Evaluation of Fast 2D NMR for Metabolomics

ARTICLE *in* ANALYTICAL CHEMISTRY · MAY 2014

Impact Factor: 5.64 · DOI: 10.1021/ac500966e · Source: PubMed

3 AUTHORS:

Adrien Le Guennec
University of Georgia
6 PUBLICATIONS **44** CITATIONS
SEE PROFILE

Patrick Giraudeau
University of Nantes
60 PUBLICATIONS **803** CITATIONS
SEE PROFILE

Stefano Caldarelli
Aix-Marseille Université
115 PUBLICATIONS **1,494** CITATIONS
SEE PROFILE

analytical chemistry

Article

pubs.acs.org/ac

Evaluation of Fast 2D NMR for Metabolomics

Adrien Le Guennec,[†,‡] Patrick Giraudeau,*[,‡] and Stefano Caldarelli*[,†,§]

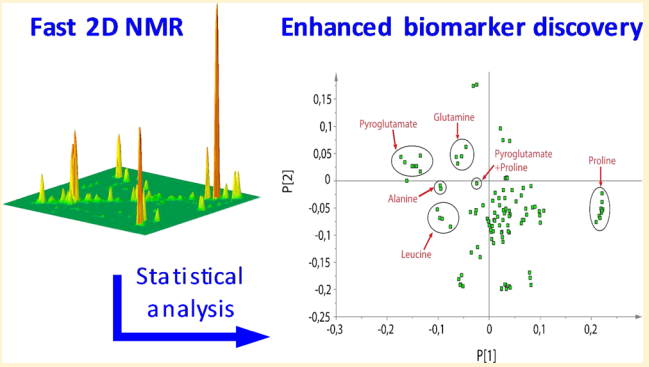
[†]Centre de Recherche CNRS de Gif-sur-Yvette, Institut de Chimie des Substances Naturelles, Laboratoire de Chimie et Biologie Structurales, UPR 2301, 1, avenue de la Terrasse, 91198 Gif-sur-Yvette, France

[‡]Université de Nantes, CNRS, CEISAM UMR 6230, BP 92208, 2 rue de la Houssinière, F-44322 Nantes Cedex 03, France

[§]Aix Marseille Université, Centrale Marseille, CNRS, iSm2 UMR 7313, 13397, Marseille, France

Supporting Information

ABSTRACT: Two-dimensional nuclear magnetic resonance (2D NMR) is increasingly explored as a tool for metabolomics because of its superior resolution compared to one-dimensional NMR (1D NMR). However, 2D NMR is characterized by longer acquisition times, which makes it less suitable for high-throughput studies. In this Article, we evaluated two methods for the acceleration of *n*D NMR, ultrafast (UF) and nonuniform sampling (NUS), in the context of metabolomics. To this end, model samples mimicking the metabolic profile variations in serum from subjects affected by colorectal cancer and controls were analyzed by 1D ^{1}H NMR along with conventional and accelerated DQF-COSY and HSQC. A statistical analysis (OPLS-DA) yielded similar results for the group separation with all techniques, but biomarker identification from 2D spectra was substantially enhanced, both in terms of number of molecules and easiness of assignment. Most interestingly, fast 2D NMR techniques lead to similar results as conventional 2D NMR, opening the way for high-throughput metabolomics studies using 2D NMR.



Metabolomics is defined as the characterization of the largest possible ensemble of metabolites (ideally all of them) as a phenotyping tool.[1] 1D ^{1}H NMR along with multivariate analysis has become one of the main analytical tools in metabolomics, in a very large range of scientific domains, ranging from toxicology and plant biology to clinical studies, to name just a few.[2−5] Today, the NMR analysis of biological samples for metabolomics has become a standardized routine.[4,6]

Despite the good sensitivity and the relatively straightforward acquisition of ^{1}H 1D spectra, this methodology has one main disadvantage: the high degree of spectral overlap because of the molecular complexity of the targeted samples associated with the low spectral range of ^{1}H resonances.[7] Overlap can confuse the subsequent statistical analysis, particularly if a biomarker signal is buried under those of nondiscriminating molecules. The large information redundancy associated with a global metabolic perturbation may partially compensate for this effect in simple phenotyping classification, but clearly the lack of an as-complete-as-possible set of biomarkers hinders the identification of the perturbed metabolic pathways and thus any further biological interpretation of the metabolomic results.

Spectral deconvolution, the determination of the individual 1D spectra of the known metabolites present in the sample,[8,9] may be a possibility to overcome the problems arising from signal overlap. But this approach presents its own limitations, namely the presence of unknown compounds and deviation of the signal from an ideal line shape. Moreover, the method relies on spectral databases of individual compounds, and consequently, it is sensitive to pH and to possible interactions between metabolites.

Spectral overlap can be efficiently reduced by increasing spectral dimensionality[10,11] and thus an increased interest has been shown toward the use of 2D NMR for quantitative analysis during the past decade.[12−16] The use of 2D NMR for metabolomics has attracted attention during these past years, demonstrating an easier identification of biomarkers than with 1D spectra.[7,17−21] A detailed description of what has been achieved in the field is given in a recent review by Bingol and Bruschweiler.[7] However, since classic 2D NMR spectra are acquired as series of 1D spectra, the former take a longer time to acquire than the latter.

Longer experimental durations have three main consequences: (i) 2D spectra are usually more sensitive to spectrometer instabilities, usually showed by the so-called t_1 noise,[22] (ii) it causes timetable constrains, especially in the case of high-throughput analysis or of rapid clinical responses, and (iii) 2D NMR is ill-suited for the study of unstable samples, such as those undergoing chemical or biochemical reactions.

Received: March 17, 2014
Accepted: May 23, 2014
Published: May 23, 2014

   dx.doi.org/10.1021/ac500966e | *Anal. Chem.* 2014, 86, 5946−5954

Table 1. Mean Concentrations for the Control Group and Diseased Group with Standard Deviations for Both Groups and Fold Change to Standard Deviation Ratio, Which Is Calculated with the Following Formula: FC/SD = (mc_{canc} − mc_{heal})/SD

metabolite	mean concentration in synthetic serum of control patients (mc_{heal}) (mM)	mean concentration in synthetic serum of diseased patients (mc_{canc}) (mM)	standard deviation (SD) (mM)	fold change to standard deviation ratio[a] (FC/SD)
3-hydroxybutyrate	1.6	2.2	1.3	0.47
acetic acid	0.9	1.3	0.3	**1.48**
alanine	8.6	6.8	1.7	**−1.04**
arginine	2.3	2.3	0.3	0.14
asparagine	1.6	1.7	0.1	0.69
aspartic acid	0.4	0.4	0.1	−0.01
betaine	1.4	1.3	0.4	−0.35
carnitine	0.8	0.9	0.2	0.43
citric acid	2.2	1.9	0.5	−0.52
creatine	0.6	0.4	0.6	−0.30
creatinine	1.6	1.7	0.4	0.41
glutamic acid	2.1	1.9	0.3	−0.64
glutamine	10.4	8.0	2.4	**−1.01**
glycerol	8.9	8.6	2.0	−0.13
glycine	6.3	6.2	2.5	−0.01
histidine	2.7	2.9	0.7	0.20
isoleucine	1.2	1.2	0.4	0.09
lactic acid	29.2	26.3	7.4	−0.39
leucine	2.0	1.7	0.2	**−1.15**
lysine	3.7	3.7	1.2	0.00
methanol	1.6	1.6	0.3	0.16
methionine	0.6	0.6	0.1	0.08
ornithine	1.4	1.2	0.3	−0.46
phenylalanine	1.5	1.8	0.4	0.58
proline	3.9	6.8	1.3	**2.30**
pyruvic acid	0.6	0.6	0.5	0.09
serine	3.1	3.0	0.5	−0.06
threonine	2.5	2.5	0.8	0.00
tryptophan	1.1	1.1	0.2	0.16
valine	4.4	3.8	1.2	−0.50

[a]Positive values for FC/SD shows indicates that the metabolite is more concentrated in samples from diseased patients and vice versa. Bold numbers indicates metabolites that are the most important for separation with OPLS-DA.

These are common shortcomings of multidimensional NMR, and indeed one of the main research challenges in this area is the development of approaches to reduce experimental time in nD NMR. These approaches can be classified in four categories. The first one focuses on reducing the actual spectral width in the indirect dimension using aliasing/folding.[13,23] The second one aims at reducing the interscan delay, like the band-selective optimized-flip-angle short-transient (SOFAST),[24] along with the similar band-selective excitation short-transient (BEST),[25] acceleration by sharing adjacent polarization (ASAP),[26,27] and small recovery times (SMART).[28] A third family of approaches exploits the sparse nature of the NMR signal in *n* dimensions, and includes nonuniform sampling (NUS), Hadamard,[29] and projection reconstruction[30] sampling. Finally, a completely different method introduces multiplexing instead of sequential sampling in the indirect dimension, by spatial encoding, and goes under the name of ultrafast (UF) 2D NMR.[31]

All of these approaches have advantages and drawbacks. Aliasing/folding, which reduces the observed spectral width in the indirect dimension relies on specific conditions in which either the folded peaks are known to occupy an empty area of the spectrum, or additional experiments are performed to determine the true chemical shift of folded/aliased peaks.[32] These conditions are increasingly harder to meet when the number of components increases. NUS allows the reconstruction of spectra similar to those obtained with conventional 2D NMR, but its performance depends of the algorithm used to reconstruct the nonsampled points prior to Fourier transform.[33] SOFAST/BEST manages to reach very short interscan delays, but only works for systems in which spin diffusion is an effective relaxation mechanism, such as macromolecules or small molecules in viscous solvents.[34] It is therefore a priori not adapted to metabolomics samples. Likewise, ASAP allows very short interscan delays (a few tens of microseconds) but is restricted to heteronuclear experiments for samples at natural abundance or slightly enriched. SMART also allows very low interscan delays, but requires rather high concentrations and access to triple axis gradients. The experimental time in Hadamard spectroscopy no longer depends on the number of t_1 increments to record, but on the number of regions of interest the spectroscopist wants to record. However, the acquisition relies on having a priori information about the peak positions. Projection--reconstruction is available only for *n*D spectra with $n > 2$, but it can allow the construction of 2D spectra with maximum resolution between close peaks.[35] Finally, UF 2D NMR offers the most important timesaving to record 2D spectra, but a compromise is necessary between spectral widths, resolution, and sensitivity.[36] The last point can be reduced by the multiscan single-shot approach (M3S).[37] Finally, whatever the methodological approach, sensitivity limitations of hetero-

dx.doi.org/10.1021/ac500966e | Anal. Chem. 2014, 86, 5946−5954

nuclear experiments can also be mitigated during sample preparation by ^{13}C- or ^{15}N-labeling.[38−40]

Despite the potential of fast 2D techniques to obtain spectra in shorter acquisition times, very few studies have been conducted on their potential for complex metabolic mixture analysis, to the best of our knowledge.[7,41−44] So far, these studies have focused on the quantification of metabolites or on targeted studies. No evaluation of fast 2D NMR for metabolomics has been yet reported. A first evaluation of this potential is presented in this study. NUS and UF 2D NMR have been selected for this evaluation, as the linearity of these approaches in the range of concentrations typically encountered in metabolomics has already been verified,[41,44] as well as the repeatability of peak volumes.[41,45] These are indispensible requirements for metabolomics. Aliasing and Hadamard have been excluded because of their main downsides mentioned above, which are potentially troublesome for the analysis of a complex mixture, while ASAP was not used in this study since the evaluation of its analytical performance has not been reported yet.

Model samples were used that mimicked the composition of serum, a rather simple biofluid, for controls, and patients affected by colorectal cancer.[46,47]

MATERIALS AND METHODS

Sample Preparation. Thirty-nine synthetic samples, each containing 30 metabolites, were prepared for metabolomic studies to reproduce typical concentration variations in the serum of patients affected by colorectal cancer (19 samples) and controls (20 samples).[46−48] Metabolite concentrations were enhanced by a factor 20 to avoid being limited by sensitivity in this first study. Standard deviations of individual metabolites were the same for the two groups. Solutions were prepared in 90% H_2O/10% D_2O with a phosphate buffer (0.1 M, pH 7), 0.5 mM TSP for chemical shift reference, and around 4 mM of NaN_3 to prevent possible bacteria proliferation. The mean concentration and standard deviations in the two groups are summarized in Table 1 and a complete table of the concentration of all metabolites in each sample is presented in the Supporting Information.

NMR Spectroscopy. A typical 1D pulse sequence in metabolomics was used for comparison with conventional 2D pulse sequences. Two different 2D pulse sequences, DQF-COSY, and HSQC, were evaluated to highlight potential differences between homonuclear and heteronuclear spectra. These two particular pulse sequences were chosen because they have already been used for complex mixture analysis and the resulting spectra are inherently sparser than the two other commonly used 2D pulse sequences, that is, *J*-resolved spectroscopy and TOCSY. This property makes these two pulse sequences more adequate for NUS. Conventional 2D NMR was performed using standard pulse sequences, acquisition and processing parameters.

For conventional DQF-COSY and HSQC spectra, the data were acquired with a Bruker Avance I 600 MHz spectrometer equipped with a 5 mm cryoprobe TXI triple resonance. However, NUS spectra could not be easily acquired with this spectrometer and UF spectra were not optimal because the available gradient strength was not sufficient for the characterization of the whole aliphatic region.[49] Thus, a Bruker Avance III 500 MHz spectrometer equipped with a 5 mm $^1H/^{13}C$ dual cryoprobe was used for UF COSY. NUS spectra and 1D spectra were subsequently acquired on a Bruker Avance III 600 MHz spectrometer equipped with a 5 mm cryoprobe TCI triple resonance.

^1H 1D spectra were acquired using the pulse sequence from the Bruker library *noesypr1d*, with 4 dummy scans (DS), 8 scans (NS), a repetition time (including acquisition) of 5.3 s, 32k points, and a short mixing time, as suggested elsewhere,[50] optimized at 1 ms for solvent suppression. Total acquisition time was 1 min 3 s. The FID was weighed with an exponential function of 0.3 Hz.

Phase sensitive DQF-COSY spectra (States-TPPI) were acquired using the sequence from the Bruker library *cosydfphpr* with DS = 16, NS = 8, 4k × 256 data points and a repetition time of 3.3 s, for a total acquisition time of 1 h 55 min 19 s. The FID was weighed with an unshifted sine-bell function in both dimensions prior to Fourier transform. Linear prediction was used to double the number of points in the indirect dimension, since it reduces the volume of the buckets and it does not affect repeatability.[45] Zero-filling was employed in the indirect dimension up to 1024 points. 2-fold zero-filling was also applied in the direct dimension. For multivariate analysis, DQF-COSY spectra were processed in magnitude mode since the phased mode leads to antiphased peaks whose total volume is zero.

For the NUS variant, only 30% of the t_1 increments were acquired, which reduced the acquisition time to 34 min 36 s. The reconstruction of the 2D spectrum was performed using the compressed sensing algorithm available within the Bruker software.[51,52]

The impact of the reconstruction algorithm for NUS data was evaluated using 50% of NUS for the two pulse sequences and two different algorithms proposed by Topspin: compressed sensing and recursive multidimensional decomposition (MDD).[53] We found that compressed sensing led to 2D spectra with less noise in the whole spectrum (data not shown). Accordingly, all NUS spectra were processed with compressed sensing.

Phase-sensitive HSQC spectra (States-TPPI) were obtained with the pulse sequence from the Bruker library *hsqcgpph*, in which water presaturation during the recovery delay was added, with 16 dummy scans, 8 scans for t_1 increment, 4k × 128 data points and a repetition time of 3.3 s, for a total acquisition time of 57 min 37 s. The FID was weighed with a $\pi/2$ shifted sine-bell function in both dimensions prior to Fourier transform. Linear prediction was used to double the number of points in the indirect dimension, followed by zero-filling up to 1024 points.

For the NUS variant, only 50% of t_1 increments were acquired, which reduced the acquisition time to 29 min 15 s. The reconstruction of the 2D spectrum was performed with compressed sensing, like with DQF-COSY.

The percentage of NUS was chosen based on 2 parameters: the fidelity of the peak intensities and the t_1 noise. For DQF-COSY, the threshold value was 30% of NUS. Lower values led to a dramatic increase of t_1 noise ridges, and the peak intensity started to drop below 25% of NUS. For HSQC, the intensity for small peaks started to drop below 50% of NUS. Accordingly, we preferred to work at 50% of NUS for HSQC, but if sensitivity is not an issue, then it may also be possible to work around 30% of NUS with HSQC.

UFCOSY spectra were recorded using a pulse sequence and parameters recently described.[41] Briefly, spatial encoding was performed using two smooth chirp pulses of 15 ms duration with a sweep range of 60 kHz, applied simultaneously with a 5.33 G/cm bipolar gradient pair. For coherence selection, gradients of opposite signs were applied with a sine shape, with

dx.doi.org/10.1021/ac500966e | Anal. Chem. 2014, 86, 5946−5954

amplitude of 58.5 G/cm and duration of 1 ms. During the acquisition, 128 gradient pairs were applied, with an amplitude 52 G/cm and a duration of 256 μs. The other parameters were NS = 176 and a relaxation time of 5.1 s for an acquisition time of 14 min 57 s. The processing was performed with a home-written routine, as detailed elsewhere,[41] using an optimized Gaussian weighting function in the spatially encoded dimension and a $\pi/4$ shifted squared sine-bell weighting function in the conventional dimension.

Before multivariate analysis, a polynomial baseline correction of order 2 was applied in all dimensions. The chemical shift was calibrated using the TSP peak, except for the UFCOSY, for which the diagonal peaks of lactate were used for calibration.[54]

Multivariate Analysis. Buckets in 1D and 2D spectra were defined manually for each spectrum. The buckets were defined to minimize interaction between peaks. Around 70−120 buckets were defined for each spectrum. 2D Peak volumes (or 1D peak areas) were extracted using the integration tool of Topspin. Volumes/areas were then normalized relatively to the total volume/area. The data sets were then imported in SIMCA, version 13, in which the data were scaled using unit variance (UV), and analyzed using supervised orthogonal partial least squares-discriminant analysis (OPLS-DA) multivariate analysis. For completeness, Pareto scaling was also tried instead of UV (see Supporting Information Figure S1), but, as expected for well-resolved spectra, this choice resulted in a less clear biomarker identification. The loading scatter plots were subsequently analyzed to highlight potential biomarkers. Multivariate analysis has also been done from the concentration of each metabolite, using the table given in the Supporting Information.

RESULTS AND DISCUSSION

Multivariate Analysis of Metabolites Concentrations Using OPLS-DA. Metabolomics relies on spectroscopic data to approximate metabolite variations. In the present case, this latter information was available, and thus a statistical model was built for reference, to which the models built on NMR will be compared for a measure of faithfulness. The OPLS-DA analysis on the metabolite concentrations is shown in Figure 1 (PCA results are also shown in Supporting Information, Figure S2). Despite the fact that only 16% of total variance was explained by the OPLS-DA model, a good separation of the two groups is still achieved with OPLS-DA. Cross-validation using CV-ANOVA showed a p-value below 10^{-3}, demonstrating that this separation is statistically relevant. The model is also robust, as shown by a permutation test on PLS-DA (Supporting Information Figure S3). Using the loading plot (Figure 1a) and the variable importance in the projection (VIP), it can be concluded that five metabolites are mainly responsible for group separation: proline, acetic acid, glutamine, leucine and alanine. From Table 1, it can be shown that, not surprisingly, these metabolites are the ones with a ratio fold change to standard deviation (FC/SD) greater than 1. Indeed, intragroup variance, as described by the standard deviation, must not exceed the intergroup one (the fold change) for a variable to be discriminant. Thus, metabolites like creatine and 3-hydroxybutyrate, which have high fold change but also high standard deviation, have little weight for group separation.

Analysis of 1D and 2D Spectra. Figure 2 shows the spectra obtained with the pulse sequences described in the Materials and Methods section. The 1D spectrum, even with these rather simple samples, is dominated by overlapping peaks. The regions between 2 and 2.5 ppm and between 3 and 4 ppm, where 4 to 7

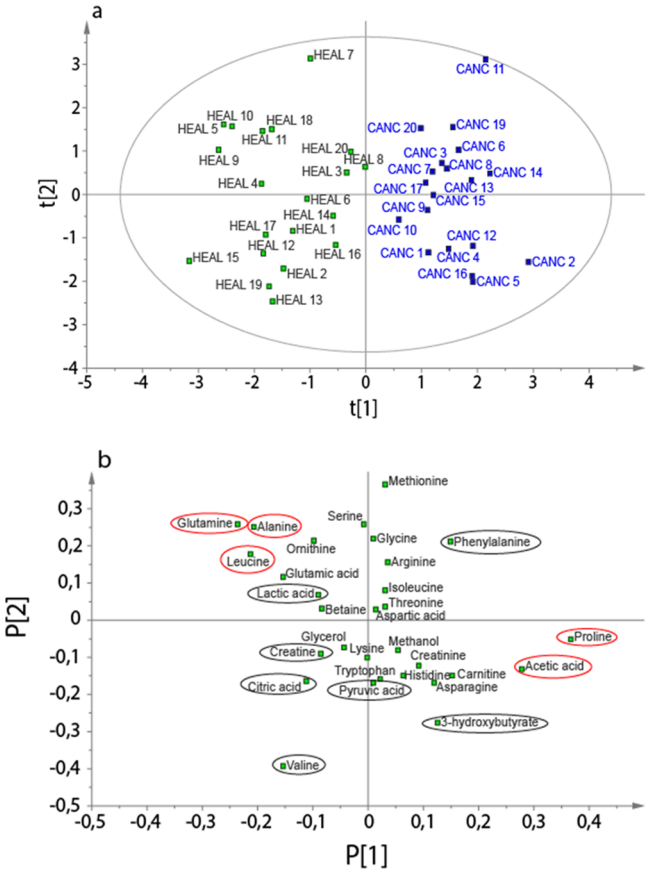


Figure 1. (a) Score scatter plot from the OPLS-DA of the concentration of individual metabolites in each sample (values in Supporting Information). (b) Loading scatter plot of the same data. Encircled metabolites are potential biomarkers. Those encircled in red are metabolites that contribute the most to the separation of groups.

multiplets can overlap, are particularly crowded. Overlap becomes much less significant in 2D spectra: most of the signals overlapping in the 1D spectrum become resolved. The remaining overlaps in 2D spectra (both COSY and HSQC) concern mainly lysine, ornithine, and arginine, as well as glutamate and glutamine, which have very similar structures.

It can be seen from Figure 1b−e that uniformly sampled (US) and NUS spectra are very similar, both in peak position and intensity. It seems that NUS spectra have less t_1 noise, which is especially visible in the DQF-COSY spectra, although it is difficult to say at this point if it is caused by the reduced experimental time or the fact that they were recorded on two different spectrometers.

Because of the somewhat low concentration values, UF-COSY was acquired using the M3S approach,[37,41] which consists in adding UF spectra to increase the signal-to-noise ratio (SNR). Under these conditions, it can be seen in Figure 1f that most of the correlation peaks for COSY are visible on the UF COSY spectrum recorded in 15 min. Note that, because of the spectral width limitation of UF COSY, only the aliphatic region was recorded. This is not an uncommon procedure in metabolomics, since this area possesses most of the signal, both in terms of number and intensity of peaks. Although new interleaved approaches to record M3S spectra with large spectral widths have recently been proposed,[55] their analytical performance is still unexplored. Of special interest are the correlation peaks of methionine, aspartic acid, and carnitine, for which the

dx.doi.org/10.1021/ac500966e | Anal. Chem. 2014, 86, 5946−5954

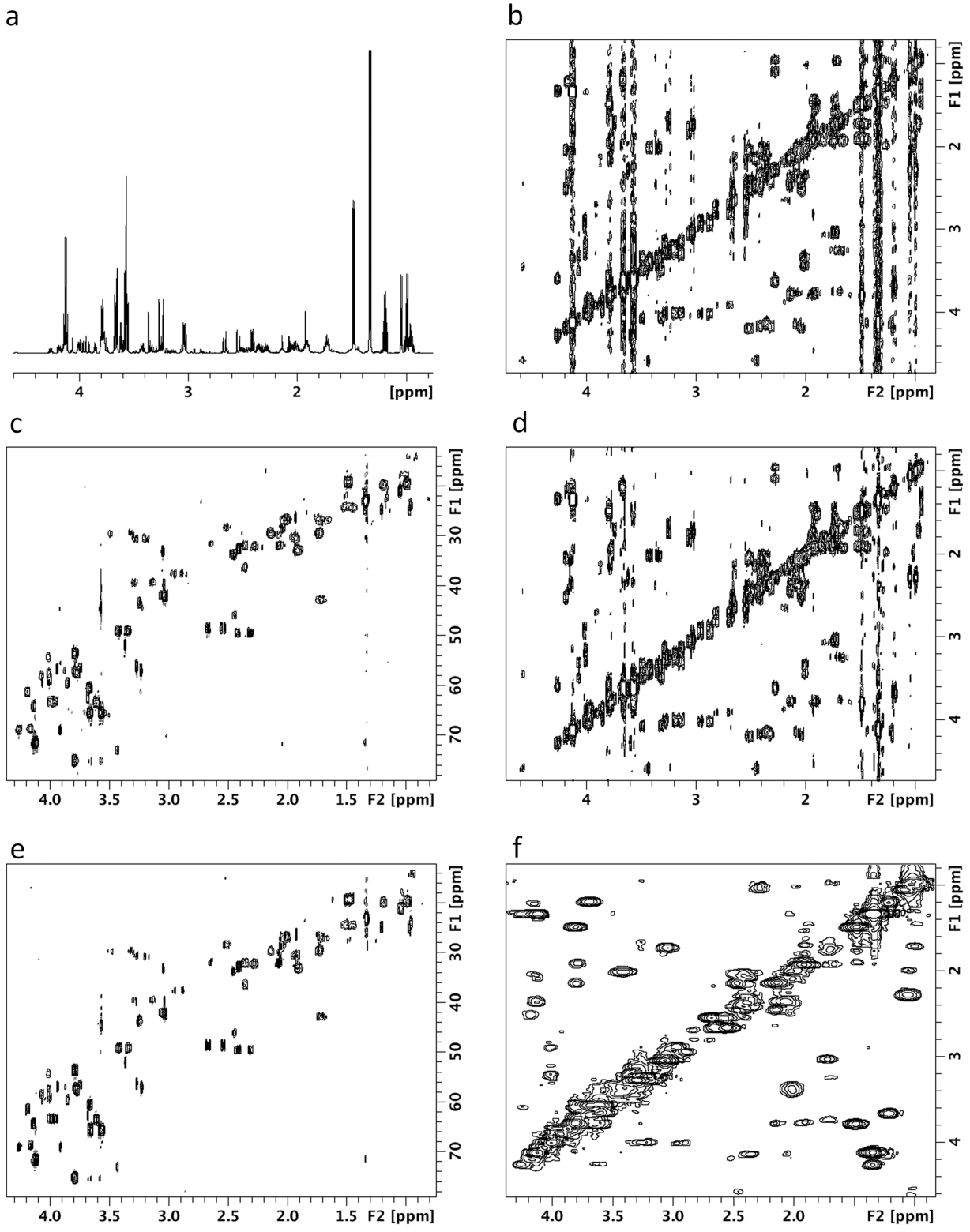


Figure 2. NMR spectra of the aliphatic region of a model synthetic blood sample (mixture of 30 metabolites) obtained with the following conditions: (a) 1D NOESY ^{1}H spectrum with water signal presaturation recorded in 1 min with 8 scans. (b) Conventional ^{1}H double quantum filtered COSY spectrum recorded in 2 h min with 8 scans and 256 t_1 increments. (c) Conventional ^{1}H−^{13}C HSQC spectrum recorded in 1 h min with 8 scans and 128 t_1 increments. (d) Nonuniformly sampled (NUS) ^{1}H double quantum filtered COSY spectrum recorded in the same conditions as b, with 30% of sparse sampling, which resulted in 35 min of acquisition. (e) Nonuniformly sampled (NUS) ^{1}H−^{13}C HSQC spectrum recorded in the same conditions as d, with 50% of sparse sampling, which resulted in 30 min of acquisition. (f) Multiscan single shot (M3S) ^{1}H COSY spectrum recorded in 15 min by repeating 176 times an ultrafast COSY pulse sequence. All the spectra were recorded on a 600 MHz spectrometer with a cryoprobe, except the M3S spectrum, which was recorded on a 500 MHz spectrometer with a cryoprobe.

dx.doi.org/10.1021/ac500966e | *Anal. Chem.* 2014, 86, 5946−5954

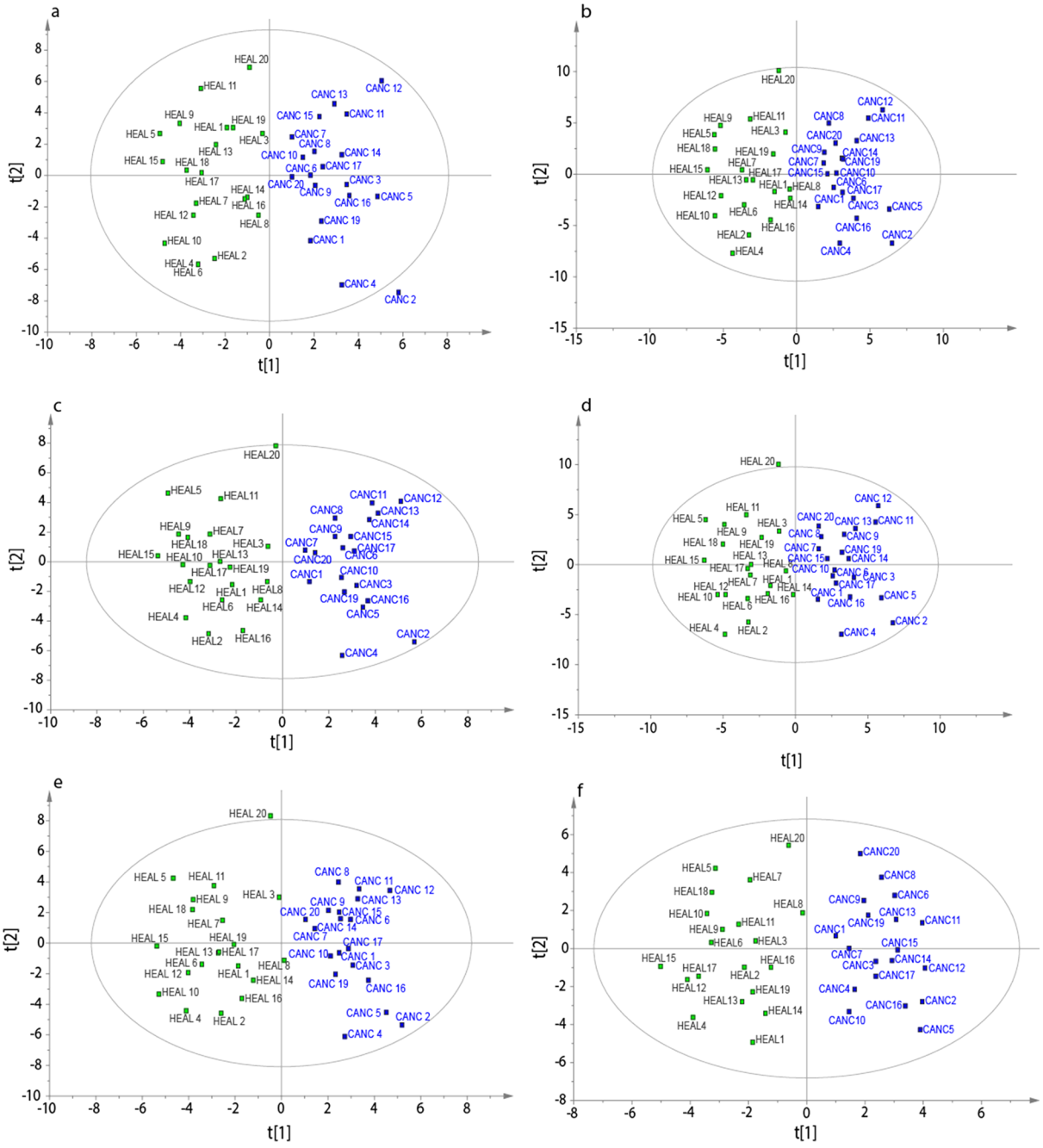


Figure 3. Score plots from OPLS-DA. The raw data was constructed by taking the surface (for 1D) or the volume (for 2D) of each bucket in the spectra of the 39 samples, using the TOPSPIN integration tool. The multivariate analysis was performed by SIMCA version 13. (a–f) Same legend as Figure 1

concentrations in the synthetic samples are almost always below 1 mM. This confirms findings from previous work[41] where submillimolar concentrations could be detected and quantified by the M3S approach using a magnetic field of 500 MHz. Instead of UF DQF-COSY, UF COSY was chosen for two main reasons: (i) UF DQF-COSY suffers from sensitivity issues compared to UF COSY, (ii) one of the main reasons for the use of DQF-COSY instead of COSY is the removal of the dispersive components from diagonal peaks, but it is not relevant in UF COSY where the peak shapes are purely absorptive.[56] The reason for this absence of dispersive components in UF COSY is that UF COSY is closer in nature to constant-time COSY[57] because of the spatial encoding strategy[58] chosen for this work. UF HSQC was also tested, but a preliminary study showed that it was not competitive in terms of sensitivity compared to its conventional counterpart.

Separation of Groups and Identification of Potential Biomarkers Using OPLS-DA. Figure 3 shows the score plots obtained from the spectra acquired with the different pulse sequences, and summarizes the good group separation capacity for all of them. Like with Figure 1a, a good group separation is achieved (p-value always below 10^{-4}) for all 1D and 2D sequences despite the low variance explained by the model (between 10% and 13% for the first parameter depending of the sequence).

dx.doi.org/10.1021/ac500966e | Anal. Chem. 2014, 86, 5946–5954

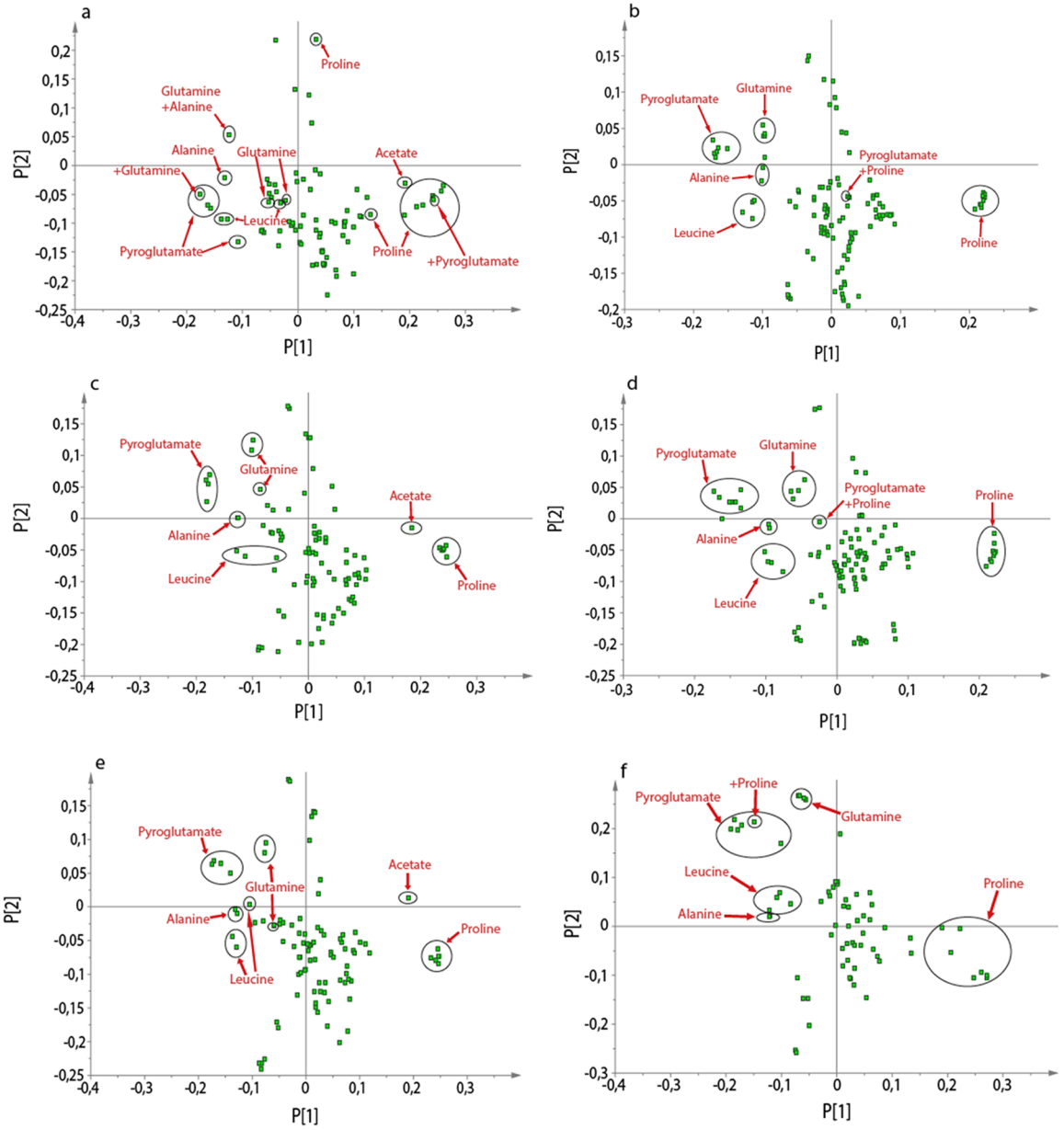


Figure 4. Loading plot from the OPLS-DA, from the same data as in Figure 2. (a–f) Same legend as Figures 2 and 3. The circled regions show where the location of the most important biomarkers for separation of groups (see Figure 1 and Results and Discussion for more details). Overlap between biomarkers is indicated with two circles for the same bucket.

It should be noted that multivariate analysis with 1D spectra also shows a very good separation of the two groups, which can be explained by the relative simplicity of the model serum samples.[46] Therefore, the interest of 2D NMR for group separation cannot be demonstrated from these samples, but it is not the main purpose of this study which concentrates on fast approaches, and proofs of better separation by 2D spectra are reported elsewhere.[17,19,59]

In contrast, the loading plots (Figure 4) show clear differences between pulse sequences. Before starting the analysis, we should point out that one of the potential biomarkers found in all cases was not in the initial list (Table 1): pyroglutamate. This metabolite is one of the degradation products of the glutamine,[60] which was catalyzed by the phosphate buffer in the samples. While the degradation has rather low kinetics, a few months passed between the preparation of samples and the final metabolomics analysis, which explains its presence in the samples. Since glutamine is expected to be a biomarker (Table 1), pyroglutamate should also be one. For the sake of discussion, the buckets of both pyroglutamate and glutamine will be thus analyzed.

The OPLS-DA analysis was parametrized for one discriminating latent variable, thus the only axis relevant for assessing the biomarker discrimination is the one reported on the abscissas. The biomarker identification is strongly impeded in the statistics arising from 1D spectra (Figure 4a). Here, because of heavy overlap, two or more metabolites, including biomarkers, can contribute to the total intensity of a bucket. In conjunction with UV normalization, this can cause some buckets containing a given metabolite to scatter across the loadings plot, away from the other signals of the same compound. Pyroglutamate, glutamine and proline are interesting examples of this situation. Indeed, in the 1D spectrum all the peaks from pyroglutamate overlap with peaks from other

dx.doi.org/10.1021/ac500966e | Anal. Chem. 2014, 86, 5946–5954

biomarkers (proline, 3-hydroxybutyrate, glutamine, and citrate). From five buckets containing pyroglutamate peaks, only three are found to be relevant to discrimination (considering the results from Figure 1b for a cutoff value) while the other two are within the regions of another biomarker. Most of the buckets containing proline peaks are within the same region, although they overlap with the signal of different metabolites. One exception is a proline peak that overlaps with one from lactate. Since lactate is the most abundant metabolite in the samples, the information from proline is lost because of the small relative weight of proline, further reduced after normalization. For glutamine, the three multiplets overlap heavily with other peaks. Four buckets contain glutamine peaks, two of them overlapping with other biomarkers. These latter two buckets have their position in the loading plot influenced by the other biomarkers while the former two others assume a low contribution, thus preventing the identification of this metabolite as a biomarker.

These examples reinforce the known fact that overlap can seriously hinder biomarker identification and discovery. While redundancy of information in NMR can help making overlap less of an issue for group discrimination, the lost part of the information can make biomarker identification less certain and create ambiguity in data analysis, since most of the buckets end up containing peaks from different species. On the other hand, the use of 2D NMR greatly reduces the overlap and, as a result, the buckets for a given biomarker are more uniquely identified. The result is that they will be assigned by the statistical model to have similar impact on the loadings, as it can be seen in the plots corresponding to the 2D experiments in Figure 4. A very important point, that has not been discussed previously, is that the profile of the loading plots from 2D spectra is closer to the one obtained directly from the molecular concentrations, that is the "ground truth" (Figure 1). Thus, the enhanced resolution makes multivariate analysis of 2D peak volumes closer to the property we are evaluating, that is, the changes in concentration. This is consistent with the quantitative nature of 2D NMR experiments.[61] Interestingly, in this context peaks from the residual glutamine appear to be slightly less discriminatory than pyroglutamate (Figure 4), and it is this latter that occupies, in the loading plots of Figure 4, a position more closely resembling the one occupied by glutamine in Figure 1b. In line with these findings, *J*-resolved spectroscopy, another widely used 2D experiment in metabolomics, provided intermediate results between the ones obtained with 1D and 2D spectra, because of its reduced discrimination power compared to other 2D techniques (Supporting Information Figure S4).

More specifically to the 2D experiment tested here, the main problem for COSY (and homonuclear sequences in general) is the lack of cross-peaks for metabolites without J-couplings, like acetic acid, which in this case is a biomarker. This is not the case for heteronuclear sequences (like HSQC), where all the protons visible in 1D are also potentially visible in HSQC. The trade-off is the reduced sensitivity. Thus, although the COSY spectra offered more sensitivity per unit of time and provided excellent group discrimination, in the present model they could not detect all discriminant biomarkers (Figure 4b and d), contrary to HSQC that provided a biomarker representation very close to the ground truth.

A second interesting aspect stemming from the analysis of Figure 4 is that there is little difference in the biomarker identification for the spectra acquired with either US (Figures 4b and c) or NUS (Figures 4d and e) approaches. For the UF COSY (Figure 4f), even if the loading plot is slightly different from the conventional DQF-COSY, the coefficients for the first parameter (the one which is important for separation) are similar to the other pulse sequences. The gathering of buckets from the same metabolite is also present in the loading plot, but the low sensitivity of some cross-peaks of proline cause some of its buckets to have a slightly lower weight than expected. Despite this, the identification of biomarkers using this methodology is still a much easier task than for 1D. From this analysis, it can be concluded that the use of the two fast 2D approaches allows retrieving the same information as for conventional 2D NMR with a smaller experimental time, which is of great interest for metabolomics.

CONCLUSION

This study on a simple but realistic model demonstrates that fast 2D NMR provides the same information as conventional 2D NMR in the case of typical metabolomics studies of biofluids. Since 2D NMR allows a more reliable identification of potential biomarkers than ^{1}H 1D NMR, fast 2D NMR can become a powerful tool for high-throughput studies thanks to acquisition times closer to those in 1D NMR. The two methodologies can be combined with realistic experimental times for enhanced information content.[62,63]

Even in the case of sensitivity limitations, the use of fast 2D approaches is still beneficial, since they give more sensitivity per unit of time.[37,41,64,65] Regarding UF spectroscopy, so far, only homonuclear experiments have shown a sufficient performance to envisage metabolomics applications, but ongoing methodological developments will hopefully help improving the sensitivity and the robustness of UF heteronuclear pulse sequences. Research perspectives also include the combination of several fast 2D approaches, either to compress further the acquisition time, like a combination of NUS and ASAP, or for enhanced performances, for example by combining NUS and UF.[66]

ASSOCIATED CONTENT

Supporting Information

Additional material as described in the text. This material is available free of charge via the Internet at http://pubs.acs.org/

AUTHOR INFORMATION

Corresponding Authors

*Tel. +33(0)2 51 12 57 09. Fax. +33(0)2 51 12 57 12. E-mail: patrick.giraudeau@univ-nantes.fr.

*E-mail: stefano.caldarelli@cnrs.fr.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are grateful to Estelle Martineau for help with multivariate analysis and Jean-Nicolas Dumez for insightful comments. This research was supported by a young investigator starting grant from the "Agence Nationale de la Recherche" (ANR grant 2010-JCJC-0804-01).

REFERENCES

(1) Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. *Trends Biotechnol.* **2004**, *22*, 245−252.

(2) Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E.; et al. *Nat. Rev. Drug Discovery* **2002**, *1*, 153−162.

(3) Duarte, I. F.; Gil, A. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2012**, *62*, 51−74.

(4) Emwas, A.-H.; Salek, R.; Griffin, J.; Merzaban, J. *Metabolomics* **2013**, *9*, 1048−1072.
(5) Rolin, D.; Deborde, C.; Maucourt, M.; Cabasson, C.; Fauvelle, F.; Jacob, D.; Canlet, C.; Moing, A. High-Resolution H-1-NMR Spectroscopy and Beyond to Explore Plant Metabolome. In *Metabolomics Coming of Age with Its Technological Diversity*; Rolin, D., Ed.; Academic Press: London, 2013; Vol. *67*, pp 1−66.
(6) Issaq, H. J.; Van, Q. N.; Waybright, T. J.; Muschik, G. M.; Veenstra, T. D. *J. Sep. Sci.* **2009**, *32*, 2183−2199.
(7) Bingol, K.; Bruschweiler, R. *Anal. Chem.* **2014**, *86*, 47−57.
(8) Wishart, D. S. *TrAC, Trends Anal. Chem.* **2008**, *27*, 228−237.
(9) Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M. *Anal. Chem.* **2006**, *78*, 4430−4442.
(10) Jeener, J. Presented at Ampere International Summer School, Basko Polje, Yugoslavia, 1971.
(11) Aue, W. P.; Bartholdi, E.; Ernst, R. R. *J. Chem. Phys.* **1976**, *64*, 2229−2246.
(12) Hu, F.; Furihata, K.; Kato, Y.; Tanokura, M. *J. Agric. Food. Chem.* **2007**, *55*, 4307−4311.
(13) Lewis, I. A.; Schommer, S. C.; Hodis, B.; Robb, K. A.; Tonelli, M.; Westler, W. M.; Sussman, M. R.; Markley, J. L. *Anal. Chem.* **2007**, *79*, 9385−9390.
(14) Martineau, E.; Tea, I.; Akoka, S.; Giraudeau, P. *NMR Biomed.* **2012**, *25*, 985−992.
(15) Bayet-Robert, M.; Loiseau, D.; Rio, P.; Demidem, A.; Barthomeuf, C.; Stepien, G.; Morvan, D. *Magn. Reson. Med.* **2010**, *63*, 1172−1183.
(16) Bingol, K.; Zhang, F.; Bruschweiler-Li, L.; Brüschweiler, R. *Anal. Chem.* **2013**, *85*, 6414−6420.
(17) Van, Q. N.; Issaq, H. J.; Jiang, Q.; Li, Q.; Muschik, G. M.; Waybright, T. J.; Lou, H.; Dean, M.; Uitto, J.; Veenstra, T. D. *J. Proteome Res.* **2008**, *7*, 630−639.
(18) Chikayama, E.; Suto, M.; Nishihara, T.; Shinozaki, K.; Hirayama, T.; Kikuchi, J. *PLoS One* **2008**, *3*, e3805.
(19) Yuk, J.; McKelvie, J. R.; Simpson, M. J.; Spraul, M.; Simpson, A. J. *Environ. Chem.* **2010**, *7*, 524−536.
(20) Ludwig, C.; Viant, M. R. *Phytochem. Anal.* **2010**, *21*, 22−32.
(21) Robinette, S. L.; Ajredini, R.; Rasheed, H.; Zeinomar, A.; Schroeder, F. C.; Dossey, A. T.; Edison, A. S. *Anal. Chem.* **2011**, *83*, 1649−1657.
(22) Morris, G. A. *J. Magn. Reson.* **1992**, *100*, 316−328.
(23) Vitorge, B.; Bieri, S.; Humam, M.; Christen, P.; Hostettmann, K.; Munoz, O.; Loss, S.; Jeannerat, D. *Chem. Commun.* **2009**, 950−952.
(24) Schanda, P. *Prog. Nucl. Magn. Reson. Spectrosc.* **2009**, *55*, 238−265.
(25) Schanda, P.; Van Melckebeke, H.; Brutscher, B. *J. Am. Chem. Soc.* **2006**, *128*, 9042−9043.
(26) Kupce, E.; Freeman, R. *Magn. Reson. Chem.* **2007**, *45*, 2−4.
(27) Schulze-Sünninghausen, D.; Becker, J.; Luy, B. *J. Am. Chem. Soc.* **2014**, *136*, 1242−1245.
(28) Vitorge, B.; Bodenhausen, G.; Pelupessy, P. *J. Magn. Reson.* **2010**, *207*, 149−152.
(29) Kupce, E.; Freeman, R. *J. Magn. Reson.* **2003**, *162*, 300−310.
(30) Kupce, E.; Freeman, R. *J. Am. Chem. Soc.* **2004**, *126*, 6429−6440.
(31) Tal, A.; Frydman, L. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57*, 241−292.
(32) Foroozandeh, M.; Jeannerat, D. *ChemPhysChem* **2010**, *11*, 2503−2505.
(33) Kazimierczuk, K.; Stanek, J.; Zawadzka-Kazimierczuk, A.; Kozminski, W. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57*, 420−434.
(34) Motta, A.; Paris, D.; Melck, D. *Anal. Chem.* **2010**, *82*, 2405−2411.
(35) Pontoizeau, C.; Herrmann, T.; Toulhoat, P.; Elena-Herrmann, B.; Emsley, L. *Magn. Reson. Chem.* **2010**, *48*, 727−733.
(36) Shrot, Y.; Frydman, L. *J. Chem. Phys.* **2009**, *131*, 224516−224516−11.
(37) Pathan, M.; Akoka, S.; Tea, I.; Charrier, B.; Giraudeau, P. *Analyst* **2011**, *136*, 3157−3163.
(38) Kikuchi, J.; Shinozaki, K.; Hirayama, T. *Plant Cell Physiol.* **2004**, *45*, 1099−1104.
(39) Gowda, G. A. N.; Tayyari, F.; Ye, T.; Suryani, Y.; Wei, S.; Shanaiah, N.; Raftery, D. *Anal. Chem.* **2010**, *82*, 8983−8990.
(40) Foston, M.; Samuel, R.; Ragauskas, A. J. *Analyst* **2012**, *137*, 3904−3909.
(41) Le Guennec, A.; Tea, I.; Antheaume, I.; Martineau, E.; Charrier, B.; Pathan, M.; Akoka, S.; Giraudeau, P. *Anal. Chem.* **2012**, *84*, 10831−10837.
(42) Hyberts, S. G.; Heffron, G. J.; Tarragona, N. G.; Solanky, K.; Edmonds, K. A.; Luithardt, H.; Fejzo, J.; Chorev, M.; Aktas, H.; Colson, K.; Falchuk, K. H.; Halperin, J. A.; Wagner, G. *J. Am. Chem. Soc.* **2007**, *129*, 5108−5116.
(43) Ludwig, C.; Ward, D. G.; Martin, A.; Viant, M. R.; Ismail, T.; Johnson, P. J.; Wakelam, M. J. O.; Günther, U. L. *Magn. Reson. Chem.* **2009**, *47*, S68−S73.
(44) Rai, R. K.; Sinha, N. *Anal. Chem.* **2012**, *84*, 10005−10011.
(45) Martineau, E.; Akoka, S.; Boisseau, R.; Delanoue, B.; Giraudeau, P. *Anal. Chem.* **2013**, *85*, 4777−4783.
(46) Psychogios, N.; Hau, D. D.; Peng, J.; Guo, A. C.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner, R.; Gautam, B.; Young, N.; Xia, J.; Knox, C.; Dong, E.; Huang, P.; Hollander, Z.; Pedersen, T. L.; Smith, S. R.; Bamforth, F.; Greiner, R.; McManus, B.; Newman, J. W.; Goodfriend, T.; Wishart, D. S. *PLoS One* **2011**, *6*, e16957.
(47) Bernini, P.; Bertini, I.; Luchinat, C.; Nincheri, P.; Staderini, S.; Turano, P. *J. Biomol. NMR* **2011**, *49*, 231−243.
(48) Bertini, I.; Cacciatore, S.; Jensen, B. V.; Schou, J. V.; Johansen, J. S.; Kruhoffer, M.; Luchinat, C.; Nielsen, D. L.; Turano, P. *Cancer Res.* **2012**, *72*, 356−364.
(49) Junior, L. H. K. Q.; Queiroz, D. P. K.; Dhooghe, L.; Ferreira, A. G.; Giraudeau, P. *Analyst* **2012**, *137*, 2357−2361.
(50) Mckay, R. T. *Concept. Magn. Reson. A* **2011**, *38*, 197−220.
(51) Holland, D. J.; Bostock, M. J.; Gladden, L. F.; Nietlispach, D. *Angew. Chem.* **2011**, *123*, 6678−6681.
(52) Kazimierczuk, K.; Orekhov, V. Y. *Angew. Chem.* **2011**, *50*, 5556−5559.
(53) Orekhov, V. Y.; Jaravine, V. A. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *59*, 271−292.
(54) Pathan, M.; Charrier, B.; Tea, I.; Akoka, S.; Giraudeau, P. *Magn. Reson. Chem.* **2013**, *51*, 168−175.
(55) Rouger, L.; Charrier, B.; Pathan, M.; Akoka, S.; Giraudeau, P. *J. Magn. Reson.* **2014**, *238*, 87−93.
(56) Shapira, B.; Lupulescu, A.; Shrot, Y.; Frydman, L. *J. Magn. Reson.* **2004**, *166*, 152−163.
(57) Bax, A.; Freeman, R. *J. Magn. Reson.* **1981**, *44*, 542−561.
(58) Pelupessy, P. *J. Am. Chem. Soc.* **2003**, *125*, 12345−12350.
(59) Mavel, S.; Nadal-Desbarats, L.; Blasco, H.; Bonnet-Brilhault, F.; Barthélémy, C.; Montigny, F.; Sarda, P.; Laumonnier, F.; Andres, C. R.; Emond, P. *Talanta* **2013**, *114*, 95−102.
(60) Khan, K.; Elia, M. *Clin. Nutr.* **1991**, *10*, 186−192.
(61) Giraudeau, P. *Magn. Reson. Chem.* **2014**, DOI: 10.1002/mrc.4068.
(62) Yuk, J.; Simpson, M. J.; Simpson, A. J. *Environ. Chem.* **2011**, *8*, 281−294.
(63) Yuk, J.; Simpson, M. J.; Simpson, A. J. *Environ. Pollut.* **2013**, *175*, 35−44.
(64) Paramasivam, S.; Suiter, C. L.; Hou, G. J.; Sun, S. J.; Palmer, M.; Hoch, J. C.; Rovnyak, D.; Polenova, T. *J. Phys. Chem. B* **2012**, *116*, 7416−7427.
(65) Rovnyak, D.; Sarcone, M.; Jiang, Z. *Magn. Reson. Chem.* **2011**, *49*, 483−491.
(66) Shrot, Y.; Frydman, L. *J. Magn. Reson.* **2011**, *209*, 352−358.