Published in final edited form as:

Chemistry. 2018 August 09; 24(45): 11535–11544. doi:10.1002/chem.201800954.

Non-Uniform and Absolute Minimal Sampling for High-Throughput Multidimensional NMR Applications

Dawei Li¹, Alexandar L. Hansen¹, Lei Bruschweiler-Li¹, and Rafael Brüschweiler^{1,2,3,*}

¹Campus Chemical Instrument Center, The Ohio State University, Columbus, Ohio 43210, U.S.A.

²Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, U.S.A.

³Department of Biological Chemistry and Pharmacology, The Ohio State University, Columbus, Ohio 43210, United States

Abstract

Many biomolecular NMR applications can benefit from the faster acquisition of multidimensional NMR data with high resolution and their automated analysis and interpretation. In recent years, a number of non-uniform sampling (NUS) approaches have been introduced for the reconstruction of multidimensional NMR spectra, such as compressed sensing, thereby bypassing traditional Fourier-transform processing. Such approaches are applicable to both biomacromolecules and small molecules and their complex mixtures and can be combined with homonuclear decoupling (pure shift) and covariance processing. For homonuclear 2D TOCSY experiments, absolute minimal sampling (AMS) permits the drastic shortening of measurement times necessary for high-throughput applications for identification and quantification of components in complex biological mixtures in the field of metabolomics. Such TOCSY spectra can be comprehensively represented by graph theoretical maximal cliques for the identification of entire spin systems and their subsequent query against NMR databases. Integration of these methods in web servers permits the rapid and reliable identification of mixture components. Recent progress is reviewed in this Minireview.

Graphical Abstract

*To whom correspondence should be addressed: Rafael Brüschweiler, Ph.D., bruschweiler.1@osu.edu.

Rapid multidimensional NMR for high-throughput screening

1. Introduction

Analysis of molecules and complex molecular mixtures by nuclear magnetic resonance (NMR) spectroscopy enjoys widespread use in many different fields, such as chemistry, biochemistry, biomedicine, and the food sciences.^[1–2] Reasons for this include (i) the high reproducibility of NMR spectra, (ii) the non-destructive nature, (iii) the quantitative character allowing the determination of concentrations, and (iv) the rich information content of NMR spectra for the characterization of molecular structures and dynamics. For samples that exceed a certain level of complexity, the use of two-dimensional (2D) or higher dimensional NMR experiments offers substantially improved spectral resolution and spin-connectivity information greatly facilitating the extraction of relevant (bio)chemical information.

However, acquisition of multidimensional NMR experiments for processing by traditional Fourier transformation (FT) often requires prolonged measurement times. This can have two fundamentally different causes that are important to distinguish in the context of this review. For NMR samples and experiments that have intrinsically low sensitivity, which is typically due to low sample concentration, a low-gamma nucleus for detection, or low natural abundance of the NMR-active spin (such as ^{13}C or ^{15}N), longer measurement times help improve the signal-to-noise ratio (S/N). Since the S/N scales with the square root of the total measurement time, if the S/N for a given sample needs to be increased, for example, by a factor of 10, it will require a 100 times longer measurement time. This situation is sometimes referred to as the “sensitivity-limited regime”. By contrast, if the total NMR time is dictated by the Nyquist sampling requirement rather than sensitivity,^[3] which defines the sampling rate and total number of data points to be collected to achieve a given spectral resolution, one enters the “sampling-limited regime”. For the sensitivity-limited regime, remedies are time-consuming and costly. These include an increase of the number of scans

(by taking advantage of the fact that each scan contributes to the sensitivity of the whole spectrum) or, if possible, higher sample concentration or advanced NMR hardware (higher magnetic field, cryogenically cooled probe).

For the sampling-limited regime, important advances have been made in recent years to substantially cut down the NMR time while retaining high spectral resolution along the indirect dimension(s). This is the main subject of the first part of this review. We will describe the philosophy and methods behind non-standard sampling approaches, which is illustrated for a pure shift TOCSY spectrum of a mixture of natural products. We then focus on absolute minimal sampling (AMS) as a method to reduce the amount of measurement time to an absolute minimum and illustrate it for heteronuclear 3D protein NMR spectra and homonuclear 2D TOCSY. The latter plays a critical role in 2D NMR applications in the field of complex mixture analysis and metabolomics for molecular identification and quantification. Next, the semi-automated analysis of the TOCSY experiments of complex mixtures is discussed based on graph-theoretical methods to extract the relevant spin system information, which is key for the unambiguous identification of the molecular species present in the mixture. Finally, the integration of these methods in web servers for the rapid and reliable identification of mixture components is discussed.

2. Uniform vs. non-uniform sampling

2.1 General overview

Standard discrete Fourier transform, as implemented in the fast Fourier transform algorithm (FFT), is bound to the Nyquist theorem,^[3] which has important practical implications. When applied along an indirect time domain t_1 with a desired digital resolution ν_1 , a maximal evolution time $t_{1,\max} = 1/\nu_1$ is required. Because in practice, the minimally required spectral width, SW_1 , is usually predetermined by the chemical shift dispersion of a sample, which cannot be easily altered, the sampling increment is given by $\Delta t_1 = 1/SW_1$, and defines the total number of complex time increments $N_1 = t_{1,\max}/\Delta t_1$. Importantly, the total measurement time of the multidimensional experiment is directly proportional to N_1 . For a homonuclear 2D ^1H - ^1H Correlation Spectroscopy (COSY)^[4] or Total Correlation Spectroscopy (TOCSY)^[5] experiment, N_1 is typically 256 – 512, leading to measurement times from several hours to overnight.

A logical approach to speed up NMR measurements is to reduce N_1 . This can be achieved in multiple ways, each with its own trade-offs. The simplest way is to collect only a subset of equally spaced t_1 points, for example only the first 25% points, amounting to a 4-fold speed up of the total measurement time. The associated four-fold reduction of $t_{1,\max}$, however, causes a lower resolution ν_1 that may prove to be problematic for a given application. In some cases, the resolution loss can be partially offset by linear prediction methods.^[6] Another possibility is to leave out t_1 points in regular intervals, e.g. collect only every 4th t_1 point. Because this amounts to a four-fold increase of Δt_1 , it is equivalent to a reduction of the spectral width SW to $SW_{1,\text{red}} = SW/4$ (see above). After Fourier transform, the peaks that lie outside of $SW_{1,\text{red}}$ will be folded back (aliased) into the spectrum and can lead to ill-defined chemical shifts and unwanted cross-peak overlaps,^[2] although algorithms exist that mitigate such issues.^[7] Both of these NMR time-saving methods, which have been routinely

used since the advent of multidimensional NMR, share an advantage, namely that processing can still be done by standard discrete Fourier transformation.

In recent years, considerable progress has been made in the development of new methods for the reconstruction of 2D (or higher dimensional) spectra with the original spectral width and high digital resolution $\propto \nu_1$ using only a fraction of increments along the indirect dimension(s). Some of these approaches are specifically suitable for 3D and higher-dimensional spectra, such as GFT,^[8] projection reconstruction,^[9–10] ADAPT-NMR,^[11] and APSY.^[12] All of these approaches employ “radial” sampling by collinearly incrementing two or more indirect evolution times, which selects 2D cross-sections of higher dimensional experiments thereby resolving signal overlap of spectra in lower dimensions. These methods are generally overdetermined in the sense that spectral processing produces uniquely defined spectra without the need for external input about the number of resonances or their shapes.

Other approaches, including filter diagonalization method (FDM),^[13] maximum entropy reconstruction (MEM),^[14–15] multi-dimensional decomposition (MDD),^[16–17] iterative soft thresholding (IST),^[18–20] and SMILE,^[21] are directly applicable also to 2D spectra. They all reconstruct the indirect frequency dimension based on an incomplete set of t_1 points that are irregularly or non-uniformly spaced, generally referred to as “non-uniform sampling” or NUS. Common to these methods is that, unlike traditional FT-based methods, they are underdetermined, i.e. the experimental data is generally consistent with many different possible spectra. In order to make the reconstructed spectra unique and closely resemble the corresponding idealized Fourier transform spectra additional information or constraints are required. Such information can be supplied in a mathematical form, such as the general parametric form of the NMR signal, a formal regularization condition, or empirical rules applied during the algorithmic reconstruction of the NMR spectrum. All of these techniques have the goal to enhance sensitivity or resolution, or both, over conventional, uniform sampling and processing methods.^[22] “Ultrafast” methods are yet another way to speed up 2D NMR experiments that are bound by resolution and sensitivity criteria other than those discussed here.^[23]

In recent years, so-called “compressed sensing” or “compressive sensing” has gained in popularity as a non-parametric approach to reconstruct NMR spectra.^[24–28] Compressed sensing solves a mathematical non-linear least squares problem, which can be expressed as follows:

$$\min_{\mathbf{S}} \{ \|FT^{-1}(\mathbf{S}) - \mathbf{s}\|_{l_2} + \mu \|\mathbf{S}\|_{l_1} \} \quad (1)$$

where $\mathbf{s}$ is the original non-uniformly sampled time-domain signal, $\mathbf{S}$ is a candidate spectrum whose inverse Fourier transform (FT^{-1}) matches $\mathbf{s}$, and hence minimizes the least-squares term (l_2 -norm) on the left, and μ is a regularization parameter that scales the importance of the last term, $\|\mathbf{S}\|_{l_1}$, which is the l_1 -norm of the spectrum corresponding to the sum of the absolute values of all its components. Because for a typical NUS data set $\mathbf{s}$, many different spectra $\mathbf{S}$ exist that are consistent with $\mathbf{s}$, the regularization term with $\mu > 0$ provides a selection criterion that prefers spectra with a “simpler” or sparser structure over spectra with

a “richer” or more complex structure. The regularization concept for general non-linear least squares problems was introduced by Tikhonov to give preference to certain solutions with desirable mathematical or physical properties.^[29] While MEM takes a similar form as Eq. (1), but with a different regularization term, the regularization condition for compressed sensing has a more rigorous mathematical foundation.^[30–31] A strength of Eq. (1) is that no assumption about the shape of the resonances of the spectrum is needed, making this approach applicable to problems where little prior knowledge about the spectrum is available.

In many solution NMR applications, however, a considerable amount of spectral information may already exist or can be obtained in a short amount of time. In the case of *covariance NMR*,^[32] the information about line positions and shapes is taken from one of the frequency dimensions that has been measured and used for the dimension that is to be reconstructed.

2.2 Covariance processing

Covariance NMR improves the resolution of homonuclear 2D NMR experiments that are inherently symmetric with respect to the two frequency axes, such as 2D TOCSY, COSY, or NOESY spectra. Covariance NMR, which was reviewed recently,^[33] can also be generalized to heteronuclear experiments.^[34–37] In *direct covariance* NMR one utilizes the high-resolution spectral content of the direct ω_2 dimension for the reconstruction of the indirect dimension by treating the t_1 modulation as a statistical property subjected to statistical covariance analysis. By using only a relatively small number of t_1 increments, e.g. 48 - 60 increments, a high-resolution spectrum can be reconstructed.^[38] The combination of covariance NMR with non-uniform sampling has been demonstrated.^[39]

Covariance processing can also be applied in an “indirect mode” by using the frequency information collected along the indirect ω_1 dimension and utilize it for the reconstruction of the direct ω_2 dimension.^[40] The application of *indirect covariance* processing to so-called pure shift experiments is particularly instructive.^[41–45] Pure shift experiments provide homonuclear decoupling, i.e. the removal of the splitting caused by scalar J-couplings between proton spins, leading to spectra with narrow lines and reduced overlap.^[46] The 2D F1-PSYCHE-TOCSY pure-shift experiment^[43] decouples the spectrum along the indirect ω_1 dimension, which is achieved by a hard 180° pulse and two low flip angle chirp pulses applied in the presence of a weak pulsed field gradient in the middle of the evolution period (t_1) of a 2D TOCSY pulse sequence. The net effect is a 2D TOCSY spectrum where multiplets in the indirect dimension (F1) are collapsed to singlets (Figure 1).^[47] The gain in resolution, however, is accompanied by a decrease in sensitivity compared to the standard experiment.

The pure shift 2D F1-PSYCHE TOCSY spectrum lends itself to indirect covariance processing to obtain a spectrum that is homonuclear decoupled along both dimensions. Mathematically, this is achieved by matrix algebra applied to the original 2D F1-PSYCHE spectrum represented by the real matrix $\mathbf{F}$:^[40]

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

where $\mathbf{F}^T$ is the transpose of $\mathbf{F}$ and the square-root denotes the matrix-square root of the product, which can be computed by matrix diagonalization or singular value decomposition (SVD).^[48] The resulting spectrum $\mathbf{C}$ is symmetric and homonuclear decoupled along both dimensions (Figure 1). The F1-PSYCHE experiment together with indirect covariance processing is only worthwhile if the spectral resolution along the indirect dimension is sufficiently high so that it is comparable (or better) than the homonuclear scalar J-couplings. For a typical TOCSY experiment, such a high resolution requires a large $t_{1,\max}$ value, which for a uniform sampling schedule translates into a very large number of t_1 increments (N_1) and therefore long measurement times. This makes NUS an attractive choice to either reduce the experimental NMR time while retaining high spectral resolution or to leverage sensitivity through the enhanced sampling of shorter t_1 values.

With covariance NMR, the resonance positions and lineshapes are inferred experimentally. If this is not easily possible, a mathematical ansatz can be made based on NMR theory or heuristics, as is the case in FDM.^[13] In solution NMR, lineshapes have in good approximation Lorentzian shape because the free induction decay (FID) of individual resonances decays (approximately) exponentially. Spectral reconstruction of a sum of exponentially decaying sinusoid functions is the basis of the CRAFT method.^[49–50] CRAFT, which was originally developed for uniformly sampled data sets, i.e. not designed to shorten NMR time, has recently been combined with NUS.^[51]

2.3 Absolute minimal sampling

As it turns out, direct fitting of time-domain data against a parametric form of the spectrum can allow spectral reconstruction using fewer data points and, hence, less measurement time. In the special case where for a given ω_2 frequency the indirect t_1 dimension contains a single resonance at position ω_0 , the frequency and amplitude of a cross-peak along t_1 can be extracted from the measurement of a single increment $t_1 > 0$ of the cosine and sine modulated parts $S_{\exp}(t_1)$ and $C_{\exp}(t_1)$, respectively. This forms the basis of the SPEED approach:^[52]

$$\omega_0 = \arctan\{S_{\exp}(t_1), C_{\exp}(t_1)\}/t_1 \quad (3)$$

where $\arctan(y,x)$ is the two-argument variant of the arctangent function. Because it is usually unknown *a priori* whether there is only a single cross-peak, this relationship is of limited use in practice, but it highlights that it is in principle possible to extract the desired spectral information along indirect dimension(s) with far fewer data points than suggested by Nyquist sampling and Fourier transformation, which in the extreme case of SPEED amounts to a single complex data point.

More generally, when M different resonances exist along t_1 , a non-linear least-squares fit in the time domain can be performed for a sum of M cosine and M sine modulated t_1 traces:

$$C(t_1) = \sum_{k=1}^M A_k \exp(-R_{2,k} t_1) \cos(\omega_k t_1) \quad S(t_1) = \sum_{k=1}^M A_k \exp(-R_{2,k} t_1) \sin(\omega_k t_1) \quad (4)$$

where each resonance k is defined by three fit parameters, the frequency ω_k , volume A_k , and transverse relaxation rate $R_{2,k}$, that can be obtained by minimizing the sum of the l_2 norms of the residuals of the cosine and sine modulated interferograms:[53]

$$\chi^2 = \min\{\|C(t_1) - C_{\text{exp}}(t_1)\| + \|S(t_1) - S_{\text{exp}}(t_1)\|\} \quad (5)$$

To extract this information for M resonances, one needs to collect

$$N_1 \geq 3M/2 \quad (6)$$

complex increments along t_1 , thereby defining the *absolute minimal sampling* (AMS) condition. In many cases, $R_{2,k}$ can be treated as a constant for all resonances, which further reduces the number of required increments to $N_1 \sim M$.^[53] In practice, it is advisable to collect a few additional t_1 increments in order to have an over-determined system of equations. This makes the fitting results more robust with respect to noise and other artifacts and, at the same time, allows one to assess via χ^2 statistics whether the assumption of M resonances contained in the signal is justified. When comparing the AMS approach of Eq. (5) with Eq. (1) it can be seen that it does not require a regularization term. Instead, AMS uses a parameterization of the NMR signal in the time domain where the setting of the number of fitted resonances, M , within the range allowed by the AMS condition naturally constrains the solutions to ones with relatively low complexity.

The AMS method was originally demonstrated for heteronuclear 3D NMR datasets.^[53] Two of the more sensitive 3D NMR experiments for protein backbone resonance assignments are the 3D HNCO and 3D HN(CA)CO. Figure 2 shows representative cross-sections of these experiments of the protein arginine kinase (42 kDa, monomeric) along the indirect carbonyl carbon C' dimension for AMS reconstructed peak positions and amplitudes (red lines) using only 4 – 6 t_1 increments along the C' dimension in comparison to 3D FT experiments collected with 40 – 48 t_1 points. It demonstrates the capability of AMS to reconstruct peak positions with high fidelity using a minimal amount of experimental data along this indirect time domain.

AMS also has useful applicability for homonuclear 2D experiments, such as 2D TOCSY, of complex mixtures as encountered in metabolomics.^[54] As a homonuclear experiment, it has significantly higher sensitivity than the 2D ^{13}C - ^1H HSQC experiment (when at ^{13}C natural abundance). Therefore, many TOCSY applications are sampling rather than sensitivity limited. For reasonably high spectral resolution along the indirect dimension, a TOCSY typically has of the order of 256 complex t_1 increments, requiring about 4 hours to record. For high-throughput applications that may involve hundreds of samples, 4 hours per sample

is impractical, which is the reason why these types of NMR applications rely mostly on 1D ^1H NMR that take only about 10 minutes per sample.

AMS of 2D TOCSY spectra of complex mixtures faces additional challenges when compared to AMS of 3D protein assignment experiments. First, TOCSY spectra display a larger and variable number of cross-peaks: along a ω_1 trace the number of cross-peaks can be as high as ~ 10 . This renders the non-linear least-squares fitting procedure (Eq. (5)) considerably harder for the search of the global minimum. Second, due to the potentially large range of compound concentrations (i.e. peak volumes), the TOCSY cross-peak intensities can display a large dynamic range, which makes the solution to the optimization problem of Eq. (5) more challenging and susceptible to noise, especially for low intensity peaks. In order to overcome these challenges, we find a moderately conservative NUS schedule most effective when paired with an extensive search for the global minimum of Eq. (5).^[54] For the NUS schedule, linear sampling for the initial increments followed by triangular sampling works particularly well with a total of about 16 increments.^[54] For triangular sampling, the spacing between adjacent t_1 points is linearly increased, which amounts to a quadratic distribution of t_1 points. An example for such a sampling schedule uses the following multiples of t_1 : $n = 0, 1, 2, 3, 4, 6, 9, 13, 18, 24, 31, 39, 48, 58, 69, 81$. It follows that for a typical spectral width of 6000 Hz, $t_{1,\text{max}} = 13.5$ ms. For $R_2 = 10 \text{ s}^{-1}$, the NMR signal decays during the evolution period at most to $\exp(-R_2 \cdot t_{1,\text{max}}) = 87\%$, which rationalizes why $R_{2,k}$ does not need to be explicitly included as an AMS fit parameter.

The optimization problem of Eq. (5) is achieved numerically by starting the minimization, e.g. using the Levenberg-Marquardt method, from a large number of different randomly or systematically chosen initial guesses for the initial fit parameters to identify local minima. Comparison of all identified local minima then yields the best estimate for the global minimum of the optimization problem. Only resonance frequencies and amplitudes are used as fit parameters, while $R_{2,k}$ can be kept constant and uniform for all resonances (for example at 10 s^{-1}). This is possible because R_2 relaxation of small molecules is slow and hence plays only a minor role, even at the longest t_1 evolution time $t_{1,\text{max}}$. In practice, the fitting starts with only one resonance per t_1 trace and the number of fitted peaks is iteratively increased until (i) the residual error drops below a certain level proportional to the noise level of the time domain spectrum or (ii) the amplitude of a newly added peak is below 2% of the highest peak in the trace, to account for the fact that real peaks do not have idealized Lorentzian shapes and are often distorted by homonuclear J-couplings and noise, including t_1 -noise. It is advisable to fulfill the AMS condition beyond the minimal requirement and collect more t_1 data points N_1 than the expected number of resonances M . For a maximal number of up to $M = 12$ expected resonances, it is recommended to set $N_1 = 16$. This not only provides more robust fitting results, but it also permits an independent assessment of the maximal number M of expected resonances per t_1 trace.

Next, the full 2D spectrum is reconstructed for downstream analysis from all of the fitted AMS peaks, which requires clustering the fitted peaks in each trace to represent the 2D peak positions observed in a FT spectrum. For this purpose, each fitted peak is convoluted with a narrow 2D Gaussian function followed by peak-picking to define cluster centers. Gaussian convolution effectively provides a smoothing step to account for the lower S/N of the AMS

spectrum, compared with a standard FT spectrum obtained from a much larger number of t_1 increments. Once cluster centers (2D peak positions) are defined, peak amplitudes are obtained by Gaussian mixture analysis under the constraint that the peak volumes are conserved, so that the quantitative nature of AMS fitting is maintained.

Figure 3 shows an example of an AMS-reconstructed 2D TOCSY spectrum (red) measured with 16 t_1 increments of a cell growth base medium (DMEM) in comparison with a standard 2D FT spectrum collected with 256 t_1 increments (blue). The positions of the AMS-derived peaks agree well with the 2D FT results, although the AMS data required only 15 minutes vs. 4 hours for the full 2D FT spectrum. Nonetheless, the effective resolution of AMS along both dimensions is well comparable to the one of the 2D FT spectrum. Moreover, AMS accurately reproduces the peak integrals of the 2D FT spectrum reflecting the relative concentrations of the mixture components. This suggests that 2D AMS TOCSY is suitable for quantitative analysis.

The above example shows that AMS is capable of reconstructing all but the weakest signals in the 2D TOCSY spectrum in a manner that is quantitative both in terms of cross-peak positions and intensities. This is achieved at a small fraction of the total measurement time required for a standard 2D FT experiment: instead of 4 hours, AMS requires only 15 minutes. Such a speed-up makes AMS reconstruction of 2D NMR experiments that are sampling but not sensitivity limited amenable to high-throughput screening and quality control for a wide range of applications. The much shorter acquisition time for AMS invariably causes a lower effective sensitivity, which is manifested in the potential disappearance of the weakest signals.

3. Automated analysis of TOCSY-type spectra

The speed up of NMR measurements is clearly valuable as it allows a larger number of NMR samples to be measured per spectrometer per day. On the other hand, time-savings should also be viewed in the context of the total NMR workflow starting with sample preparation and ending with data analysis and interpretation. If data analysis is the bottleneck taking a multiple of the time used for data collection, e.g. of a suite of NUS 3D protein assignment spectra or a 2D AMS TOCSY spectrum of a mixture, the availability of ever faster NMR methods will have diminishing returns for the entire project. It is therefore only logical that significant efforts be devoted to the automated analysis of multidimensional NMR spectra to ease such bottlenecks. In the case of protein assignments, a number of software exist that assist manual analysis or are fully automated.^[55–61] For the automated analysis of multi-dimensional NMR spectra of metabolomics samples, available software and webserver are more limited.^[62–63]

For 2D TOCSY (or 3D HSQC-TOCSY experiments) of complex mixtures, the large number of cross-peaks makes the manual analysis of such spectra a formidable challenge. A first step in spectral analysis is the identification of cross-peaks and their positions through peak-picking. Most NMR software packages use their own peak-picker.^[64–67] For crowded spectra as encountered in complex mixtures, a good peak-picker should be able to (i) identify all true cross-peaks and their positions that have a S/N that markedly exceed the

noise level and are reasonably well separated, (ii) identify cross-peak candidates that partially overlap with other signals, (iii) provide a list of potential cross-peaks that could be true or could correspond to noise, and (iv) identify cross-peak-like features that are artifacts, such as t_1 -noise. The power of automated peak-picking can be seen in Figure 4. A “smart” peak-picker, which is integrated in the COLMAR webserver, is capable of distinguishing between true peaks, including shoulder peaks, t_1 -noise, and weak features that are likely to belong to true cross-peaks.

In a subsequent step, the picked peaks need to be placed in context with respect to each other for the construction of entire spin systems. Recent research has indicated new ways to address this problem using mathematical graph theory^[68] along with available algorithms for the analysis of subgraphs with certain properties that relate to individual molecules. For sufficiently long mixing times ($\tau_m \approx 100$ ms) the 2D ^1H - ^1H TOCSY spectrum of a molecule with a single spin system has cross-peaks connecting each resonance with each other.^[5] This spectral structure can be converted into a graph that represents each diagonal peak (resonance) by a *node* and each cross-peak by an *edge*. Identification of such graphs where each node is connected to each other by an edge, which are known as complete graphs,^[68] from the graph generated from the diagonal and cross-peaks of a TOCSY spectrum represents a recipe for the interpretation of TOCSY spectra in terms of entire spin systems^[69–70] which can be assigned to known and unknown mixture components.^[71] In practice, however, complex mixtures may have compounds whose resonances overlap, which leads to graphs that are no longer complete. As a remedy, one can search for subgraphs that are complete in themselves, which are termed cliques. Hence, the identification of maximal cliques returns the spin systems of interest.^[72] Sophisticated algorithms exist that can identify all maximal cliques in complex graphs, which includes the popular Bron-Kerbosch algorithm.^[73] Figure 5 shows a TOCSY spectrum of a 19-metabolite model mixture (plus DSS for referencing) with its entire connectivity graph, which was derived by automated peak-picking using the peak-picker of Figure 4 and which contains 306 nodes and 2600 edges. Although the graph looks very complex to the human eye, the maximal clique algorithm is able to accurately extract entire spin systems that can be assigned to individual molecules (Figure 5C).

To facilitate the application of the maximal clique algorithm, the COLMAR TOCSY web server is now publicly available (<http://spin.ccic.ohio-state.edu/index.php/tocsy>) (Figure 5). It follows a similar design as the COLMARm server,^[63] allowing users to first upload their frequency domain TOCSY spectrum. The web server will perform peak picking and produce a peak list with peak amplitudes for quantification. In the case of AMS TOCSY, the user can also upload the peak list (together with peak integrals) so that the server can directly use the peak list for analysis and use the uploaded spectrum for visualization only. As an additional option, the web server also provides spectral referencing based on pattern matching of commonly found metabolites.^[63]

The server provides two database query methods, which can be selected by the user. Method 1 performs direct matching, where expected ^1H - ^1H TOCSY peaks of each compound in our TOCCATA database^[74] are matched against all experimental cross-peaks (but not diagonal peaks). This algorithm is highly efficient and works well for relatively simple TOCSY

spectra, but it is less optimal for more complex TOCSY spectra with a large number of overlapping peaks, where it may produce false matches due to (nearly) degenerate resonances.

In Method 2, cross-peaks are first subjected to *maximal clique analysis* to obtain spin systems. Next, each spin system is matched against all database spin systems to identify known compounds. The web server also returns those spin systems that do not have any good database match, indicating that they belong to unknown metabolites, which can be further studied, for example, by the SUMMIT MS/NMR approach.^[75–76] Method 2 is computationally more expensive. Depending on the number of cliques, it typically takes several minutes and sometimes hours to complete, whereas Method 1 takes only seconds.

This example demonstrates the power of automated analysis of TOCSY spectra, which are rich in molecular information but whose manual analysis can be tedious and time-consuming. Homonuclear ^1H - ^1H TOCSY is intrinsically endowed with high sensitivity, which makes it ideally suitable (i) for the analysis of low abundance compounds that escape detection, for example, by 2D ^{13}C - ^1H HSQC (at ^{13}C natural abundance) and for (ii) high-throughput applications for the monitoring of higher concentration compounds ($> 100\ \mu\text{M}$) by AMS.

4. Concluding remarks and outlook

NMR spectroscopy is widely credited for its ability to provide quantitative high-quality data on a broad range of molecular systems in a highly reproducible manner. A drawback often associated with NMR, especially when using 2D or higher dimensional NMR, are the long measurement times. In this review we described how such limitations are being addressed with recent progress for the shortening of the measurement of 2D and higher dimensional NMR data sets of experiments that are primarily sampling limited, which includes heteronuclear 3D NMR experiments for resonance assignments of isotopically labeled proteins and many homonuclear 2D experiments, such as 2D TOCSY.

Maximal speed can be obtained by the combination of parametric approaches, such as AMS, which assume certain mathematical properties of the resonances based on experience or NMR theory, in combination with non-uniform sampling (NUS). The AMS TOCSY-type experiments provide high resolution and quantitative cross-peak volumes, together with intramolecular connectivity information allowing one to identify unique spin systems and molecules. We illustrated how these steps can be largely automated and interlinked by direct time-domain fitting procedures and potent graph theoretical analysis of the resulting spectra. The quantitative nature of both traditional and 2D AMS TOCSY spectra can be used to monitor concentration changes of individual components as a function of time or across a cohort of samples. Identification of components is easiest performed by querying the spin systems with their chemical shifts against NMR databases, such as HMDB,^[77] BMRB,^[78] or COLMAR.^[79] For unknown molecules, i.e. molecules whose spectral information is not contained in the databases, such spin system information provides a useful starting point toward a full structural characterization even in the context of a complex mixture.^[80]

Multidimensional NMR has for many years been a powerful source of molecular information. With traditional approaches, the generation and extraction of such information is often quite time-consuming and labor-intensive. New developments in NUS NMR together with dedicated reconstruction algorithms and automation procedures, including those described in this review, have started to chip away at these long-standing limitations. It readily opens the door for the faster and easier access to high-resolution spectral NMR information for the identification, quantification, and structural and dynamic characterization of biomacromolecules and molecules in complex biological mixtures. These advances will likely broaden their use for routine screening and other high-throughput applications bringing to bear the full power of nuclear magnetic resonance to a wide range of new applications.

Acknowledgment

This work was supported by the National Institutes of Health (grant R01 GM 066041).

About the authors:

Da-Wei Li is a research scientist at the Campus Chemical Instrument Center (CCIC) NMR Facility at the Ohio State University. He received both his B.S. degree in Physics and Ph.D. in Condensed Matter Physics from Nanjing University, Jiangsu Province, China. He then focused on computational chemistry and biophysics as a postdoctoral fellow at Clark University, Worcester, MA, and Florida State University, Tallahassee, FL. At OSU, he continues his journey in scientific computing for the structural dynamic interpretation of NMR data of proteins and NMR-based metabolomics methods developments.

Alexandar L. Hansen is a research scientist at the Campus Chemical Instrument Center (CCIC) NMR Facility at the Ohio State University. He received his B.A. degree in Chemistry and Music at Carthage College, Kenosha, WI. This was followed by his Ph.D. in Physical Chemistry at the University of Michigan, Ann Arbor, MI developing NMR methods for studying the dynamics of nucleic acids. He then worked as a postdoctoral fellow at the University of Toronto, Canada, where he further developed methods for studying excited states in proteins by NMR. At OSU, Alex continues to pursue his interests in NMR methods development and data analysis pertaining to biomolecular structure and dynamics.

Lei Bruschweiler-Li is a research scientist at the Campus Chemical Instrument Center (CCIC) NMR Facility at the Ohio State University. She received her B.S. degree in Biochemical Engineering at Shanghai University of Science and Technology (now Shanghai University), Shanghai, China and continued with a Ph.D. in Biochemistry at ETH, Zurich, Switzerland. Supported by a postdoctoral fellowship from the Swiss National Science Foundation, she joined U. Massachusetts Medical School, Worcester, MA, before moving to Florida State University, Tallahassee, FL, as Scholar/Scientist. At OSU, she pursues NMR-based research with a wide range of CCIC users on proteins and metabolomics.

Rafael Brüschweiler is a full professor at the Departments of Chemistry & Biochemistry and Biological Chemistry & Pharmacology at the Ohio State University, an Ohio Research

Scholar, and Executive Director of CCIC NMR. He received his B.S. degree in Physics and his Ph.D. in Physical Chemistry at ETH, Zurich, Switzerland. He was a postdoctoral fellow at the Scripps Research Institute, La Jolla, CA, which was followed by a Habilitation at ETH, Zurich. Before joining OSU, he was the Carlson Chair of Chemistry at Clark University, Worcester, MA, the George Matthew Edgar Professor of Chemistry and Biochemistry at Florida State University and the Director for Biophysics at the National High Magnetic Field Laboratory in Tallahassee, FL. He is an elected fellow of the American Physical Society and the AAAS. His research group studies protein dynamics, interactions, and function as well as metabolomics using NMR spectroscopy and high-performance computation by developing and applying new experimental and computational methods.

References

- [1]. Ernst RR, Bodenhausen G, Wokaun A, Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Clarendon Press, 1991.
- [2]. Cavanagh J, Fairbrother WJ, Rance M, Palmer AG, Skelton NJ, Protein NMR Spectroscopy : Principles and Practice, Academic Press, 2007.
- [3]. Proakis JG, Manolakis DG, Digital Signal Processing, Pearson Prentice Hall, 2007.
- [4]. Rance M, Sørensen OW, Bodenhausen G, Wagner G, Ernst RR, Wüthrich K, Biochem. Biophys. Res. Commun 1983, 117, 479–485. [PubMed: 6661238]
- [5]. Braunschweiler L, Ernst RR, J. Magn. Reson 1983, 53, 521–528.
- [6]. Led JJ, Gesmar H, Chem. Rev 1991, 91, 1413–1426.
- [7]. Foroozandeh M, Jeannerat D, Magn. Reson. Chem 2015, 53, 894–900. [PubMed: 26289946]
- [8]. Kim S, Szyperski T, J. Am. Chem. Soc 2003, 125, 1385–1393. [PubMed: 12553842]
- [9]. Kup e , Freeman R, J. Am. Chem. Soc 2003, 125, 13958–13959. [PubMed: 14611222]
- [10]. Coggins BE, Venters RA, Zhou P, J. Am. Chem. Soc 2004, 126, 1000–1001. [PubMed: 14746450]
- [11]. Lee W, Hu K, Tonelli M, Bahrami A, Neuhardt E, Glass KC, Markley JL, J. Magn. Reson 2013, 236, 83–88. [PubMed: 24091140]
- [12]. Hiller S, Fiorito F, Wuthrich K, Wider G, Proc. Natl. Acad. Sci. U. S. A 2005, 102, 10876–10881. [PubMed: 16043707]
- [13]. Mandelshtam VA, Taylor HS, Shaka AJ, J. Magn. Reson 1998, 133, 304–312. [PubMed: 9716473]
- [14]. Hyberts SG, Takeuchi K, Wagner G, J. Am. Chem. Soc 2010, 132, 2145–2147. [PubMed: 20121194]
- [15]. Hoch JC, Maciejewski MW, Mobli M, Schuyler AD, Stern AS, Acc. Chem. Res 2014, 47, 708–717. [PubMed: 24400700]
- [16]. Orekhov VY, Ibraghimov I, Billeter M, Biomol J. NMR 2003, 27, 165–173.
- [17]. Orekhov VY, Jaravine VA, Prog. Nucl. Magn. Reson. Spectrosc 2011, 59, 271–292. [PubMed: 21920222]
- [18]. Papoulis A, IEEE Trans. Circuits Syst 1975, 22, 735–742.
- [19]. Stern AS, Donoho DL, Hoch JC, J. Magn. Reson 2007, 188, 295–300. [PubMed: 17723313]
- [20]. Hyberts SG, Milbradt AG, Wagner AB, Arthanari H, Wagner G, Biomol J. NMR 2012, 52, 315–327.
- [21]. Ying J, Delaglio F, Torchia DA, Bax A, Biomol J. NMR 2017, 68, 101–118.
- [22]. Palmer MR, Suiter CL, Henry GE, Rovnyak J, Hoch JC, Polenova T, Rovnyak D, J. Phys. Chem. B 2015, 119, 6502–6515. [PubMed: 25901905]
- [23]. Giraudeau P, Frydman L, Annu. Rev. Anal. Chem 2014, 7, 129–161.
- [24]. Kazimierczuk K, Orekhov VY, Angew. Chem. Int. Ed 2011, 50, 5556–5559.

- [25]. Holland DJ, Bostock MJ, Gladden LF, Nietlispach D, Angew. Chem. Int. Ed 2011, 123, 6678–6681.
- [26]. Kazimierczuk K, Orekhov VY, J. Magn. Reson 2012, 223, 1–10. [PubMed: 22960668]
- [27]. Sun S, Gill M, Li Y, Huang M, Byrd RA, Biomol J. NMR 2015, doi: 10.1007/s10858-015-9923-x.
- [28]. Stern AS, Hoch JC, Magn. Reson. Chem 2015, 53, 908–912. [PubMed: 26256110]
- [29]. Tikhonov AN, Arsenin VI, Solutions of Ill-Posed Problems, Winston, 1977.
- [30]. Donoho DL, Comm. Pure Appl. Math 2006, 59, 797–829.
- [31]. Candès EJ, Romberg JK, Tao T, Comm. Pure Appl. Math 2006, 59, 1207–1223.
- [32]. Brüschweiler R, Zhang F, J. Chem. Phys 2004, 120, 5253–5260. [PubMed: 15267396]
- [33]. Jaeger M, Aspers RLEG, Annu. Rep. NMR Spectrosc 2014, 83, 271–349.
- [34]. Lescop E, Brutscher B, J. Am. Chem. Soc 2007, 129, 11916–11917. [PubMed: 17845051]
- [35]. Snyder DA, R. Brüschweiler, J. Phys. Chem. A 2009, 113, 12898–12903. [PubMed: 19810742]
- [36]. Martin GE, Thomas Williamson R, Blinov KA, Anklin CG, Bermel W, Magn. Reson. Chem 2012, 50, 691–695. [PubMed: 22930548]
- [37]. Harden BJ, Frueh DP, in Protein NMR: Methods and Protocols, doi: 10.1007/978-1-4939-7386-6_16 (Ed.: Ghose R), Humana Press, New York, NY, 2018, pp. 353–373.
- [38]. Chen Y, Zhang F, Bermel W, Brüschweiler R, J. Am. Chem. Soc 2006, 128, 15564–15565. [PubMed: 17147346]
- [39]. Lafon O, Hu B, Amoureux J-P, Lesot P, Chem. Eur. J 2011, 17, 6716–6724. [PubMed: 21563219]
- [40]. Zhang F, Brüschweiler R, J. Am. Chem. Soc 2004, 126, 13180–13181. [PubMed: 15479045]
- [41]. Morris GA, Aguilar JA, Evans R, Haiber S, Nilsson M, J. Am. Chem. Soc 2010, 132, 12770–12772. [PubMed: 20795722]
- [42]. Aguilar JA, Colbourne AA, Cassani J, Nilsson M, Morris GA, Angew. Chem. Int. Ed 2012, 51, 6460–6463.
- [43]. Foroozandeh M, Adams RW, Nilsson M, Morris GA, J. Am. Chem. Soc 2014, 136, 11867–11869. [PubMed: 25111063]
- [44]. Fredi A, Nolis P, Cobas C, Martin GE, Parella T, J. Magn. Reson 2016, 266, 16–22. [PubMed: 27003379]
- [45]. Fredi A, Nolis P, Cobas C, Parella T, J. Magn. Reson 2016, 270, 161–168. [PubMed: 27494746]
- [46]. Zangger K, Prog. Nucl. Magn. Reson. Spectrosc 2015, 86-87, 1–20. [PubMed: 25919196]
- [47]. Kakita VMR, Hosur RV, Chemphyschem 2016, 17, 2304–2308. [PubMed: 27124764]
- [48]. Trbovic N, Smirnov S, Zhang F, Brüschweiler R, J. Magn. Reson 2004, 171, 277–283. [PubMed: 15546754]
- [49]. Krishnamurthy K, Magn. Reson. Chem 2013, 51, 821–829. [PubMed: 24154986]
- [50]. Krishnamurthy K, Sefler AM, Russell DJ, Magn. Reson. Chem 2017, 55, 224–232. [PubMed: 27160956]
- [51]. Krishnamurthy K, Hari N, Magn. Reson. Chem 2017, doi: 10.1002/mrc.4664.
- [52]. Kup e E, Freeman R, Magn. Reson. Chem 2007, 45, 711–713.
- [53]. Hansen AL, Brüschweiler R, Angew. Chem. Int. Ed 2016, 55, 14169–14172.
- [54]. Hansen AL, Li D, Wang C, Brüschweiler R, Angew. Chem. Int. Ed 2017, 129, 8261–8264.
- [55]. Bartels C, Billeter M, Guntert P, Wuthrich K, J. Biomol. NMR 1996, 7, 207–213. [PubMed: 22911044]
- [56]. Bartels C, Güntert P, Billeter M, Wüthrich K, J. Comput. Chem 1997, 18, 139–149.
- [57]. Moseley HNB, Monleon D, Montelione GT, Methods Enzymol 2001, 339, 91–108. [PubMed: 11462827]
- [58]. Jung Y-S, Zweckstetter M, J. Biomol. NMR 2004, 30, 11–23. [PubMed: 15452431]
- [59]. Bahrami A, Assadi AH, Markley JL, Eghbalnia HR, PLoS Comp. Biol 2009, 5, e1000307.
- [60]. Schmidt E, Güntert P, J. Am. Chem. Soc 2012, 134, 12817–12829. [PubMed: 22794163]
- [61]. Lee W, Markley JL, Bioinformatics 2017, doi: 10.1093/bioinformatics/btx785.

- [62]. Xia J, Bjorn Dahl TC, Tang P, Wishart DS, BMC Bioinformatics 2008, 9, 507. [PubMed: 19040747]
- [63]. Bingol K, Li D-W, Zhang B, Bruschweiler R, Anal. Chem 2016, 88, 12411–12418. [PubMed: 28193069]
- [64]. Garrett DS, Powers R, Gronenborn AM, Clore GM, J. Magn. Reson 1991, 95, 214–220.
- [65]. Koradi R, Billeter M, Engeli M, Güntert P, Wüthrich K, J. Magn. Reson 1998, 135, 288–297. [PubMed: 9878459]
- [66]. Alipanahi B, Gao X, Karakoc E, Donaldson L, Li M, Bioinformatics 2009, 25, i268–i275. [PubMed: 19477998]
- [67]. Würz JM, Güntert P, Biomol J. NMR 2017, 67, 63–76.
- [68]. Gross JL, Yellen J, Graph Theory and Its Applications, 2nd ed., Chapman & Hall/CRC, Boca Raton, 2006.
- [69]. Oschkinat H, Holak TA, Cieslar C, Biopolymers 1991, 31, 699–712. [PubMed: 1932568]
- [70]. Zhang F, Bruschweiler-Li L, Bruschweiler R, J. Am. Chem. Soc 2010, 132, 16922–16927. [PubMed: 21062057]
- [71]. Bingol K, Zhang F, Bruschweiler-Li L, Bruschweiler R, J. Am. Chem. Soc 2012, 134, 9006–9011. [PubMed: 22540339]
- [72]. Li D-W, Wang C, Bruschweiler R, J. Biomol. NMR 2017, 68, 195–202. [PubMed: 28573376]
- [73]. Bron C, Kerbosch J, Commun. ACM 1973, 16, 575–577.
- [74]. Bingol K, Bruschweiler-Li L, Li DW, Bruschweiler R, Anal. Chem 2014, 86, 5494–5501. [PubMed: 24773139]
- [75]. Bingol K, Bruschweiler-Li L, Yu C, Somogyi A, Zhang FL, Bruschweiler R, Anal. Chem 2015, 87, 3864–3870. [PubMed: 25674812]
- [76]. Wang C, He L, Li D-W, Bruschweiler-Li L, Marshall AG, Bruschweiler R, J. Proteome Res 2017, 16, 3774–3786. [PubMed: 28795575]
- [77]. 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 Res 2007, 35, D521–D526. [PubMed: 17202168]
- [78]. 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]
- [79]. Zhang F, Robinette SL, Bruschweiler-Li L, Bruschweiler R, Magn. Reson. Chem 2009, 47, S118–S122. [PubMed: 19634130]
- [80]. Robinette SL, Bruschweiler R, Schroeder FC, Edison AS, Acc. Chem. Res 2012, 45, 288–297. [PubMed: 21888316]

Figure 1.

Comparison of standard 2D TOCSY (left) of mixture of testosterone and estradiol (structures shown as inserts in right panel) with F1-PSYCHE-TOCSY (right) using 5% NUS sampling using a Poisson gap schedule after IST reconstruction and indirect covariance processing. The homonuclear decoupling along with NUS sampling allowing longer $t_{1,\max}$ yields significantly improved spectral resolution along ω_1 , which is mapped onto ω_2 by indirect covariance processing. Figure adopted from Ref. ^[47] with permission.

Figure 2.

Absolute Minimal Sampling (AMS) applied to 3D HNCO (Panels A, B) and 3D HN(CA)CO experiments (Panels C, D) of arginine kinase (42 kDa) for protein resonance assignments. The 4 panels belong to 2 different residues of AK as indicated in the figures. The solid black lines correspond to 1D cross-section along the $\omega_1(\text{CO})$ dimension of a regular 3D FT spectrum with 44 complex t_1 points, whereas the red bars correspond to the AMS-derived frequencies and amplitudes obtained with only 4 t_1 points (Panels A, B) or 6 t_1 points (Panels C, D). The heights of the red bars reflect fitted resonance amplitudes A_k (Eq. (4)). Figure adopted from Ref. [53] with permission.

Figure 3.

Comparison of a regular 2D FT TOCSY spectrum (blue) of a cell growth medium with AMS reconstruction (red). In A,B, the traditional FT spectrum collected with 256 complex t_1 increments is shown with blue contours. In A, the AMS result with only 16 t_1 increments is shown as red dots. In B, a Gaussian reconstruction of the AMS result (see main text) is shown with red contours. Individual ω_1 cross-sections taken at 3.02 (C) and 1.43 (D) ppm in ω_2 , as indicated by black dashed lines in A and B, highlight the quantitative nature of the AMS results. The blue dashed line indicates the cumulative sum of the FT spectrum, while

the red bars are the relative amplitudes and positions of the AMS peaks. Note that the AMS amplitudes correspond to integrals of the FT resonances.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4.

Illustration of smart peak-picker applied to 2D TOCSY spectrum of the hydrophilic *E.coli* lysate. The colored dots correspond to the picked cross-peaks returned by the peak-picker along with color coding, which represents the assessment by the software of the likely nature of the peak. The blue dots are considered as “true” cross-peaks, the green dots as “false” cross-peaks (mostly t_1 artifacts), and the “red” cross-peaks as potential true peaks, which have low intensity or are shoulder peaks of stronger cross-peaks.

Figure 5.

Illustration of maximal clique method for the extraction of spin systems from redundant cross-peak information of a 2D TOCSY spectrum of a complex mixture. A. Region of a 2D ^1H - ^1H TOCSY spectrum of a 20-compound model mixture consisting of 19 metabolites and DSS. The red, blue, and green circles belong to lysine, arginine, and ascorbate, respectively, with lysine and arginine showing overlaps of three of their resonances. B. Representation of the 804 cross-peaks picked in the 2D TOCSY spectrum as a graph consisting of a total of 2600 edges and 306 nodes after removing all maximal cliques of size 2 (two nodes

connected by an edge). C. Analysis of the graph of Panel B by the maximal clique method produces a non-redundant set of connected and disconnected maximal cliques that can be directly assigned to all spin systems of the mixture components that contain > 2 spins. Figure reprinted with permission from Ref. [72].

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 6.

Spectral analysis and display by COLMAR TOCSY webserver of 2D TOCSY spectra in terms of maximal cliques (Method 2, see main text) exemplified for a complex metabolite mixture that contains valine. Each of the 9 subpanels depicts an enlarged spectral region with the correlated cross-peaks labeled by symbols with different colors. Magenta circles reflect picked peaks that match cross-peak patterns (red ellipses with 0.02 ppm width along both dimensions) belonging to the corresponding database compound. The blue squares belong to the maximal clique defined by the experimental TOCSY cross-peaks. The interactive webserver interface allows users to select different cliques and query them against the COLMAR TOCSY database with > 700 metabolites.