

# Extended Data - Automated analysis for multiplet identification from ultra-high resolution 2D- $^{13}\text{C}$ , $^1\text{H}$ -HSQC NMR spectra.

Laura Ferrante,<sup>†</sup> Kashif Rajpoot,<sup>†,§</sup> Mark Jeeves,<sup>‡,||</sup> and Christian Ludwig\*,<sup>¶,||</sup>

<sup>†</sup>*School of Computer Sciences, University of Birmingham, Birmingham, B15 2TT, UK*

<sup>‡</sup>*Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, B15 2TT, UK*

<sup>¶</sup>*Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, B15 2TT, UK*

<sup>§</sup>*University of Birmingham Dubai Dubai International Academic City, Dubai, UAE*

<sup>||</sup>*Authors contributed equally*

E-mail: C.Ludwig@bham.ac.uk

Below we detail the fundamental steps in the software pipeline: (1) the ICA problem formulation (2) the hypotheses underpinning the algorithm choice for its application to 2D- $^{13}\text{C}$ ,  $^1\text{H}$ -HSQC NMR spectra.

## Independent Component Analysis for mutiplets identification

The problem of identifying the metabolites multiplets in spectra is in this work considered a signal unmixing problem and it is solved using independent component analysis (ICA).

Within signal unmixing methods, ICA aims to explain  $y_1(\omega), y_2(\omega), \dots, y_N(\omega)$  observations through a smaller number of so called latent or hidden variables  $s_1(\omega), s_2(\omega), \dots, s_N(\omega)$ , assuming that each observation  $y_i$  is a mixture of latent variables weighted by unknown coefficients  $w_i$  as follows:

$$y_i = w_{i1}s_1 + w_{i2}s_2 + \dots + w_{iN}s_N \quad (1)$$

Using a vector-matrix notation and dropping the frequency dependence without loss of generality, the equation can be rewritten to describe all the observed signals:

$$\mathbf{y} = \mathbf{W}\mathbf{s} \quad (2)$$

The generative model described in 2 is called the ICA model and the latent variables are called independent components. Readers are referred to<sup>1</sup> for more details on the algorithm, while below we explain how the ICA is used to solve the multiplet identification problem. In the case of a 2D-NMR spectrum of a metabolite mixture, each multiplet of each metabolite signal is a latent variable and the 2D-HSQC spectrum is the observation space. Therefore in our case only one observation (i.e. a 2D spectrum) is available. We address this shortcoming by utilising the proton dimension as the observation dimension. Consider the Cartesian reference system where the axis x, y, z, are the proton, the carbon and the intensity of the peaks dimensions, respectively. Consider the surface  $\Omega \subset \mathbb{R}_3$  (Figure S1), for a fixed proton shift  $1H0$  the projection on the y-z ( $^{13}\text{C}$ -Intensity) plane is defined as  $A_{1H0} = (y, z) \in \mathbb{R}^2 : (x, y, z) \in \Omega$  and is given by the intersection of  $\Omega$  with the plane  $x = 1H0$ . An example concerning glutamate carbon 2 is represented in Figure S1. In discrete space, it can be demonstrated that the union along the  $^1\text{H}$ -dimension of all the 1D  $^{13}\text{C}$ -NMR spectra gives the obtained 2D spectrum  $\Omega$ . All the  $^{13}\text{C}$  NMR spectra at different proton chemical shifts are considered as observations, therefore the proton dimension is the observation dimension, and the problem can be solved using 1D-ICA. The signal unmixing problem is therefore formulated as in Eq.2, where  $\mathbf{y} \in \mathbb{R}^{M \times N}$  are all the observations which constitute the 2D spectrum (e.g. the carbon spectra slices at each proton resonance shift),  $\mathbf{W}$

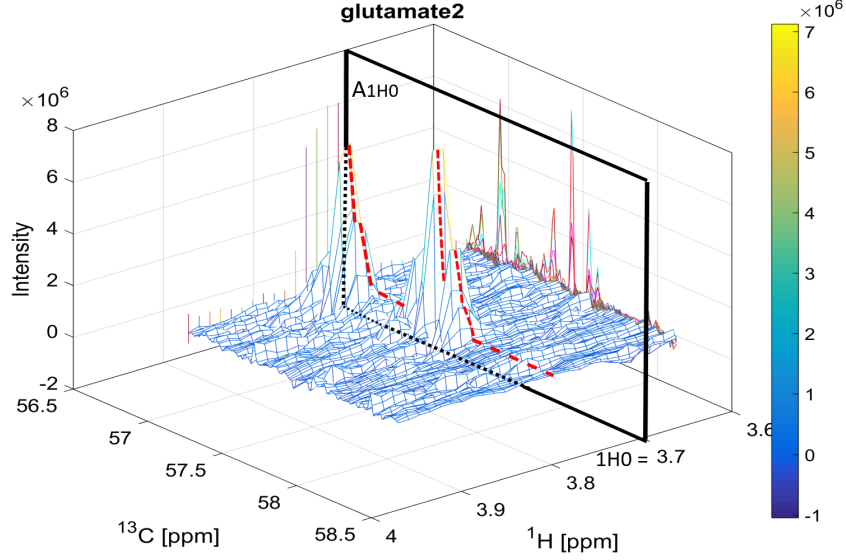


Figure S1: The spectral area of interest  $\Omega$  considering the theoretical resonance of glutamate carbon 2 is shown. On the plane  $^{13}\text{C}$ -Intensity, the 1D  $^{13}\text{C}$ -NMR spectra obtained by intersection of the spectral surface  $\Omega$  and the plane A1H0 are shown, where A1H0 is the plane sliding along the proton dimension (1H0 are the different proton shifts). On the plane  $^1\text{H}$ -Intensity the projection obtained by intersection of the spectral surface  $\Omega$  and the plane  $x = 1\text{H0}$  are shown.

is the unknown linear mapping matrix from the observation to the latent space and  $\mathbf{s}$  are the underlying spectral components (e.g. the mutiplets of the metabolites). The  $K$  components are assumed to be statistically independent as required for the ICA algorithm.<sup>1</sup> This assumption is also physically/chemically plausible, as there is no interaction between metabolites at different  $^1H$ -shifts.<sup>2</sup> Moreover, the linearity of the model can be justified, as the metabolites mix in an additive way in the spectrum.<sup>2</sup> These assumptions justify the choice of using ICA among other unmixing algorithms such as non-negative matrix decomposition (NMF), since ICA finds a decomposition of the observed data to retrieve latent components which are as independent as possible. ICA is performed using FastICA routine<sup>3</sup> with non-linearity given in the equation below 3, which, in comparison to other non-linearities commonly used, gave the best outcome (e.g. lowest rate of mismatching) during the "matching step" explained in the main paper.<sup>4</sup> Given that  $\mathbf{s}$  is a independent variable, the non linearity function  $g(\mathbf{s})$  can be written as below:

$$g(\mathbf{s}) = \mathbf{s}^2 \quad (3)$$

To recover the independent components, FastICA algorithm uses an optimisation approach based on the Newton method and faster deflationary approach.<sup>3</sup> The algorithm used here has a cubic convergence and the independent components are found in up to 12 iterations. ICA is run multiple times, considering different random initialisations of the weight matrix  $\mathbf{W}$ , in order to ensure stability of the independent components, and evaluate the stability of the algorithm to converge to the same solution. The ICA algorithm capability to produce independent signals was tested by computing the mutual information (MI).<sup>5</sup> Preliminary experiments showed that the values are  $< 0.1$ , which is in accordance to our expectations indicating that the components are independent. These experiments were carried out during the initial framework development to validate the choice of hyperparameters and non-linear function. We think that the underlying white noise as well as residual noise originating from non-uniform sampling schedules increase the MI values. While this leads to noise regions being identified as potentially valid independent components as demonstrated in IC 3 of Figure 3 in the main paper, correlations will be quite small and therefore this will not have any significant influence in practice.

Table S1: List of variables names and correspondent description.

Variable Name	Description
$\tilde{X}$	2D spectrum
$\tilde{X}$	restricted 2D spectrum
mName	Metabolite name
mSpin	Metabolite Spin number
maxWidth1H	Upper limit [ppm] to restrict the searching area along the proton dimension
maxWidth13C	Upper limit [ppm] to restrict the searching area along the carbon dimension
nReps	Number of ICA runs
K	Number of independent components
Y	Matrix of Independent components
$x_H$	Vector containing the proton location in the spectrum of each ICA component
$x_C$	Vector containing the carbon chemical shift of each multiplet in the ICA components
corrScore	Vector containing the correlation between each independent component and the corresponding 1D-H spectrum
$\Gamma_{adjusted}$	relative distance weighting: $20 \times \gamma_H / \gamma_C$
$\rho$	Coefficient of determination
s	simulated multiplet components

## Pseudocode

```

input : X, mName, mSpin, maxWidth1H, maxWidth13C, nReps, K
output: Data structure
 $\tilde{X} \leftarrow \text{setSearchArea}(X, mName, mSpin, maxWidth1H, maxWidth13C);$ 
for  $i = 1 : nReps$  do
     $Y \leftarrow \text{fastICA}(\tilde{X}, K);$ 
    for  $j = 1 : K$  do
         $x_{H_j}, \text{corrScore}_j \leftarrow \text{getHshift}(X, Y_{j,:});$ 
    end
    store all  $Y, x_H, \text{corrScore};$ 
end
 $\bar{Y} \leftarrow$  get the set of  $Y$  which contains the component with highest correlation;
 $\overline{x_H}, \overline{\text{corrScore}} \leftarrow$  get  $x_{H_j}$  and  $\text{corrScore}$  for each latent component in  $\bar{Y}$ ;
 $s \leftarrow \text{simMultiplets}(mName, mSpin);$ 
for each latent component in  $\bar{Y}$  do
    if  $\overline{\text{corrScore}_K} > 0.5$  then
         $s_{\text{norm}} \leftarrow$  normalize  $s$  to the range  $[\min(\bar{Y}_{k,:}) \max(\bar{Y}_{k,:})];$ 
         $s_{\text{aligned}} \leftarrow$  align  $s_{\text{norm}}$  to  $\bar{Y}_{k,:}$ ;
         $Y_{\text{estim}_K,:}, \rho_K \leftarrow \text{linRegress}(s_{\text{aligned}}, X_{\overline{x_H},:});$ 
         $\overline{x_C} \leftarrow \text{getCShift}(Y_{\text{estim}_K,:});$ 
         $\Delta_{\text{ppm}} \leftarrow \sqrt{(\overline{x_H} - \text{HShiftLib})^2 + (\frac{\overline{x_C} - \text{CShiftLib}}{\Gamma_{\text{adjusted}}})^2};$ 
         $\rho_{\text{adjusted}_K} \leftarrow \frac{\rho^{(k)}}{\Delta_{\text{ppm}}^2};$ 
    else
        go back to the beginning of current section;
    end
end
 $Y_{\text{final}}, x_{H_{\text{final}}} \leftarrow$  choose  $Y_{\text{estim}}$  with highest  $\rho_{\text{adjusted}}$ ;
 $x_{H_{\text{final}}} \leftarrow \text{hillClimbing}(Y_{\text{final}}, x_{H_{\text{final}}});$ 
 $s_{\text{aligned}} \leftarrow$  find best alignment to  $Y_{\text{final}}$  for each component of  $s$ ;
 $Y_{\text{final}}, \rho_{\text{final}} \leftarrow \text{linRegress}(s_{\text{aligned}}, X_{x_{H_{\text{final}}},:});$ 

```

**Algorithm 1:** Overview of the algorithm

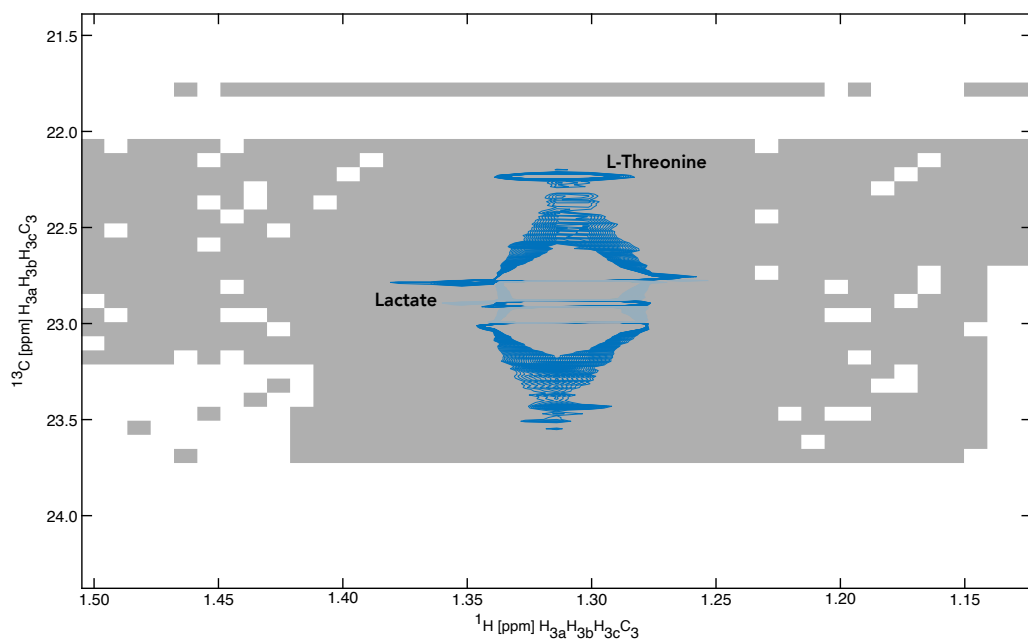


Figure S2: Region of the spectrum showing the multiplet peak for lactate C(3). The gray area underlying the spectrum indicates the highest coefficient of determination computed for lactate C(3) as a function of the chemical shift offset.

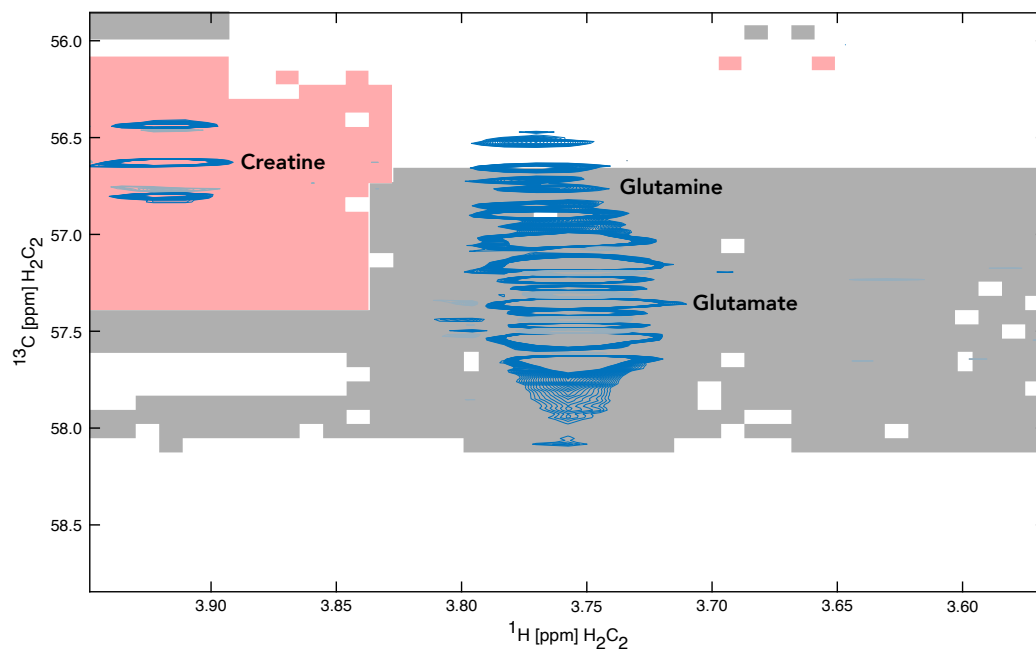


Figure S3: Region of the spectrum showing the multiplet peak for glutamate C(2). The gray area underlying the spectrum indicates the highest coefficient of determination computed for glutamate C(2) as a function of the chemical shift offset.

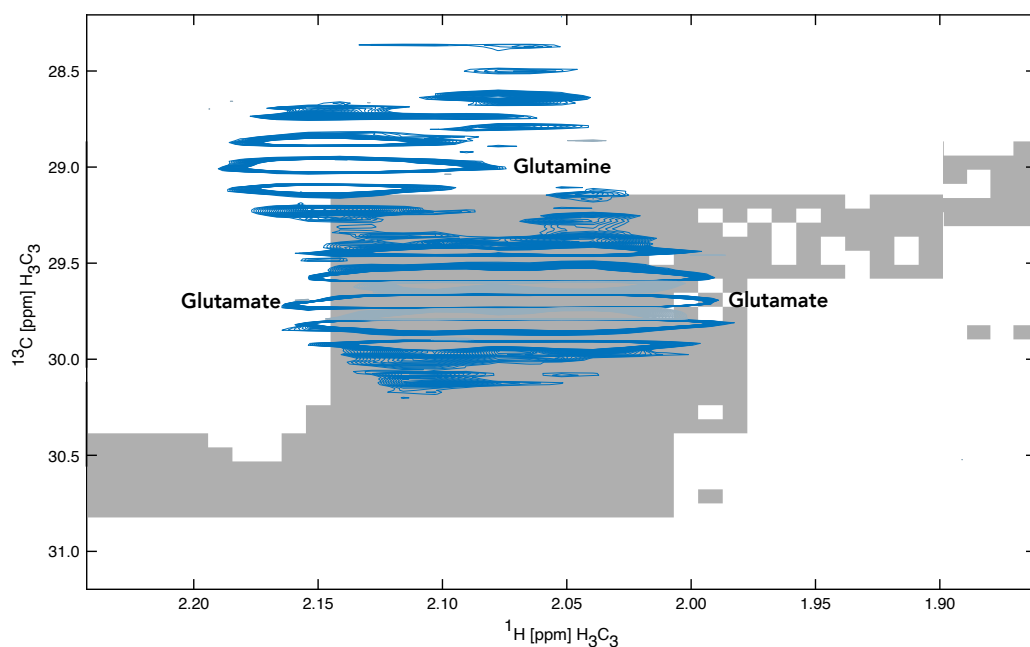


Figure S4: Region of the spectrum showing the multiplet peak for glutamate C(3). The gray area underlying the spectrum indicates the highest coefficient of determination computed for glutamate C(3) as a function of the chemical shift offset.

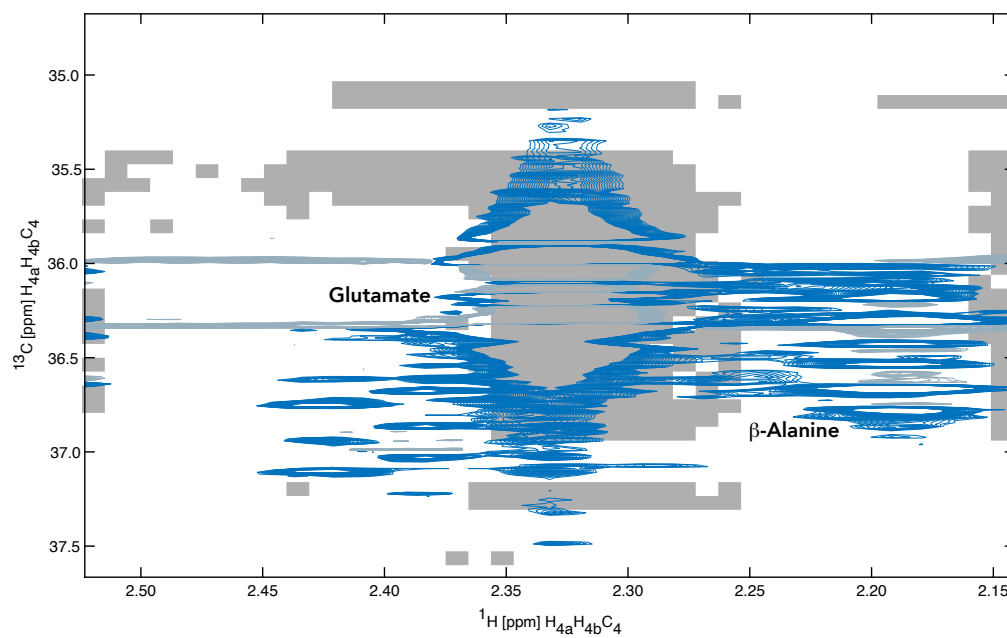


Figure S5: Region of the spectrum showing the multiplet peak for glutamate C(4). The gray area underlying the spectrum indicates the highest coefficient of determination computed for glutamate C(4) as a function of the chemical shift offset.

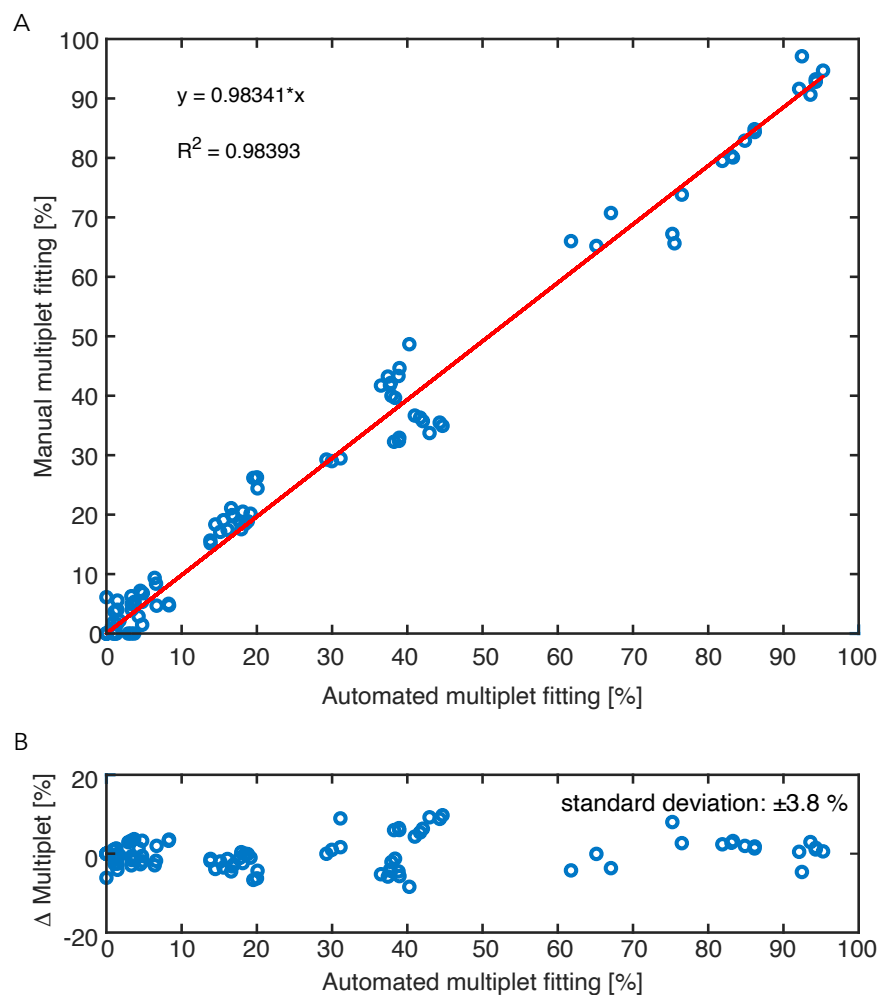


Figure S6: Manual vs automated analysis: (A) Multiplet percentages obtained with the automated algorithm are plotted on the horizontal axis, whereas previously published values<sup>6</sup> from a manual multiplet analysis are plotted on the vertical axis. (B) The difference between the fitted red line in (A) and the experimental values are plotted on the vertical axis.



Data Set	Exp.	Tracer	Sampling Rate [%]	J Enhancement	$^1\text{H}$ Freq [MHz]
MTBSL182	1-3	[1,2- $^{13}\text{C}$ ] Glucose	25	1	600
eqhn3		[1,2- $^{13}\text{C}$ ] Glucose & [U- $^{13}\text{C}$ , U- $^{15}\text{N}$ ] Gln			
	1		25	1	600
	2		25	2	600
	3		25	4	600
	4		25	8	600
qtmge		[U- $^{13}\text{C}$ ] Glucose			
	1		50	1	800
	2		25	1	800
	3		30	1	800
	4		20	1	800
	5		10	1	800
	6		5	1	800

Table S2: Dataset characteristics

Metabolite MTBLS241	Carbon #	$^1\text{H}$	$^{13}\text{C}$	$\rho_{final}$ [%]	Contribution [%]			
		[ppm]	[ppm]		1	2	3	4
alanine	2	3.821	52.965	98.8	6.4	0	93.6	0
alanine	3	1.467	18.973	89.2	17	83	-	-
aspartate	2	3.923	54.908	90.3	36.6	44.4	17.6	1.5
aspartate	3	2.765	39.308	99	37.8	17.8	41.1	3.3
aspartate	3	2.765	39.308	99	37.8	17.8	41.1	3.3
glutamate	2	3.796	57.36	82	38.8	38.1	14.5	8.5
glutamate	3	2.039	29.709	98.5	61.3	0	30	8.7
glutamate	3	2.039	29.709	98.5	61.3	0	30	8.7
glutamate	4	2.332	35.653	97.2	19.3	1.7	75.1	3.9
lactate	2	4.139	71.24	97.9	4.5	1.2	94.3	0
lactate	3	1.314	22.897	84.7	13.9	86.1	-	-
alanine	2	3.821	52.965	98.2	6.6	1.3	92.1	0
alanine	3	1.467	18.973	91.5	18.2	81.8	-	-
aspartate	2	3.923	54.908	94.9	37.9	42.8	19.3	0
aspartate	3	2.765	39.308	92.5	38.5	18.3	41.7	1.5
aspartate	3	2.765	39.308	92.5	38.5	18.3	41.7	1.5
glutamate	2	3.796	57.36	83.9	38.7	38.8	15.6	6.9
glutamate	3	2.039	29.709	97.4	65	0	31.2	3.8
glutamate	3	2.039	29.709	97.4	65	0	31.2	3.8
glutamate	4	2.332	35.653	93.4	19.8	0	76.6	3.6
lactate	2	4.139	71.24	95.9	4.8	0.8	94.3	0
lactate	3	1.314	22.897	95.9	15.1	84.9	-	-
alanine	2	3.834	52.965	98.7	4.7	0	95.3	0
alanine	3	1.467	18.973	93.3	15.9	84.1	-	-
aspartate	2	3.948	54.908	93.2	37.9	44.8	16.4	0.9
aspartate	3	2.765	39.308	95.1	37.7	18.4	42.2	1.7
aspartate	3	2.765	39.308	95.1	37.7	18.4	42.2	1.7
glutamate	2	3.808	57.36	80.6	40.2	38.9	16.1	4.8
glutamate	3	2.052	29.709	97.1	67.1	0	29.5	3.4
glutamate	3	2.052	29.709	97.1	67.1	0	29.5	3.4
glutamate	4	2.332	35.653	90.5	18.7	1.4	75.7	4.2
lactate	2	4.152	71.24	97.5	4.3	3.3	92.4	0
lactate	3	1.314	22.897	97.1	13.9	86.1	-	-

Table S3: Dataset MTBLS241

Metabolite eqhn3	Carbon #	$^1\text{H}$	$^{13}\text{C}$	$\rho_{final}$ [%]	Contribution [%]			
		[ppm]	[ppm]		1	2	3	4
alanine	2	3.783	52.965	97.9	2.2	2.6	70.9	24.3
alanine	3	1.467	18.973	99.1	11.8	88.2	-	-
aspartate	2	3.885	54.908	93.2	25.3	46.4	13.6	14.7
aspartate	3	2.778	39.308	98.3	26.6	15.1	43.5	14.8
aspartate	3	2.778	39.308	98.3	26.6	15.1	43.5	14.8
glutamate	2	3.758	57.36	88.2	25.5	42.1	15.7	16.7
glutamate	3	2.103	29.709	97.2	45.7	0	41.6	12.6
glutamate	3	2.103	29.709	97.2	45.7	0	41.6	12.6
glutamate	4	2.332	35.653	89.6	14.5	2.3	70.3	12.8
lactate	2	4.101	71.24	99.5	3.2	3	76.4	17.4
lactate	3	1.314	22.897	99	12	88	-	-
alanine	2	3.783	52.965	97.3	2	2.2	73.2	22.6
alanine	3	1.467	18.973	99.1	11.8	88.2	-	-
aspartate	2	3.885	54.908	96.1	27.4	44.9	12.3	15.5
aspartate	3	2.778	39.308	99.1	26.3	15.4	43.4	14.9
aspartate	3	2.778	39.308	99.1	26.3	15.4	43.4	14.9
glutamate	2	3.758	57.36	94.1	26.3	41.3	15.1	17.3
glutamate	3	2.103	29.709	98.7	44.7	41.6	0	13.7
glutamate	3	2.103	29.709	98.7	44.7	41.6	0	13.7
glutamate	4	2.332	35.653	96.8	14.3	1.5	71.4	12.8
lactate	2	4.101	71.24	98.6	2.8	2.6	77.9	16.7
lactate	3	1.314	22.897	99.5	12.1	87.9	-	-
alanine	2	3.783	52.965	86.6	2.6	3.5	75.2	18.8
alanine	3	1.467	18.973	98.9	11.7	88.3	-	-
aspartate	2	3.898	54.908	95.2	31.5	44.8	12.1	11.6
aspartate	3	2.778	39.308	97.6	26.6	14.9	43.7	14.7
aspartate	3	2.778	39.308	97.6	26.6	14.9	43.7	14.7
glutamate	2	3.758	57.36	94.6	26.6	40.9	15.9	16.6
glutamate	3	2.052	29.709	98.6	47.1	0	42.5	10.4
glutamate	3	2.103	29.709	98.7	48.7	0	42.1	9.1
glutamate	4	2.332	35.653	98.3	14	1.7	71.4	12.8
lactate	2	4.101	71.24	96.3	2.1	3.4	79.7	14.8
lactate	3	1.314	22.897	99.5	12	88	-	-
alanine	2	3.783	52.965	56.7	5.2	9.7	64.5	20.6
alanine	3	1.467	18.973	98.9	11.4	88.6	-	-
aspartate	2	3.885	54.908	92.4	32.7	45.5	11.2	10.6
aspartate	3	2.778	39.308	96.3	27.9	13.8	46	12.3
aspartate	3	2.778	39.308	96.3	27.9	13.8	46	12.3
glutamate	2	3.758	57.36	88.1	29.6	43.9	13.3	13.3
glutamate	3	2.103	29.709	98.3	51.7	42.2	0	6
glutamate	3	2.103	29.709	98.3	51.7	42.2	0	6
glutamate	4	2.332	35.653	98.3	14.5	1.6	72	12
lactate	2	4.101	71.24	81.1	3	4.1	77.1	15.8
lactate	3	1.314	22.897	99.5	11.8	88.2	-	-

Table S4: dataset eqhn3

Name	C#	<sup>1</sup> H	<sup>13</sup> C	$\rho_{final}$ [%]	Contribution [%]			
		[ppm]	[ppm]		1	2	3	4
qtmge								
alanine	2	3.805	52.965	96.9	2.6	0	0	97.4
alanine	3	1.482	18.973	99.3	2.2	97.8	-	-
aspartate	2	3.912	54.908	55	34.1	40.9	14.1	10.9
aspartate	3	2.735	39.308	70.5	38.9	6.4	46.1	8.5
aspartate	3	2.735	39.308	70.5	38.9	6.4	46.1	8.5
glutamate	2	3.912	57.36	47.8	100	0	0	0
glutamate	3	2.155	29.709	93.2	81.4	3.6	0	15
glutamate	3	2.048	29.709	90.1	72.4	27.6	0	0
glutamate	4	2.338	35.653	94.8	13.2	7.9	70	8.9
lactate	2	4.126	71.24	99.4	1.9	1.5	2.4	94.2
lactate	3	1.314	22.897	89.8	0.7	99.3	-	-
alanine	2	3.805	52.965	95.9	5.3	1.9	3.7	89.2
alanine	3	1.467	18.973	96	1.4	98.6	-	-
aspartate	2	3.912	54.908	62.6	32.5	49.8	7.8	9.9
aspartate	3	2.781	39.308	59.2	46.4	8.3	28.5	16.8
aspartate	3	2.781	39.308	59.2	46.4	8.3	28.5	16.8
glutamate	2	3.79	57.36	94.2	80.7	3.5	3.8	12.1
glutamate	3	2.093	29.709	95.1	69.7	30.3	0	0
glutamate	3	2.093	29.709	95.1	69.7	30.3	0	0
glutamate	4	2.323	35.653	87.7	14.5	9.1	68.5	7.9
lactate	2	4.111	71.24	99.3	2.3	1.6	2.7	93.5
lactate	3	1.314	22.897	94.9	2.1	97.9	-	-
alanine	2	3.805	52.965	90.1	4.8	0	0	95.2
alanine	3	1.467	18.973	99.4	1.4	98.6	-	-
aspartate	2	3.729	54.908	22.5	75.7	5.6	0	18.7
aspartate	3	2.72	39.308	68.6	44.7	0	55.3	0
aspartate	3	2.72	39.308	68.6	44.7	0	55.3	0
glutamate	2	3.759	57.36	20.2	51	26.3	0	22.7
glutamate	3	2.155	29.709	39.5	100	0	0	0
glutamate	3	2.032	29.709	94.1	80.6	19.4	0	0
glutamate	4	2.323	35.653	86	12.9	8.7	65.6	12.8
lactate	2	4.111	71.24	98.7	2	1.3	2.1	94.6
lactate	3	1.314	22.897	98.3	2.6	97.4	-	-
alanine	2	3.79	52.965	80.8	0.3	0	0	99.7
alanine	3	1.467	18.973	97.6	1.6	98.4	-	-
aspartate	2	3.912	54.908	63.8	93.2	3.3	3.5	0
aspartate	3	2.781	39.308	40.6	47.8	3.3	15.4	33.5
aspartate	3	2.781	39.308	40.6	47.8	3.3	15.4	33.5
glutamate	2	3.79	57.36	90	82.6	0	0	17.4
glutamate	3	2.155	29.709	88.9	65.6	0	4.2	30.2
glutamate	3	2.032	29.709	91.8	71.9	28.1	0	0
glutamate	4	2.323	35.653	89.7	11.1	6.5	73.6	8.7
lactate	2	4.111	71.24	97.9	2	1.1	2.1	94.9
lactate	3	1.314	22.897	98.3	0.4	99.6	-	-
alanine	2	3.805	52.965	0	0	0	36.6	63.4
alanine	3	1.467	18.973	98.8	1.9	98.1	-	-

aspartate	2	3.943	54.908	63.8	93.2	3.3	3.5	0
aspartate	3	2.72	39.308	60.1	42.1	17.3	29.6	10.9
aspartate	3	2.72	39.308	60.1	42.1	17.3	29.6	10.9
glutamate	2	3.79	57.36	35.5	100	0	0	0
glutamate	3	2.093	29.709	87.4	67.8	32.2	0	0
glutamate	3	2.093	29.709	87.4	67.8	32.2	0	0
glutamate	4	2.323	35.653	82.9	14.7	11.8	73.5	0
lactate	2	4.111	71.24	98.2	1.7	0.9	0	97.4
lactate	3	1.314	22.897	99.3	1.2	98.8	-	-
alanine	2	3.805	52.965	59.9	88.9	3.4	3.3	4.4
alanine	3	1.467	18.973	92.9	2.4	97.6	-	-
aspartate	2	3.729	54.908	63.8	93.2	3.3	3.5	0
aspartate	3	2.781	39.308	60.9	43.7	39.2	17.1	0
aspartate	3	2.781	39.308	60.9	43.7	39.2	17.1	0
glutamate	2	3.805	57.36	33.4	61.3	4.5	34.2	0
glutamate	3	2.109	29.709	83.2	69.6	4.9	0	25.5
glutamate	3	2.063	29.709	91.6	69.1	4.2	0	26.7
glutamate	4	2.338	35.653	91.6	14.1	7.6	74.4	3.9
lactate	2	4.126	71.24	99.2	2	0.4	3.3	94.3
lactate	3	1.314	22.897	84.2	1.2	98.8	-	-

Table S5: Dataset qtmge

## MetaboLab Script - MTBLS241.ml

```
% Processing script for MetaboLab
% Comments start with a percentage sign
% Everything between START and END MLScript is executed
% everything outside is ignored
% The script starts in the next line
START MLScript

autoHsqcMA
    metabolites:      alanine, aspartate, glutamate, glutamine, lactate
    maxWidth1H:       0.10 % [ppm]
    maxWidth13C:      1.00 % [ppm]
    minCorr:          0.50 %
    rangeC:           0.80 % [ppm]
    maxRange:         1    % [points]
    nReps:            25    %
    R2:               1.5   % [Hz]
    echoTime:         1.85 % [ms]
    dataSets:         all   %
    experiments:      all   %
    report:           on    %
    outputDir:        /Volumes/Home/ludwigc/Desktop/autoHsqcMAREport
    outputName:       mlReport_MTBLS241
endAutoHsqcMA

% End of script
END MLScript
```

## MetaboLab Script - qtmge.ml

```
% Processing script for MetaboLab
% Comments start with a percentage sign
% Everything between START and END MLScript is executed
% everything outside is ignored
% The script starts in the next line
START MLScript

autoHsqcMA
    metabolites:      alanine, aspartate, glutamate, lactate
    maxWidth1H:       0.10 % [ppm]
    maxWidth13C:      1.00 % [ppm]
    minCorr:          0.50 %
    rangeC:           0.80 % [ppm]
    maxRange:         1    % [points]
    nReps:            25    %
    R2:               1.5   % [Hz]
    echoTime:         1.95 % [ms]
    dataSets:         all   %
    experiments:      all   %
    report:           on    %
    outputDir:        /Volumes/Home/ludwigc/Desktop/autoHsqcMAREport
    outputName:       mlReport_qtmge
endAutoHsqcMA

% End of script
END MLScript
```

## MetaboLab Script - EQHN3.ml

```
% Processing script for MetaboLab
% Comments start with a percentage sign
% Everything between START and END MLScript is executed
% everything outside is ignored
% The script starts in the next line
START MLScript

autoHsqcMA
    metabolites:      alanine, aspartate, glutamate, lactate
    maxWidth1H:       0.10 % [ppm]
    maxWidth13C:      1.00 % [ppm]
    minCorr:          0.50 %
    rangeC:           0.80 % [ppm]
    maxRange:         1    % [points]
    nReps:            25    %
    R2:               1.5   % [Hz]
    echoTime:         1.85 % [ms]
    dataSets:         all   %
    experiments:      all   %
    report:           on    %
    outputDir:        /Volumes/Home/ludwigc/Desktop/autoHsqcMAREport
    outputName:       mlReport_eqhn3
endAutoHsqcMA

% End of script
END MLScript
```



## References

- (1) Comon, P. Independent component analysis, A new concept? *Signal Processing* **1994**, *36*, 287 – 314, Higher Order Statistics.
- (2) Levitt, M. H. *Spin dynamics: basics of nuclear magnetic resonance*; John Wiley & Sons, 2001.
- (3) Hyvarinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE transactions on Neural Networks* **1999**, *10*, 626–634.
- (4) Hyvarinen, A. One-unit contrast functions for independent component analysis: A statistical analysis. Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop. 1997; pp 388–397.
- (5) Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Physical review E* **2004**, *69*, 066138.
- (6) Chong, M.; Jayaraman, A.; Marin, S.; Selivanov, V.; de Atauri Carulla, P. R.; Tennant, D. A.; Cascante, M.; Günther, U. L.; Ludwig, C. Combined analysis of NMR and MS spectra (CANMS). *Angewandte Chemie International Edition* **2017**, *56*, 4140–4144.