deconvoluted manually, which is a tedious process that is impractical for high-throughput applications. Our lab has developed semi-automated approaches to deconvolute 2D TOCSY spectra of complex mixtures into TOCSY traces of individual mixture components, which can be directly searched against NMR databases for identification. One of these techniques is *DemixC*, which extracts 1D cross-sections (traces) of a 2D TOCSY that contain little or no peak overlaps by different spin systems.³⁹ Although *DemixC* works well for mixtures of moderate complexity, metabolomics samples have frequently a level of complexity with dozens to hundreds of compounds, which makes peak overlaps very common. For this purpose, the *DeCoDeC* technique was developed, which can successfully handle the deconvolution of mixtures with higher complexity.⁴⁰ The *DeCoDeC* technique identifies common peaks in selected pairs of TOCSY 1D cross-sections (traces) in order to eliminate overlapping peaks belonging to other metabolites. In a TOCSY spectrum, represented by $N_1 \times N_2$ matrix $\mathbf{F}$ (or, alternatively, the covariance spectrum $\mathbf{C}$), there are many trace-pairs that do not correspond to a single metabolite. In order to select only meaningful trace pairs, a peak-picking procedure is used. Peak-picking of the cross-peaks of matrix $\mathbf{F}$ yields a list (k, k') where k and k' denote the position of a certain cross-peak along the two frequency axes. For each cross-peak entry (k, k') , the consensus trace $\mathbf{q}^{(kk')}$ is determined as follows:

$$q_j^{(kk')} = \min(F_{kj}, F_{k'j}) \quad (6)$$

where index j goes over all N_2 columns. Eq. (6) ensures that the consensus trace contains only peaks that are present in *both* input traces of $\mathbf{F}$; therefore, on average it is less affected by peak overlap than each of the two input traces and it is more likely to represent a 'clean' 1D spectrum of a spin system. The next step aims at data reduction to extract a non-redundant set of clean 1D spectra representing the molecules in the mixture. For this purpose, the complete set of consensus traces is subjected to clustering for identification of those traces that represent 1D spectra of individual spin systems. In Figure 5A, *DeCoDeC* is capable to successfully deconvolute a uniformly ^{13}C labeled *E. coli* cell lysate ^{13}C - ^{13}C CT-TOCSY spectrum into 98 unique 1D ^{13}C NMR consensus traces in a semi-automated manner.³⁵ The *DeCoDeC* approach is also applicable to certain heteronuclear experiments, such as 2D ^{13}C - ^1H HSQC-TOCSY.⁴⁰ Furthermore, it is applicable to the analysis of a combination of 2D experiments, namely 2D ^{13}C - ^1H HSQC-TOCSY with 2D ^{13}C - ^1H

HSQC.⁴⁰ In this case, instead of 1D NMR traces, 2D ^{13}C - ^1H HSQC planes of individual metabolites are obtained by using the triple-rank spectroscopy method.⁴¹

Backbone topology determination

Many biological samples contain significant numbers of unknown metabolites that are not catalogued in databanks. Systematic identification and structural characterization of uncatalogued metabolites is an important task of metabolomics. NMR-based *de novo* structure elucidation of metabolites is based on 2D NMR methods that provide structural information. For sensitivity reasons, so far, the majority of these studies have been based on 2D ^1H NMR experiments taking advantage of the high natural abundance of ^1H spins and their relatively large magnetic moment.⁴² However, by combining 2D ^1H - ^1H COSY and 2D ^{13}C - ^1H HSQC spectra, a ^{13}C - ^{13}C correlation spectrum can be constructed, by means of linear algebraic tools, with very high resolution along both dimensions, from which the topology of protonated carbons of individual molecules can be derived.⁴³ This strategy, termed doubly-indirect covariance processing, was successfully applied to the analysis of metabolites in a prostate cancer cell extract.

A drawback of 2D ^1H NMR is that the strong conformation dependence of vicinal $^3J(^1\text{H}, ^1\text{H})$ -couplings can cause uneven magnetization transfer in TOCSY and COSY spectra, thereby impeding the assignment of cross-peaks to individual spin systems and entire molecules. Furthermore, the spectral information of protons may not be sufficient for the complete reconstruction of the carbon backbone of metabolites and their bonding topology, which is a prerequisite for structure determination.

In order to obtain information directly about the carbons and their connectivity, fully ^{13}C -labeled metabolic samples can be used instead. A protocol was developed, which identifies traces in long mixing time ^{13}C CT-TOCSY spectrum that are unique for individual mixture components by *DeCoDeC* (Figure 5A) and then assembles for each consensus trace the corresponding carbon-bond topology network by using short mixing time CT-TOCSY and COSY. This led to the determination of 112 topologies of unique metabolites in *E. coli* from a single sample. They constitute the “topolome” of a cell, which is depicted in Figure 5B. The observed occurrences of topologies, which are given in Figure 5C, are dominated by carbon topologies of carbohydrates (34.8%) and amino acids (45.5%) both of which can represent building blocks of more complex biopolymers.³⁵

Database searching

Since the topology determination does not require any database information, the approach is not limited to mixtures whose components are already catalogued. However, for those mixture components that are present in metabolite databases, it is important to be able to accurately and rapidly identify them.

There are several 1D ^1H NMR and 1D ^{13}C NMR databases for metabolite identification, some of them are public⁹⁴⁴⁴⁵⁴⁶¹⁰ and others are commercial (Chenomx NMR Suite (Edmonton, AB, Canada), Bruker AMIX (Billerica, MA, USA)). However, compound identification from a single 1D spectrum of a complex mixture can introduce ambiguities for

two reasons: (1) limited discrimination power because of peak overlaps and the lack of connectivity information between peaks belonging to the same compound and (2) changes in peak positions between mixture and database spectra. Moreover, if the mixture spectrum is measured at a different magnetic field strength than the spectra in the database, a mismatch in peak appearance vs peak-to-peak distance will cause additional complications. For example, two peaks that overlap at low magnetic field strength may be easily identifiable as two separate signals at high field.

The use of 2D NMR spectra can overcome some of these issues. For the matching of 2D NMR spectra against database information a number of different strategies have been proposed. 2D ^1H - ^{13}C HSQC spectra can be matched cross-peak by cross-peak against database entries.⁹⁴⁴¹⁰⁴⁷ Although the resolution is increased by the introduction of the indirect ^{13}C dimension, the lack of connectivity information between the different ^1H - ^{13}C pairs belonging to the same molecule causes similar types of challenges for peak annotation and metabolite identification as in the case of 1D NMR. Connectivity information between resonances stemming from different parts of a molecule is available in TOCSY spectra collected at long mixing times.

Since TOCSY traces only correlate resonances with each other that belong to the same spin system, for molecules with multiple spin systems or multiple isomeric forms that are in slow exchange, they yield only part of the entire 1D NMR spectrum. This is exemplified in Figure 6 with galactose. Although the 1D ^{13}C NMR spectrum consists of signals of both α - and β -galactose (Figure 6A), the ^{13}C TOCSY traces (collected at long mixing time) consist of signals of only one isomer (Figures 6B and 6C). Therefore, a query against a NMR database consisting of full 1D NMR spectra of metabolites leads to imperfect matches, carrying the risk of false interpretations. Moreover, depending on the scoring function used, often molecules with a large number of resonances are returned since they have a higher chance to match the resonances of the query trace. Because NMR databases do not sort spins into individual spin systems or multiple slowly exchanging isomers for separate queries, a customized metabolite database, TOCCATA, has been developed, which is specifically geared toward the query of ^{13}C TOCSY traces with the goal to optimize the matching accuracy.⁴⁸ TOCCATA uses ^{13}C chemical shift information for the reliable identification of metabolites, their isomeric states and spin systems. TOCCATA, whose spectral information was derived from the BMRB and HMDB databases and the literature, currently contains 463 compounds and 801 spin systems, and it can be used through a publicly accessible web server. Out of the 463 compounds, 171 contain more than one spin system and 37 exist in multiple isomeric forms. In addition to chemical shift information, TOCCATA contains the peak multiplet pattern of each individual carbon resonance in the databank assuming that the neighboring carbons are all ^{13}C labeled. In a uniformly ^{13}C -labeled compound, with all protons decoupled, a ^{13}C multiplet reports on the number of directly bonded ^{13}C atoms. A primary, secondary, tertiary, or quaternary carbon possesses a multiplet with intensity ratios of 1:1, 1:2:1 (or 1:1:1:1), 1:3:3:1 and 1:4:6:4:1, respectively. Inspection of multiplet patterns along the ω_2 -detection dimension in the CT-TOCSY spectrum has proven useful for the independent validation of the top matches returned by database query.³⁵ TOCCATA allows the identification of metabolites in the submillimolar range from ^{13}C - ^{13}C TOCSY experiments of complex mixtures as has been demonstrated for different uniformly ^{13}C -

labeled mixtures. For an *E. coli* cell extract, querying with TOCCATA provides a 41% improvement of the accuracy over the BMRB and a 32% improvement over COLMAR⁴⁵. The same customization strategy can be extended to ¹H-¹H TOCSY spectral databases, work that is in progress in our lab.

Methods: Quantification

1D ¹H NMR experiments are widely applied for the extraction of quantitative concentrations of individual chemical species in solution provided that the spectra are well-resolved. As mentioned above, a key advantage of 1D ¹H spectra is that the integral of a given peak is directly proportional to the concentration of the compound it belongs to.¹¹

In the presence of strong peak overlaps, which are typical for complex mixtures such as ones encountered in metabolomics, alternative methods are required. While the resolution issue can often be overcome by 2D NMR spectroscopy, the quantification of 2D spectra is hindered by the variability of cross-peak intensities due to uneven magnetization transfer during the preparation, evolution, or mixing periods caused by differences in scalar J-couplings and spin relaxation.⁴⁹ This prevents the direct use of cross-peak integrals as quantitative measures of sample concentrations. Therefore, more elaborate 2D NMR quantification methods have been developed, which can be divided into two main groups based on their strategy to deal with the variability of cross-peak intensities. The first group uses an internal standard for each type of molecule. This approach has been demonstrated for heteronuclear 2D ¹³C-¹H HSQC^{50,51,52} and for homonuclear 2D ¹H-¹H TOCSY⁵² and 2D ¹H-INADEQUATE experiments.⁵³ It requires the preparation and measurement of a large number of standard samples, often under multiple conditions. Obviously, molecules identified in a sample cannot be quantified if their standard is unknown as is also the case for newly discovered molecules.

The second approach aims at minimizing the variability in cross-peak intensities by modification of ¹³C-¹H HSQC experiments^{54,55,56,57} without requiring an internal standard for each molecule. One of these approaches is time zero ¹³C HSQC (HSQC₀), which extrapolates a series of ¹³C HSQC experiments to a virtual HSQC spectrum at “zero time”, which is free of signal attenuation during the coherence transfer period. Therefore, its peak intensities can be directly converted to absolute concentrations.⁵⁶ The approach has the advantage that it does not require correction factors specific for each metabolite. The idea behind the HSQC₀ approach is shown in Figure 7.

The quantification techniques mentioned so far have been applied to metabolite samples at natural ¹³C abundance. In the case of uniformly ¹³C labeled metabolites, 2D ¹³C-¹³C CT-TOCSY NMR spectra can be directly quantified, without requirement of a standard for each molecule, provided that the backbone carbon topology is known.⁵⁸ The TOCSY cross-peak volumes are computed by direct integration of the Liouville-von Neumann equation describing the spin dynamics in terms of quantum-mechanics during the 2D experiment and compared with the experimental values. The approach works for 2D ¹³C-¹³C CT-TOCSY spectra collected at short ($\tau_m = 4.7$ ms) or long mixing times ($\tau_m = 47$ ms). Figure 8 illustrates comparison of experimental and simulated cross-peak integrals of fructose β -

furanose, glucose β -pyranose, isoleucine, and lysine at long mixing-time. The peak integrals in Fig. 8 align well along the diagonal with a correlation coefficient $R > 0.95$. At shorter mixing times, the accuracy is slightly reduced because of the smaller number of cross-peaks leading to larger statistical errors and distorted peak shapes caused by the presence of zero-quantum effects. For the short-mixing time TOCSY experiments, the numerical results can be also approximated by analytical relationships providing a rapid means for quantification.

Overall, these techniques show that quantification by 2D NMR is feasible overcoming the requirement of well-resolved resonances in the 1D spectrum, which should prove particularly useful for applications to highly complex mixtures with strongly overlapping peaks as is typical for real-world metabolic samples.

Concluding Remarks

Accurate identification and quantification of the metabolites in complex mixtures are key steps in metabolomics, which significantly affect downstream analysis and interpretation. Metabolite identification is performed in two steps. After the NMR spectrum of a metabolite mixture has been deconvoluted into fingerprints of individual components, the metabolites are identified from the fingerprints either *de novo* or by databank searching. In this review, recent approaches are described to improve and speed-up these steps using multidimensional methods. The most convenient platforms are the obviously the ones that permit identification and quantification using the same data set(s). For example, the same 2D ^{13}C - ^{13}C CT-TOCSY experiment used for identification of metabolites (deconvolution, backbone topology construction and database querying) can also be used for quantification through back-calculation of the peak volume of each metabolite. While metabolomics studies with uniformly ^{13}C -labeled samples are not yet widespread, the ease and reliability of interpretation provides an incentive for this approach. As ^{13}C labeling of whole organisms, such as bacteria, yeast, and plants, is becoming increasingly common, the emergence of a wealth of new chemical and biological information including both natural product chemistry and metabolomics can be expected. This also makes the development of labeling strategies for more complex, fully ^{13}C -labeled organisms a promising direction of research.

One of the challenges for NMR remains sensitivity. Dynamic nuclear polarization (DNP) methods hold promise to increase the signal-to-noise ratio of NMR spectra by several orders of magnitude. The most common DNP technique in solution, dissolution DNP, first polarizes the sample frozen in a solution containing free radicals by microwave radiation. Subsequently, the sample is dissolved and transferred to the NMR spectrometer for data acquisition.^{59,60,61} A drawback of the approach is that the recycle time, which includes irradiation, sample transfer and data acquisition, takes between one and several hours and the process needs to be repeated many times for the measurement of a traditional 2D NMR spectrum. An alternative approach relies on ultrafast techniques, which acquire a multidimensional NMR experiment in a single scan by encoding spin evolution spatially by using pulsed magnetic field gradients.⁶² Another DNP technique, temperature-jump (TJ-DNP), performs polarization and the NMR experiment in the same physical environment by melting the sample with CO_2 laser irradiation. The recycle time of the approach is shorter

(one or several minutes), but the approach is demanding on the hardware.⁶³ All DNP approaches have a common drawback namely that polarization of atoms in a molecule tends to be non-uniform depending on the T_1 's of the polarized atoms.⁶⁴ Nevertheless, improvements in sensitivity by DNP not only reduce the amount of NMR sample required, but they also allow studying of low abundance nuclei. Sufficient time resolution (seconds) to observe ongoing reactions is achieved by using DNP with small pulse flip angle excitation. By combining this approach with isotope labeling, the path of a metabolite in an organ or a whole organism can be followed in real time.⁶⁵ Furthermore, ^{13}C labeled circulating drugs at submicromolar concentrations can be detected by DNP in complex body fluids with minimum sample preparation and minimum spectral background, because all other molecules are at natural ^{13}C abundance.⁶⁶ For mass-limited samples, sensitivity can also be increased by using a small radio-frequency coil size.⁶⁷ For instance, a 1-mm triple resonance high-temperature-superconducting probe was built to study metabolites and natural products with $\sim 10\ \mu\text{l}$ volume size, which is 1/60 of a regular NMR sample in a 5 mm tube.⁶⁸ Sensitivity of natural abundance samples can be increased by chemoselective isotope tagging approaches, which was demonstrated by adding compounds with ^{13}C and ^{15}N labels and allowing them to react with metabolites containing amine and carboxyl groups, respectively.^{69,70,71} Finally, sensitivity together with resolution can be enhanced by using larger magnetic fields, whereby the increase in instrument costs has to be taken into consideration. As discussed above, multidimensional experiments increase the resolution besides providing valuable chemical connectivity information. However, they require longer acquisition times because of the sampling along the indirect frequency dimensions, especially for 3D and higher dimensional experiments. In situations where sampling is the limiting factor, but not sensitivity, alternative sampling schemes can be employed, such as GFT, projection reconstruction, and non-uniform sampling.^{72,73,74,75,76} These and other rapid advances in measurement technology can be combined with the spectral analysis methods described in this review. They will provide ever more powerful capabilities for the rapid and yet accurate analysis of complex mixtures with broad applications for metabolomics studies from medicine to biofuels.

Acknowledgments

This work was supported by the National Institutes of Health (Grant R01 GM066041).

References

1. Robinette SL, Brüschweiler R, Schroeder FC, Edison AS. *Acc Chem Res.* 2012; 45:288–297. [PubMed: 21888316]
2. Lenz EM, Wilson ID. *J Proteome Res.* 2007; 6:443–458. [PubMed: 17269702]
3. Wishart DS. *TrAC, Trends Anal Chem.* 2008; 27:228–237.
4. Holmes E, Loo RL, Stalder J, Bictash M, Yap IKS, Chan Q, Ebbels T, De Iorio M, Brown IJ, Veselkov KA, Daviglus ML, Kestleloot H, Ueshima H, Zhao L, Nicholson JK, Elliot P. *Nature.* 2008; 453:396–400. [PubMed: 18425110]
5. Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ, Westerhoff HV, van Dam K, Oliver SG. *Nat Biotechnol.* 2001; 19:45–50. [PubMed: 11135551]
6. Jeener, J. Presented at the Ampere International Summer School. Basko Polje, Yugoslavia: 1971.
7. Aue WP, Bartholdi E, Ernst RR. *J Chem Phys.* 1976; 64:2229–2246.

8. Wishart DS. *Bioanalysis*. 2011; 3:1769–1782. [PubMed: 21827274]
9. 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, Wenger RK, Yao HY, Markley JL. *Nucleic Acids Res*. 2008; 36:D402–D408. [PubMed: 17984079]
10. Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, Mandal R, Sinelnikov I, Xia JG, Jia L, Cruz JA, Lim E, Sobsey CA, Shrivastava S, Huang P, Liu P, Fang L, Peng J, Fradette R, Cheng D, Tzur D, Clements M, Lewis A, De Souza A, Zuniga A, Dawe M, Xiong YP, Clive D, Greiner R, Nazzyrova A, Shaykhutdinov R, Li L, Vogel HJ, Forsythe I. *Nucleic Acids Res*. 2009; 37:D603–D610. [PubMed: 18953024]
11. Pauli GF, Jaki BU, Lankin DC. *J Nat Prod*. 2005; 68:133–149. [PubMed: 15679337]
12. Kim HK, Choi YH, Verpoorte R. *Nat Protoc*. 2010; 5:536–549. [PubMed: 20203669]
13. Trygg J, Holmes E, Lundstedt T. *J Proteome Res*. 2007; 6:469–479. [PubMed: 17269704]
14. Saric J, Li JV, Utzinger J, Wang Y, Keiser J, Dirnhofer S, Beckonert O, Sharabiani MT, Fonville JM, Nicholson JK, Holmes E. *Mol Syst Biol*. 2010; 6:396. [PubMed: 20664642]
15. Veselkov KA, Lindon JC, Ebbels TMD, Crockford D, Volynkin VV, Holmes E, Davies DB, Nicholson JK. *Anal Chem*. 2009; 81:56–66. [PubMed: 19049366]
16. Cloarec O, Dumas ME, Craig A, Barton RH, Trygg J, Hudson J, Blancher C, Gauguier D, Lindon JC, Holmes E, Nicholson JK. *Anal Chem*. 2005; 77:1282–1289. [PubMed: 15732908]
17. Robinette SL, Lindon JC, Nicholson JK. *Anal Chem*. 2013; 85:5297–5303. [PubMed: 23614579]
18. Pungaliya C, Srinivasan J, Fox BW, Malik RU, Ludewig AH, Sternberg PW, Schroeder FC. *Proc Natl Acad Sci USA*. 2009; 106:7708–7713. [PubMed: 19346493]
19. Robinette SL, Ajredini R, Rasheed H, Zeinomar A, Schroeder FC, Dossey AT, Edison AS. *Anal Chem*. 2011; 83:1649–1657. [PubMed: 21314130]
20. Johnson CS. *Prog Nucl Magn Reson Spectrosc*. 1999; 34:203–256.
21. Nilsson M, Botana A, Morris GA. *Anal Chem*. 2009; 81:8119–8125. [PubMed: 19715300]
22. Novoa-Carballal R, Fernandez-Megia E, Jimenez C, Riguera R. *Nat Prod Rep*. 2011; 28:78–98. [PubMed: 20936238]
23. Jeener J, Meier BH, Bachmann P, Ernst RR. *J Chem Phys*. 1979; 71:4546–4553.
24. Simpson AJ, Woods G, Mehrzad O. *Anal Chem*. 2008; 80:186–194. [PubMed: 18052233]
25. Farooq H, Soong R, Courtier-Murias D, Anklin C, Simpson A. *Anal Chem*. 2012; 84:6759–6766. [PubMed: 22788933]
26. Bothner-By AA, Stephens RL, Lee J, Warren CD, Jeanloz RW. *J Am Chem Soc*. 1984; 106:811–813.
27. Bax A, Davis DG. *J Magn Reson*. 1985; 63:207–213.
28. Rance M, Sorensen OW, Bodenhausen G, Wagner G, Ernst RR, Wüthrich K. *Biochem Biophys Res Commun*. 1983; 117:479–485. [PubMed: 6661238]
29. Braunschweiler L, Ernst RR. *J Magn Reson*. 1983; 53:521–528.
30. Bodenhausen G, Ruben DJ. *Chem Phys Lett*. 1980; 69:185–189.
31. Brüscheweiler R, Zhang F. *J Chem Phys*. 2004; 120:5253–5260. [PubMed: 15267396]
32. Chikayama E, Suto M, Nishihara T, Shinozaki K, Hirayama T, Kikuchi J. *PloS One*. 2008; 3:e3805. [PubMed: 19030231]
33. Strelko CL, Lu W, Dufort FJ, Seyfried TN, Chiles TC, Rabinowitz JD, Roberts MF. *J Am Chem Soc*. 2011; 133:16386–16389. [PubMed: 21919507]
34. An YJ, Xu WJ, Jin X, Wen H, Kim H, Lee J, Park S. *ACS Chem Biol*. 2012; 7:2012–2018. [PubMed: 23043523]
35. Bingol K, Zhang F, Brüscheweiler-Li L, Brüscheweiler R. *J Am Chem Soc*. 2012; 134:9006–9011. [PubMed: 22540339]
36. Eletsky A, Moreira O, Kovacs H, Pervushin K. *J Biomol NMR*. 2003; 26:167–179. [PubMed: 12766412]
37. Zhang F, Brüscheweiler R. *J Am Chem Soc*. 2004; 126:13180–13181. [PubMed: 15479045]
38. Zhang F, Brüscheweiler-Li L, Brüscheweiler R. *J Magn Reson*. 2012; 225:10–13. [PubMed: 23079590]

39. Zhang F, Brüscheiler R. *Angew Chem, Int Ed.* 2007; 46:2639–2642.
40. Bingol K, Brüscheiler R. *Anal Chem.* 2011; 83:7412–7417. [PubMed: 21848333]
41. Bingol K, Salinas RK, Brüscheiler R. *J Phys Chem Lett.* 2010; 1:1086–1089. [PubMed: 20436937]
42. Forseth RR, Schroeder FC. *Curr Opin Chem Biol.* 2011; 15:38–47. [PubMed: 21071261]
43. Zhang F, Bruschweiler-Li L, Brüscheiler R. *J Am Chem Soc.* 2010; 132:16922–16927. [PubMed: 21062057]
44. 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]
45. Robinette SL, Zhang F, Bruschweiler-Li L, Brüscheiler R. *Anal Chem.* 2008; 80:3606–3611. [PubMed: 18422338]
46. Akiyama K, Chikayama E, Yuasa H, Shimada Y, Tohge T, Shinozaki K, Hirai MY, Sakurai T, Kikuchi J, Saito K. *In Silico Biol.* 2008; 8:339–345. [PubMed: 19032166]
47. 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]
48. Bingol K, Zhang F, Bruschweiler-Li L, Brüscheiler R. *Anal Chem.* 2012; 84:9395–9401. [PubMed: 23016498]
49. Pauli GF, Godecke T, Jaki BU, Lankin DC. *J Nat Prod.* 2012; 75:834–851. [PubMed: 22482996]
50. Lewis IA, Schommer SC, Hodis B, Robb KA, Tonelli M, Westler WM, Sussman MR, Markley JL. *Anal Chem.* 2007; 79:9385–9390. [PubMed: 17985927]
51. Gronwald W, Klein MS, Kaspar H, Fagerer SR, Nurnberger N, Dettmer K, Bertsch T, Oefner PJ. *Anal Chem.* 2008; 80:9288–9297. [PubMed: 19551947]
52. Gowda GAN, Tayyari F, Ye T, Suryani Y, Wei SW, Shanaiah N, Raftery D. *Anal Chem.* 2010; 82:8983–8990. [PubMed: 20879716]
53. Martineau E, Tea I, Akoka S, Giraudeau P. *NMR Biomed.* 2012; 25:985–992. [PubMed: 22331830]
54. Rai RK, Tripathi P, Sinha N. *Anal Chem.* 2009; 81:10232–10238. [PubMed: 19919088]
55. Koskela H, Heikkila O, Kilpelainen I, Heikkinen S. *J Magn Reson.* 2010; 202:24–33. [PubMed: 19853484]
56. Hu K, Westler WM, Markley JL. *J Am Chem Soc.* 2011; 133:1662–1665. [PubMed: 21247157]
57. Rai RK, Sinha N. *Anal Chem.* 2012; 84:10005–10011. [PubMed: 23061661]
58. Bingol K, Zhang F, Bruschweiler-Li L, Brüscheiler R. *Anal Chem.* 2013; 85:6414–6420. [PubMed: 23773204]
59. Ardenkjaer-Larsen JH, Fridlund B, Gram A, Hansson G, Hansson L, Lerche MH, Servin R, Thaning M, Golman K. *Proc Natl Acad Sci USA.* 2003; 100:10158–10163. [PubMed: 12930897]
60. Ludwig C, Marin-Montesinos I, Saunders MG, Günther UL. *J Am Chem Soc.* 2010; 132:2508–2509. [PubMed: 20131776]
61. Bowen S, Sekar G, Hilty C. *NMR Biomed.* 2011; 24:1016–1022. [PubMed: 21387448]
62. Giraudeau P, Shrot Y, Frydman L. *J Am Chem Soc.* 2009; 131:13902–13903. [PubMed: 19743849]
63. Joo CG, Casey A, Turner CJ, Griffin RG. *J Am Chem Soc.* 2009; 131:12–13. [PubMed: 18942782]
64. Day IJ, Mitchell JC, Snowden MJ, Davis AL. *Appl Magn Reson.* 2008; 34:453–460.
65. Golman K, in 't Zandt R, Thaning M. *Proc Natl Acad Sci USA.* 2006; 103:11270–11275. [PubMed: 16837573]
66. Lerche MH, Meier S, Jensen PR, Hustvedt SO, Karlsson M, Duus JO, Ardenkjaer-Larsen JH. *NMR Biomed.* 2011; 24:96–103. [PubMed: 20862657]
67. Fratila RM, Velders AH. *Annu Rev Anal Chem.* 2011; 4:227–249.
68. Brey WW, Edison AS, Nast RE, Rocca JR, Saha S, Withers RS. *J Magn Reson.* 2006; 179:290–293. [PubMed: 16423543]
69. Shanaiah N, Desilva MA, Gowda GAN, Raftery MA, Hainline BE, Raftery D. *Proc Natl Acad Sci USA.* 2007; 104:11540–11544. [PubMed: 17606902]

70. Ye T, Mo H, Shanaiah N, Gowda GAN, Zhang S, Raftery D. *Anal Chem.* 2009; 81:4882–4888. [PubMed: 19518144]
71. Ye T, Zhang S, Mo H, Tayyari F, Gowda GAN, Raftery D. *Anal Chem.* 2010; 82:2303–2309. [PubMed: 20180538]
72. Kim S, Szyperski T. *J Am Chem Soc.* 2003; 125:1385–1393. [PubMed: 12553842]
73. Kupce E, Freeman R. *J Am Chem Soc.* 2003; 125:13958–13959. [PubMed: 14611222]
74. 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]
75. Pontoizeau C, Herrmann T, Toulhoat P, Elena-Herrmann B, Emsley L. *Magn Reson Chem.* 2010; 48:727–733. [PubMed: 20648569]
76. Billeter, M.; Orekhov, VY. *Novel Sampling Approaches in Higher Dimensional NMR.* Springer; Heidelberg: 2012.

Biographies

Kerem Bingol studied Chemistry and Molecular Biology & Genetics at the Middle East Technical University in Turkey and received his B.Sc. degrees in 2008. He then entered the graduate program in Molecular Biophysics at Florida State University and obtained his Ph.D. in 2013 under the supervision of Prof. Rafael Brüscheiler. Currently, he is working as a postdoctoral associate in the Brüscheiler Lab. His research focuses on the analysis of complex biological mixtures in the context of metabolomics by NMR spectroscopy and mass spectrometry.

Rafael Brüscheiler studied Physics at ETH Zürich and received his Ph.D. in 1991 at the Laboratory for Physical Chemistry at ETH under the supervision of Prof. Richard R. Ernst. He then worked at the Scripps Research Institute, La Jolla, as a postdoctoral fellow and at ETH as an Oberassistent. In 1998, he became Associate Professor of Chemistry and Biochemistry and Carlson Chair at Clark University in Worcester, Massachusetts. In 2004, he joined the Department of Chemistry and Biochemistry at Florida State University in Tallahassee as a full professor and the National High Magnetic Field Laboratory as Associate Director for Biophysics. In 2013, he assumed the position of Ohio Research Scholar and full professor at The Ohio State University in Columbus, Ohio, with joint appointments at the Department of Chemistry and Biochemistry and at the College of Medicine. He also serves as NMR Executive Director of OSU's Campus Chemical Instrument Center. His current research activities include the development and application of NMR and computational methods for studying the structure, dynamics, interactions, and function of proteins and as well as for the analysis of complex biological mixtures in the context of metabolomics.

Figure 1.
1D NMR spectra of *E. coli* cell lysate. (A) 1D ^1H spectrum at natural ^{13}C abundance; (B) 1D ^{13}C spectrum of fully ^{13}C labeled sample.

Figure 2.

Representative plots of (A) principal component analysis (PCA) and (B) partial least square discriminant analysis (PLS-DA) obtained from 1D ¹H NMR spectra of brain extracts of *F. hepatica*-infected rats (red circles) and uninfected rats (blue circles) permit the discrimination between the two groups. Reprinted with permission from Saric, J.; Li, J. V.; Utzinger, J.; Wang, Y.; Keiser, J.; Dirnhofer, S.; Beckonert, O.; Sharabiani, M. T.; Fonville, J. M.; Nicholson, J. K.; Holmes, E. *Mol. Syst. Biol.* **2010**, *6*, 396. Copyright 2010 Nature Publishing Group.

Figure 3.

Representative STOCSY plot of a urinary NMR data set. A STOCSY cross-section (A) extracted from 2D STOCSY plot (B) shows correlations of 3-hydroxybutyric acid signals with itself and other metabolites. Positive correlations (red) are arising from structural connectivity between the various 3-hydroxybutyric acid signals, whereas mostly negative correlations (green) belong to pathway connectivities to other metabolites. In the 2D STOCSY plot, high positive correlations between lactate, alanine and glucose signals indicate coordinated excretion of these metabolites. Reproduced from Robinette, S. L.; Lindon, J. C.; Nicholson, J. K. *Anal. Chem.* **2013**, 85, 5297-5303. Copyright 2013 American Chemical Society.

Figure 4.

Comparison of Fourier transformation (FT) and covariance processing (Cov). (A) High resolution FT ^1H - ^1H TOCSY spectrum of four-compound mixture consisting of isoleucine, lysine, carnitine, and shikimate, where the number of increments along the indirect dimension is 512 complex points. (B) Low-resolution FT ^1H - ^1H TOCSY spectrum of four-compound mixture, where the number of increments along the indirect dimension is 96 complex points. (C) High-resolution covariance ^1H - ^1H TOCSY spectrum (Cov) of four-compound mixture obtained using the 2D FT spectrum of Panel (B) as input.

Figure 5.

Backbone topology determination of metabolites identified in *E. coli*. (A) 98 semi-automatically determined *DeCoDeC* consensus traces of 2D ^{13}C - ^{13}C CT-TOCSY along ω_2 that are used to determine the topologies. (B) Display of all unique backbone carbon topologies belonging to 112 spin systems of *E. coli* identified by the combination of the semi-automatic and manual analysis. (C) List of the different topologies identified together with the number of their occurrences (“Occ.”). Compounds with specific names accurately matched BMRB database compounds, whereas compounds referred to as “others”, “amino-acid like”, and “saccharides” were not contained in the database. Reproduced from Bingol, K.; Zhang, F.; Bruschweiler-Li, L.; Brüschweiler, R. *J. Am. Chem. Soc.* **2012**, *134*, 9006-9011. Copyright 2012 American Chemical Society.

Figure 6. 1D ^{13}C NMR spectra of galactose. (A) Regular 1D ^{13}C spectrum; (B) 1D cross-section of 2D ^{13}C CT-TOCSY for galactose α -pyranose; (C) 1D cross-section of 2D ^{13}C CT-TOCSY for galactose β -pyranose.

Figure 7.

Graphical representation of the time zero ^{13}C HSQC (HSQC_0) approach, which extrapolates a series of ^{13}C HSQC experiments with finite coherence transfer times, labeled as HSQC_1 , HSQC_2 and HSQC_3 in the figure, to a virtual HSQC_0 spectrum, which is free of signal attenuation during the coherence transfer period. Reproduced from Hu, K.; Westler, W. M.; Markley, J. L. *J. Am. Chem. Soc.* **2011**, *133*, 1662-1665. Copyright 2011 American Chemical Society.

Figure 8.

Quantitative comparison of experimental and simulated cross-peak integrals of 2D ^{13}C - ^{13}C constant-time (CT) TOCSY acquired at long mixing-time ($\tau_m = 47$ ms). The molecules are (A) fructose β -furanose, (B) glucose β -pyranose, (C) isoleucine and (D) lysine. Reproduced from Bingol, K.; Zhang, F.; Brüscheiler-Li, L.; Brüscheiler, R. *Anal. Chem.* **2013**, 85, 6414-6420. Copyright 2013 American Chemical Society.