

Supporting Information

Extraction of Individual Spectra from Mixture data Based on High-Resolution ^{13}C - ^{13}C NMR Correlation Spectrum and DECODE Procedure

Jin Wook Cha^{†,‡} and Sunghyoun Park^{*,†}

[†]College of Pharmacy, Natural Product Research Institute, Seoul National University, Seoul, 08826, Korea

[‡]Natural Product Informatics Research Center, KIST Gangneung Institute of Natural Products, Gangneung, 25451, Korea

E-mail: psh@snu.ac.kr.

*To whom correspondence should be addressed

Table of Contents

Detailed explanation on DECODE procedure.....	S3
Mathematical implementation of the DECODE procedure.....	S5
Figures.....	S8
Reference.....	S10

Detailed explanation on DECODE procedure

Since the NUS-mHMBC-iCov approach did not provide satisfactory results for mixture deconvolution, we devised a stepwise operation to make it better suited for providing molecule-wide spin network information. The first consideration was that a single HMBC focusing on typical $^2\text{-}^3J_{\text{CH}}$ of 8 Hz may not give all the carbon correlation information on the HMBC-iCov spectra, because some carbon-proton pairs have lower coupling constants. Therefore, two HMBC spectra with delays optimized for 8 Hz and 4 Hz were obtained separately and merged in the frequency domain. On top of this, an HSQC spectrum was also merged so that much-isolated carbons, i.e. methoxy carbons, can also appear on HMBC-iCov spectrum at its own chemical shift position. The actual spectral merging required slight modifications in the HSQC and HMBC pulse sequence so that they have the same pulse length as shown in Figure S1B. This modification is to ensure that the peaks in the different spectra have the same J_{HH} modulation and multiplet patterns for the merging. In the merging process, the maximum intensities were taken as the final values to avoid repetitive addition of the same peak on different spectra. Although spectroscopically simple, the addition of HSQC and two HMBC's is clearly advantageous in giving the complete molecular correlation network, as shown in Figure 3A. Second, as partial proton overlaps on the F_2 projections of HMBC peaks may generate false positive intermolecular correlation on the iCov spectrum, they were addressed by the spectral moment filter (here, the first moment filter) before the iCov operation. The principle of the first moment filter was demonstrated before¹ and conceptually similar to peak centroiding in mass spectrometry. Every signal on the spectrum were associated with the mean position value of the corresponding peak. When performing the iCov operation, the mean position values of each element of the two vectors were considered. If they differ by a certain margin, i.e. elements from different peaks, the products of the two elements were multiplied by 0 (filtered). Therefore, this filter prevents the false positive correlation due to proton projection overlaps on the HMBC spectrum. The third consideration is about the disproportionately large signals dwarfing others. This was addressed by signal intensity normalization using a sigmoid function to reduce the range of intensity difference. The parameter for the inflection point of the sigmoid function was user-customizable. The fourth consideration is about the overlaps between peak projections in the carbon dimension on the iCov spectrum due to cross-shaped ridges. This issue was addressed by peak digitization based on the peak position information from the

one-dimensional (1D) carbon spectrum of the mixture. Usually, one can simply use the diagonal slice of the NUS-mHMBC-iCov which is essentially the 1D ^{13}C spectrum of the mixture with reasonable resolution. For compounds with high carbon overlaps, as in the rotenone and brucine mixture, one can obtain the actual 1D ^{13}C spectrum of the mixture with higher resolution. Even for this case, the diagonal slice of the NUS-mHMBC-iCov gave well-separated individual spectra except for a total of 4 unresolved peaks (data not shown) which could be resolved by the use of the actual 1D spectrum of the mixture. It should be noted that the spectra of the individual compounds are never needed. From the 1D spectrum, one can tabulate every peak position by peak-picking and use them to represent all the iCov spectral peaks with three adjacent points centered on the peak position. The spectrum will have positive values at the picked position and the positions offset by one point from the peak-picked position, and all other regions will have zero values. This takes advantage of the ridge-free sharp lines of ^{13}C 1D spectra, and effectively removes spectral ridges on the resulting 2D spectrum. This digitized spectrum is then used as the input for the eigendecomposition for spectral deconvolution at the final step of the DECODE procedure.

Mathematical implementation of the DECODE procedure

1. Spectral merging

The spectral merging step constructs an n -by- m matrix, \mathbf{H} , which takes the maximum magnitude from the individual spectrum.

$$H_{ij} = \max(S^1_{ij}, S^2_{ij}, S^3_{ij} \cdots S^k_{ij}), \text{ where, } \mathbf{S}^k \text{ is an } n\text{-by-}m \text{ matrix which represents the individual spectrum.}$$

Here, $\max(x, y, \dots, z)$ returns a maximum magnitude value among variables.

2. Calculation of the first moment for the spectral moment filtering

The first moment matrix $\mathbf{\mu}$ was calculated as described previously.¹ Here, we used the power spectrum \mathbf{P} instead of the original spectral matrix \mathbf{H} to construct the first moment matrix ($P_{ij} = H_{ij}^2$).

$$\mu_{ij} = \sum_{k=-M}^M (j+k) P_{i,j+k} / \sum_{k=-M}^M P_{i,j+k}$$

where $M = 7$ (15 pts ($= 2M+1$) in 4096 pts correspond to ca. 37 Hz for 12 ppm spectral width at 850 MHz). This first moment was used during the covariance calculation as an argument for the filter function

3. Indirect covariance calculation

The indirect covariance matrix, \mathbf{C} (n -by- n), was obtained as below

$$W_{ij} = \sum_{k=1}^n \text{abs}(H_{ik}) \cdot \text{abs}(H_{jk}) \cdot f(\mu_{ik} - \mu_{jk})$$

$$\mathbf{C} = \mathbf{W}^{1/2}$$

where \mathbf{H} is the merged spectral matrix, $\mathbf{\mu}$ is the first spectral moment obtained above, and n is the number of columns in \mathbf{H} . $\text{Abs}(\mathbf{H})$ returns the absolute values of the matrix \mathbf{H} . If the J -modulation based

overlap filter is to be used with the pulse sequence S1B, the spectral derivative of **H**, instead of the absolute value matrix should be used as below.

$$W_{ij} = \sum_{k=1}^n D_{ik} \cdot D_{jk} \cdot f(\mu_{ik} - \mu_{jk})$$

where **D** is the spectral derivative of **H** along the row (F_2 domain). f is a filter function defined as

$$f(x) = \frac{1}{1 + \exp\{-10(1.4 - x)\}}$$

4. Dataset normalization

The elements of the covariance matrix were normalized with a sigmoid function as below.

$$L_{ij} = \frac{1}{1 + \exp\{-30(\log C_{ij}/\log C_{max}) + \delta\}}$$

where C_{max} and δ are the maximum value of the matrix **C** and the inflection point of the sigmoid function (user-customizable), respectively. For the logarithm calculation, if $C_{ij} < 1$ then, $C_{ij} = 1$.

5. Peak-digitization

Step 1.

Define an array **P** with positional indices of valid peaks* over a user defined threshold in ^{13}C spectrum.

Next, define a binary array **B** and its diagonal matrix **D** (=diag(**B**)) such that

for all x , the positional indices of the ^{13}C spectrum

$$B[x] = \begin{cases} 1, & \text{if } x \vee x \pm 1 \in \text{array } \mathbf{P} \\ 0, & \text{otherwise} \end{cases}$$

$$\mathbf{D} = \text{diag}(\mathbf{B})$$

Step 2.

for all k , the indices of non-zero elements of \mathbf{B} ,

Define an array \mathbf{M} with the positional indices of valid peaks* over 0.01 threshold in L_k (L_k is the k^{th} 1D slice (row/column) of the matrix \mathbf{L})

for all i , index of 1D spectrum array of L_k

$$L_k[i] = \begin{cases} L_k[i], & \text{if } i \in \text{array } \mathbf{M} \\ 0, & \text{otherwise} \end{cases}$$

Step 3.

Constructing the ridge-free matrix \mathbf{F} by following matrix operation

$$\mathbf{F} = \mathbf{D}\mathbf{L}\mathbf{D}$$

* Throughout the pick-digitization process a function 'scipy.signal.find_peak' from the SciPy library from python was used for the peak-detection.

6. Extracting eigenmodes

An eigendecomposition of the matrix \mathbf{F} ($\mathbf{F} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$) yields eigenmodes. Here, each eigenmode is V_k , where V_k is the column vector (the eigenvector of the matrix \mathbf{F}) of \mathbf{V} . The number of the desired eigenvectors (1D spectra) can be specified. The eigenvectors are calculated in descending order of the associated eigenvalues.

Figures

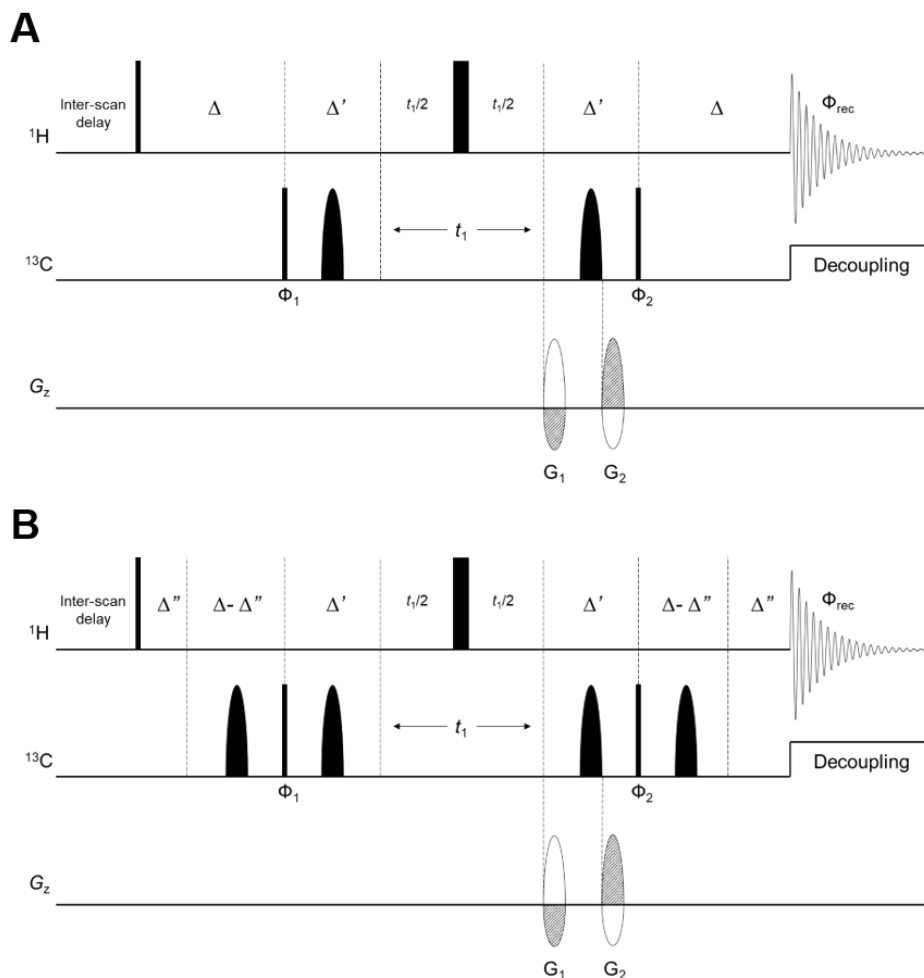


Figure S1. Pulse sequences for mHMBC for multiple bond correlation (A) and one bond correlation experiment (B) that has the same J_{HH} modulation as mHMBC. For rotenone, (A) with Δ of 62.5 ms ($^nJ_{\text{CH}} = 8$ Hz) was used. For rotenone and brucine mixture, (A) with Δ of 62.5 ms ($^nJ_{\text{CH}} = 8$ Hz) and Δ of 125 ms ($^nJ_{\text{CH}} = 4$ Hz) were used. (B) was used for sucrose/quinic acid mixture for J -modulation filter. In (B), the delay Δ'' was set to 3.45 ms for HSQC-type one-bond correlation. For both of (A) and (B), the semi-ellipsoid boxes represent 180° shaped pulses for inversion or refocusing of ^{13}C magnetization. Narrow and wide bars represent 90° and 180° hard pulses, respectively. For the phase-sensitive acquisition in the indirect domain, an echo-antiecho detection mode was employed. Phases of Φ_1 and Φ_{rec} were shifted by 180° at every domain-increment. Blank semi-ellipse boxes in the gradient channel represent gradient pulses for odd-numbered increments and gray semi-ellipse boxes for even-numbered increments. All pulses are of phase, x unless otherwise indicated. The phase cycling is as follows. $\Phi_1 = x, -x$; $\Phi_2 = -x, -x, x, x$; $\Phi_{\text{rec}} = -x, x, x, -x$. Gradient ratios: $G_1 : G_2 = 5 : -3$ (odd-numbered) and $-3 : 5$ (even-numbered).

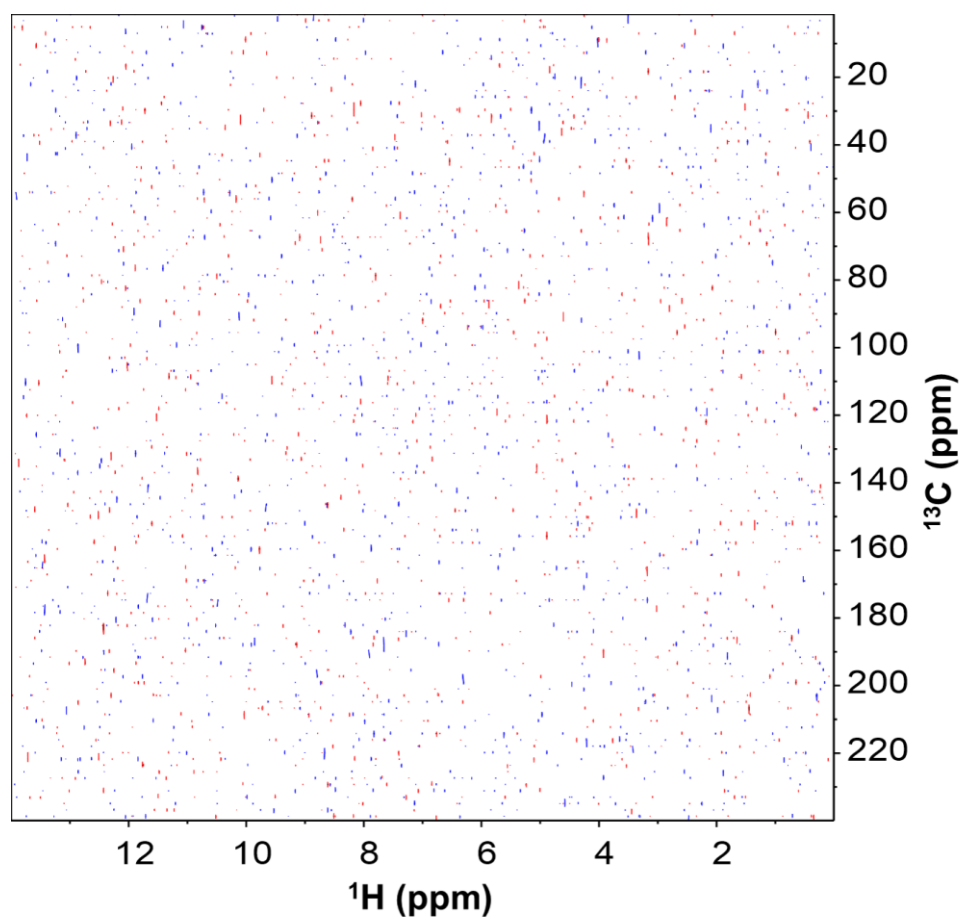


Figure S2. An ^1H , n -ADEQUATE spectrum of rotenone (2 mg in 600 μL of CDCl_3)

Reference

- (1) Bingol, K.; Salinas, R. K.; Bruschweiler, R. Higher-Rank Correlation NMR Spectra with Spectral Moment Filtering. *J. Phys. Chem. Lett.* **2010**, 1 (7), 1086.