Supporting Information

TOCCATA: A Customized Carbon TOCSY NMR Metabolomics Database

Kerem Bingol,[1,2] Fengli Zhang, [2] Lei Bruschweiler-Li,[2,3] and Rafael Brüschweiler[*1,2,3]

[1]Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306
[2]National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32310
[3]Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32306

Spectral Analysis

The deconvolution of the 2D ^{13}C-^{13}C CT-TOCSY, represented by a N_1xN_2 matrix **T**, was performed by adapting the *DeCoDeC* approach to ^{13}C-^{13}C TOCSY (*DeCoDeC* stands for Demixing by Consensus Deconvolution and Clustering,[1] which was originally developed for the analysis of ^{1}H and natural abundance ^{13}C NMR spectra) as described previously.[2] Peak picking of the cross-peaks of matrix **T** yielded a list (k,k') where k and k' denote the cross-peak position along the two frequency axes. In order to minimize the influence of those parts of **T** that are close to the diagonal, the intensities of all diagonal peaks were set to the largest peak intensity of the rest of the spectrum (see below). Next, for each cross-peak pair (k,k') and (l,l'), which are placed symmetrically with respect to the diagonal, the k^{th} and l^{th} row are extracted from **T** to obtain the consensus trace:

$$q_j^{(kl)} = \min(T_{kj}, T_{lj}) \tag{S1}$$

where index j = 1,..., N_2. The enlargement of the diagonal peaks of **T** ensures that Eq. (S1) is dominated by cross-peaks rather than diagonal peaks. The complete set of consensus traces $\mathbf{q}^{(kl)}$ was subsequently subjected to clustering for the identification of those traces that represent 1D ^{13}C spectra of individual spin systems. For this purpose, 1D ^{13}C consensus traces $\mathbf{q}^{(kl)}$ were quantitatively compared to each other via the inner product

$$P_{kl,mn} = \sum_{j=1}^{N_2} q_j^{(kl)} q_j^{(mn)} \Big/ \left(\left\| q^{(kl)} \right\| \cdot \left\| q^{(mn)} \right\| \right) \tag{S2}$$

where the L2-norm of a consensus trace is given by

$$\left\| q^{(kl)} \right\| = \left[\sum_{j=1}^{N_2} (q_j^{kl})^2 \right]^{1/2} \tag{S3}$$

$1 - P_{kl,mn}$ defines a similarity measure between pairs of traces, which permits clustering, e.g., using the agglomerative hierarchical cluster algorithm as implemented in the subroutine 'linkage' of the Matlab software package. The clustering result can be displayed as a dendrogram.

Preparation and experimental details for amino-acid mixture

A uniformly ^{13}C-labeled algal amino acid mixture, purchased from Sigma-Aldrich, was prepared by dissolving 0.5 mg mixture in 2 ml D_2O. The resulting suspension was centrifuged and the supernatant was used for measurements.

A 2D ^{13}C-^{13}C CT-TOCSY[3] data set was collected with 512 N_1 and 2048 N_2 complex data points, with 38 ms FLOPSY-16 mixing[4] at 700 MHz proton frequency with 110 pm ^{13}C spectral width at 25^{o} C. The NMR data were zero-filled, Fourier transformed, phase and baseline corrected using NMRPipe,[5] and converted to a Matlab-compatible format for subsequent processing and analysis.

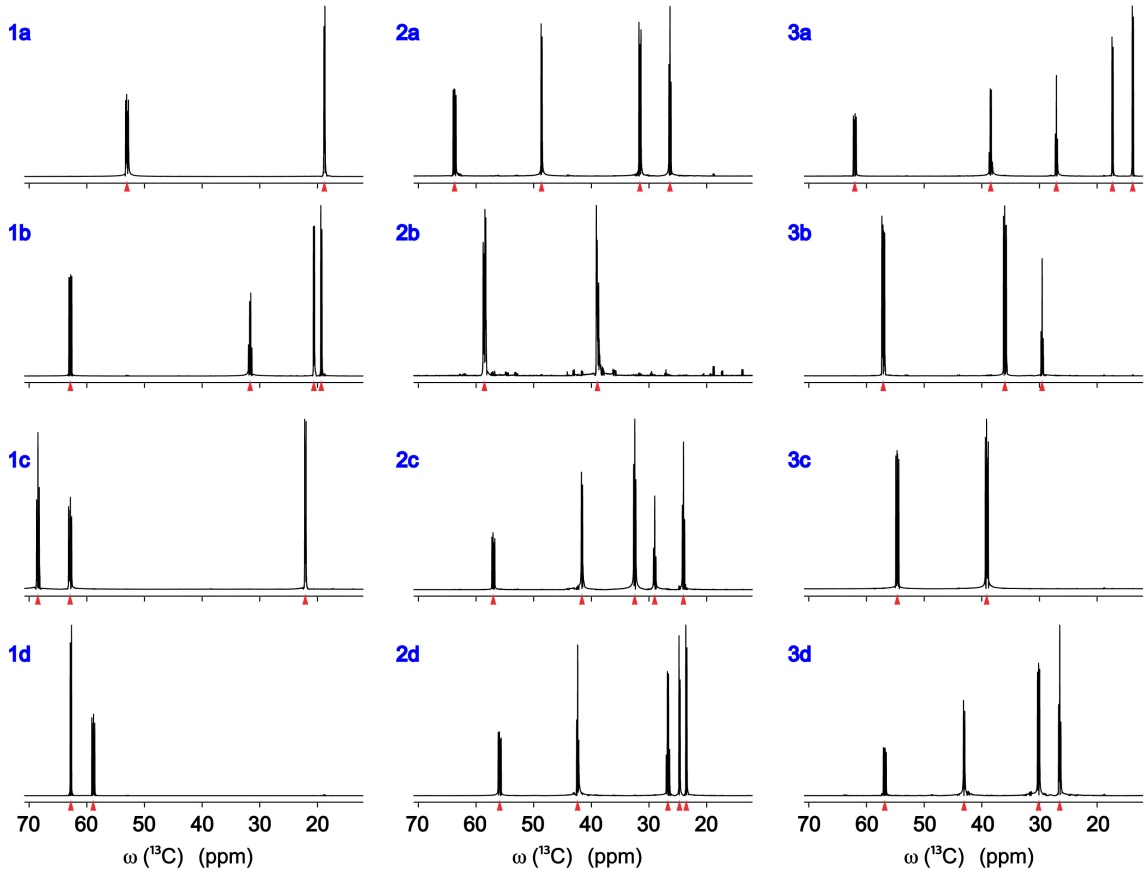

Figure S-1. Deconvolution and TOCCATA database querying of 2D ^{13}C-^{13}C CT-TOCSY spectrum of amino acid mixture. The resulting deconvoluted 1D ^{13}C TOCSY traces belong to: alanine (1a), valine (1b), threonine (1c), serine (1d), proline (2a), phenylalanine (2b), lysine (2c), leucine (2d), isoleucine (3a), glutamate (3b), aspartate (3c) and arginine (3d).

Table S-1. TOCCATA query results of ^{13}C TOCSY traces of amino acid mixture of Figure S-1. Parameters RMSD, M, Shift are defined as in Table 1 and Table 2 of the main text.

	RMSD	**M**	**Shift**		**RMSD**	**M**	**Shift**
Alanine	0.056	0	0.126	**Lysine**	0.023	0	0.142
Valine	0.072	0	0.129	**Leucine**	0.064	0	0.123
Threonine	0.080	0	0.181	**Isoleucine**	0.068	0	0.118
Serine	0.037	0	0.174	**Glutamate**	0.034	0	0.218
Proline	0.027	0	0.152	**Aspartate**	0.066	0	0.182
Phenylalanine	0.025	0	0.181	**Arginine**	0.032	0	0.131

References

(1) Bingol, K.; Brüschweiler, R. *Anal. Chem.* **2011**, *83*, 7412-7417.
(2) Bingol, K.; Zhang, F.; Bruschweiler-Li, L.; Brüschweiler, R. *J. Am. Chem. Soc.* **2012**, *134*, 9006-9011.
(3) Eletsky, A.; Moreira, O.; Kovacs, H.; Pervushin, K. *J. Biomol. NMR* **2003**, *26*, 167-179.
(4) Kadkhodaie, M.; Rivas, O.; Tan, M.; Mohebbi, A.; Shaka, A.J. *J. Magn. Reson.* **1991**, *91*, 437-443.
(5) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277-293.