# Exploring the complementarity of fast multipulse and multidimensional NMR methods for metabolomics: a chemical ecology case study

Aurore Michaud, Samuel Bertrand, Serge Akoka, Jonathan Farjon, Estelle Martineau, Nicolas Ruiz, Thibaut Robiou Du Pont, Olivier Grovel, Patrick Giraudeau

ARTICLE

# Exploring the complementarity of fast multipulse and multidimensional NMR methods for metabolomics: a chemical ecology case study

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Aurore Michaud[a,b], Samuel Bertrand[b,c], Serge Akoka[a], Jonathan Farjon[a], Estelle Martineau[a], Nicolas Ruiz[b], Thibaut Robiou Du Pont[b], Olivier Grovel[b**], Patrick Giraudeau[a*]

This study investigates the potential and complementarity of high-throughput multipulse and multidimensional NMR methods for metabolomics. Through a chemical ecology case study, three methods are investigated, offering a continuum of methods with complementary features in terms of resolution, sensitivity and experiment time. Ultrafast 2D COSY, adiabatic INEPT and SYMAPS HSQC are shown to provide a very good classification ability, comparable to the reference 1D $^{1}$H NMR method. Moreover, a detailed analysis of discriminant buckets upon supervised statistical analysis shows that all methods are highly complementary, since they are able to highlight discriminant signals that could not be detected by 1D $^{1}$H NMR. In particular, fast 2D methods appear very efficient to discriminate signals located in highly crowded regions of the $^{1}$H spectrum. Overall, the combination of these recent methods within a single NMR metabolomics workflow allows to maximize the accessible metabolic information, and also raises exciting challenges in terms of NMR data analysis for chemical ecology.

## Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy is a major analytical tool for the analysis of complex mixtures.[1–4] It is particularly powerful for metabolomics, where NMR offers exceptional advantages for the profiling of complex biological samples such as plant, cell or bacterial extracts, or biofluids like urine, plasma, etc.[5] Thanks to its high reproducibility and throughput, and to its ability to provide both structural and quantitative information, NMR metabolomics has found applications in many areas of science, from (pre-)clinical studies to environmental or plant sciences, as well as in microbiology, food authentication, etc.

NMR metabolic profiling studies rely on a well-defined, multi-step workflow which starts by a careful design of the biological study, which aims at maximizing one or several biological differences between sample groups, while limiting confounding factors.[6] Then, biological samples are prepared for analysis (typically involving an extraction step) and analyzed by high-throughput NMR experiments. After signal processing, analytical data are subjected to statistical analysis steps that aim at observing the variability between sample groups, then identifying the signals that can explain this discrimination, to determine biomarkers of specific biological conditions and/or to better understand metabolic pathways.

The overwhelming majority of NMR metabolomics studies rely on the acquisition of 1D $^{1}$H fingerprints, which are recorded with standard pulse sequences involving solvent signal suppression.[7–10] 1D $^{1}$H spectroscopy has the advantage of being high-throughput (typically a few minutes per sample) and relatively simple to process. However, 1D $^{1}$H spectra are hampered by severe and ubiquitous peak overlaps arising from the diversity of analytes present in biological samples, associated with a limited frequency range and to the spreading of $^{1}$H multiplets. This issue is further compounded by the peak shifting that occurs within sample groups due to pH, concentration or ionic strength variations, making peak integration a complex task.

Fortunately, the NMR toolbox includes several promising alternatives for the analysis of complex mixtures[4], and new tools are constantly being developed.[11–14] The acquisition of $^{13}$C NMR spectra is an efficient way of improving the separation of overlapped metabolite signals,[15] however, the low sensitivity of $^{13}$C detection makes it impractical for metabolomics at natural abundance. A modest but significant sensitivity improvement by 3 to 4 can be achieved relying on polarization transfer methods such as INEPT (Insensitive Nuclei Enhanced by Polarization Transfer), whose analytical performance has recently been optimized to yield an efficient profiling tool.[16] In a slightly more distant future, hyperpolarization methods could change the sensitivity paradigm of $^{13}$C NMR metabolomics, albeit at the price of expensive instrumentation and operating costs.[13,17] The simplification of $^{1}$H

[a.] Nantes Université, CNRS, CEISAM UMR 6230, F-44000 Nantes, France. E-mail : patrick.giraudeau@univ-nantes.fr
[b.] Nantes Université, Institut des Substances et Organismes de la Mer, ISOMer, UR 2160, F-44000 Nantes, France. E-mail : olivier.grovel@univ-nantes.fr
[c.] Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France
Supplementary Information available:. See DOI: 10.1039/x0xx00000x

NMR spectral patterns through pure-shift NMR is another promising alternative, whose first applications to metabolomics have been recently reported.[18,19] However, pure-shift 1D $^{1}$H NMR suffers from a major sensitivity penalty and is also prone to artefacts in the case of strongly coupled spin systems. A more general and versatile alternative is the use of two-dimensional (2D) NMR spectroscopy, which has the advantage of offering an additional spectroscopic dimension compared to 1D NMR, thus offering an improved separation of overlapped peaks.[20] Moreover, 2D NMR also offers additional molecular identification capabilities by providing information on atomic connectivities, most often through *J*-couplings. Finally, 2D NMR offers a great variety of homo- or heteronuclear pulse sequences among which one can choose to obtain the best compromise between sensitivity and resolution for a given sample.[21]

On the one hand, 2D NMR has been recognized for decades as a useful structure elucidation tool in metabolomics. It is widely used to determine the structure of unknown metabolites, often after a purification step,[22] but also directly in complex mixtures.[23] On the other hand, while it has been shown that 2D NMR can provide improved metabolomics classification upon statistical analysis,[24] conventional 2D NMR experiments are rarely used as a profiling tool in untargeted metabolomics workflows, since the acquisition of 2D spectra on an entire sample cohort would take an unreasonable time, contradictory with the high-throughput need of metabolomics studies. This paradigm has been slightly changing since recent years have witnessed an increasing number of examples where fast 2D NMR experiments have been incorporated into such workflows.[25] Thanks to accelerated 2D NMR approaches such as non-uniform sampling (NUS), ultrafast (UF) 2D NMR, or fast-pulsing strategies, it has become possible to systematically record 2D data on a series of samples within a reasonable time. For instance, Watermann *et al.* have shown that the use of ASAP (Acceleration by Sharing Adjacent Polarization) HSQC (Heteronuclear Single Quantum Correlation) combined with NUS provided a good classification and an improved marker identification for the discrimination of the geographical origin of walnut samples.[26] NUS HSQC was also used for metabolic flux analysis.[27] Marchand *et al.* developed an untargeted lipidomic fingerprinting workflow relying on two fast homonuclear 2D NMR approaches that helped in the identification of biomarkers in a food chemical safety study.[28] In 2020, Féraud *et al.* reported an untargeted metabolomics study on biofluids with the use of NUS applied to a $^{1}$H COSY pulse sequence.[29] Such fast 2D NMR approaches have also been applied to the targeted quantification of metabolites in mixtures,[30] and also to isotope profiling, either in enriched samples[31] or at natural abundance.[32]

While the above-mentioned examples demonstrated the potential of advanced NMR pulse sequences to be efficiently incorporated into metabolomics workflows, very few actually compared the classification performance of such methods and their complementarity. In this context, the aim of this study is to evaluate and compare the classification performance of fast multiple pulse and/or multidimensional NMR profiling methods in a metabolomics context. We chose to focus on three of the most recent multiple pulse methods that were previously developed and optimized in terms of precision and rapidity for the analysis of complex mixtures: UF 2D $^{1}$H COSY,[30] adiabatic 1D $^{13}$C INEPT,[33,34] and symmetric adiabatic pure-shift $^{1}$H-$^{13}$C HSQC.[35,36] These three methods (further referred to as UF COSY, INEPT and SYMAPS HSQC) were chosen based on their potential for the profiling of complex mixtures, and on their complementarity in terms of resolution, sensitivity and throughput. Other methods could have been considered, but the choice was made to keep the overall data acquisition time below 90 min per sample, including pre-acquisition settings. As further described in the manuscript, experiment parameters were thus optimized to yield minimal experiment times for each NMR pulse sequence (10 min, 20 min and 45 min for UF COSY, INEPT and SYMAPS HSQC), offering a continuum of analytical methods that were benchmarked to a standard $^{1}$H 1D NMR experiment recorded in 5 min. A challenging set of samples was chosen to evaluate and compare their performance, stemming from a chemical ecology example aiming at characterizing the variation of the metabolome of an *Aspergillus chevalieri* fungal strain when it is placed in presence or absence of its original host, the marine sponge *Tetilla* sp, in the culture medium. Indeed, holobionts such as sponges host microbial communities capable of producing molecules involved in chemical communication and defence against predators or other microbes. The use of environmentally relevant culture conditions applied to marine-sponge-associated fungal strains is a manner of mimicking the natural holobiont system.[37–39] By tracking changes in the fungal extract profiles, it is possible to identify molecules whose production varies in response of the presence of the sponge. However, while this model raises numerous and interesting biological questions which are out of the scope of this analytical chemistry study, it also provides an impressive spectral complexity to challenge new analytical methods.

## Material and methods

### General experimental materials and chemicals

Ethyl acetate, dichloromethane and anhydrous sulfate sodium were purchased from Carlo Erba SAS (Val de Reuil, France) and distilled before use. The membrane filters used for extraction were regenerated Cellulose (RC) membrane Filters (0.45 µm, diam. 47 mm, Sartorius, Göttingen, Germany). Deuterated solvents were purchased from Eurisotop (Saint-Aubin, France) and Hexamethylcyclotrisiloxane from Sigma-Aldrich (Darmstadt, Germany).

### Cultivation of biological samples

The fungal strain was isolated from the Red Sea sponge *Tetilla* sp. harvested during Marine Science and Technology III campaign (October 1996) and identified as *Aspergillus chevalieri* on the basis of morphological features and molecular method by sequencing DNA partial sequence of β-tubulin (Genbank accession number MW729302.1). The isolated strain is deposited in the ISOMer marine fungal collection under the reference number MMS1599. *A. chevalieri* MMS1599 was cultivated on two media. The first medium, called "Control medium", was prepared by mixing DCA medium (dextrose 40 g, enzymatic digest of casein 10 g, agar 15 g)

together with reef crystal sea salts (36 g) into distilled water (1 L). The second medium, called "Sponge medium", was prepared as the Control medium, but freeze-dried *Tetilla* sponge powder, the original host (10 g/L), was added. Cultures were performed in 250 mL Erlenmeyer flasks containing 50 mL of medium. They were inoculated in 3 points by conidia from an *A. chevalieri* stock culture and incubated at 27°C with natural light variations through a glass door incubator. After 13 days, 6 culture replicates and 2 blank replicates from both Control and Sponge cultures were stopped as described in the extraction subsection. After 20 days, the 4 last replicates of Control and Sponge cultures were stopped and extracted.

**Extraction**

Fungal cultures were extracted twice using 100 mL of ethyl acetate/dichloromethane (1:1 *v*/*v*) under 30 min of sonication. Both extracts were pooled, dried with anhydrous sodium sulfate and filtered again using regenerated cellulose membrane filters, before the solvent being evaporated under vacuum to provide a crude extract.

**NMR Sample preparation**

Deuterated water was buffered by phosphoric acid ($H_2PO_4$) to reach pH = 2. Samples were prepared on the day of their analysis by dissolving 60 mg of fungal crude extract into 700 µL of the solvent mixture (deuterated chloroform/methanol/water 60:35:5 *v*/*v*/*v*). Then 7 µL of hexamethylcyclotrisiloxane solution (0.6 mM), the reference used for normalization, were added to the solubilized extract to obtain a final concentration of 6 µM into the analyzed solution. The solution was filtered on cotton to remove any particle and put into a 4-inch long NMR tube sealed by a ball.

A quality control (QC) sample were prepared to ensure that the analytical variations were much smaller than the biological variations. The QC was prepared from a DCA medium culture extract, obtained after the samples of interest, which has been prepared according to the same protocol as the samples of interest and which has been run several times in the analysis sequence. The resulting analyses of this reference sample were used to verify that this sample gave points which were grouped on the Principal Component Analysis (PCA) score plot (Supplementary Figure SI1).

**NMR acquisition pulse sequences and parameters**

**General NMR acquisition parameters.** All the spectra were automatically recorded using IconNMR (Bruker Biospin) on a 16.4 T Bruker Avance-III HD spectrometer operating at a $^1$H frequency of 700.13 MHz, equipped with an inverse $^1$H/$^{13}$C/$^{15}$N/$^2$H cryogenically cooled probe. The sample temperature was set at 293 K and a SampleJet auto-sampler set at 293 K without light, requiring 4-inches long NMR tubes in 96-well plates, was used. The pulse lengths were calibrated at the beginning of the batch, autoshim was applied during analysis of the sample cohort.

**1D $^1$H pulse sequence.** 1D $^1$H NMR spectra were recorded by applying a pulse-acquire sequence with the following acquisition parameters: 64 scans (NS) and 4 dummy scans (DS), a 20 ppm spectral width centered at 6.3 ppm, an acquisition time of 2 s and a recovery delay of 3 s. The experiment time was 5 min per spectrum.

The number of scans was adjusted to obtain a sufficient SNR (SNR > 3 on the smallest peak around 12 ppm) in a duration compatible with high-throughput screening as in metabolomics.

**1D $^{13}$C adiabatic INEPT pulse sequence.** $^{13}$C NMR adiabatic INEPT spectra were recorded using the pulse sequence described in Ref. [33] with adiabatic proton decoupling.[40] Typical acquisition parameters were as follows: $^{13}$C and $^1$H offsets were set at 80 ppm and 9 ppm, respectively, 90° $^1$H and $^{13}$C high power pulse width were 10 µs and 11 µs respectively. For INEPT acquisitions, 192 scans were accumulated with a repetition time of 5.8 s (5 s of recovery delay and 0.8 s of acquisition time). The polarization transfer delay was adjusted to 2.4 ms and the refocusing delay to 1.5 ms. Adiabatic full passage pulses were generated using Mathcad 8 (MathSoft, Inc.). They were designed with a cosine amplitude modulation of the RF field ($\omega_2^{max}$ = 157.1 kHz or 93.89 kHz for $^{13}$C or $^1$H, respectively) and an offset independent adiabaticity (OIA) by optimizing the frequency sweep ΔF (ΔF = 39 kHz or 17 kHz for $^{13}$C or $^1$H, respectively) according to the procedure published in Ref. [40]. For inversion pulses, adiabatic full passage pulses were used.[33] For refocusing pulses, composite adiabatic pulses were used.[41] The experiment time was 20 min per spectrum.

**2D UF COSY pulse sequence.** Ultrafast COSY spectra were recorded following a multiple scan pulse sequence already described in the literature,[42] completed with an interleaved acquisition scheme to increase the observable spectral width.[43] A constant-time spatial encoding scheme was used thanks to a pair of 15 ms smoothed chirp encoding pulses with a 15 kHz sweep range. For the spatial encoding, the gradient amplitude was set to ±1.6 G/cm and for the mixing period, opposite coherence-selection gradients were used (sine shape; 50.1 G/cm, 1 ms). For the echo-planar spectroscopic imaging acquisition, 32 gradients pairs were applied for 799 µs with an amplitude of 40.7 G/cm. Sixteen Free Induction Decays (FID) were recorded – with 8 scans each and a repetition delay of 5 s – while incrementing a pre-acquisition delay, then processed in an interleaved fashion to yield a COSY spectrum with a *ca.* 10 x 10 ppm apparent spectral width. The experiment time was 10 min per spectrum.

**2D SYMAPS HSQC pulse sequence.** For the symmetric adiabatic HSQC with a pure-shift acquisition (SYMAPS HSQC),[35,44] 16 scans per increment and 8 dummy scans were recorded in the indirect dimension. For the adiabatic INEPT blocks, the $^1$J coupling evolution delay was Δ = 2 ms. For inversion pulses, adiabatic full-passage pulses were used. For refocusing pulses, adiabatic composite pulses were applied,[33] these RF pulses were the same as those used for INEPT acquisitions. $^{13}$C decoupling was performed using an optimized phase cycle and adiabatic full-passage RF pulses with a cosine square amplitude modulation ($\omega_{max}^2$ = 110.6 kHz) and offset-independent adiabaticity with an optimized frequency sweep (ΔF = 14 kHz) as described in Ref. [45].

The pure shifting of the acquisition for homodecoupling the $^1$H dimension of the SYMAPS HSQC was performed with 13 chunks of 38 ms and BIRD/180° elements were used to ensure the absence $J_{HH}$

evolution during the acquisition time (492 ms).[35] The 2D HSQC maps of 11 x 65 ppm were recorded with 7568 × 128 data points.

For the NUS version of the SYMAPS HSQC experiment, 50% of the $t_1$ increments (64) were acquired. The experiment time was 45 min per spectrum.

**NMR pre-processing methods**

The NMR data processing workflow (Supplementary Figure SI2) was divided into different steps. First, spectra pre-processing was carried out on TopSpin 4.0.5 (FID apodisation, Fourier transform, phasing, baseline correction and calibration). Then bucketing was applied to 1D spectra with NMRProcflow software (INRA UMR 1332 BFP, Bordeaux Metabolomics Facility, France, https://nmrprocflow.org/)[46] and with TopSpin for 2D spectra.

**1D $^{1}$H.** The FID were processed with the Bruker TopSpin software, version 4.0.5. Before Fourier Transform, data were weighted with a trapezoidal function with a left limit at 0 and a right limit at 30% of the FID total length, then zero filled to 131072 points. The spectra underwent zero and first order manual phase correction followed by a degree 2 polynomial automatic baseline correction. The chemical shifts were finally referenced on all spectra to the hexamethylcyclotrisiloxane singlet at 0.165 ppm, which was used as an internal standard.

**$^{13}$C adiabatic INEPT.** For each $^{13}$C NMR spectrum, before Fourier Transform, data were weighted with a trapezoidal function with a left limit at 0 and a right limit at 0.4, then zero filled to 524288 points. Spectra were phased, the baseline was corrected automatically by a polynomial function of zero order. Spectra were calibrated on the $CDCl_3$ signal at 77 ppm.

**UF COSY.** The processing was carried out thanks to a home-written algorithm in TopSpin 4.0.5, as described in Ref. [47]. It uses an optimized Gaussian weighting function for spatial apodization in the spatially encoded dimension[48] and a sine-bell weighting function in the conventional dimension. Spectra were processed in magnitude mode. Due to the specificities of UF experiments, the chemical shift was finally calibrated using two diagonal signals: the 1.25 ppm and the 6.78 ppm peaks.

**SYMAPS HSQC.** Before Fourier Transform, sinusoidal apodization functions (SSB = 2) were applied in dimension F2 and F1. The data matrices were zero-filled to 16384 points in F2 and 1024 points in F1. Spectra were processed in magnitude mode. An automatic polynomial baseline correction (n 2) was applied in both dimensions. Calibration was adjusted on the hexamethylcyclotrisiloxane singlet at 0.165 ppm in F2 and 9.904 ppm in F1 (internal standard). Reconstruction of the 2D spectrum was performed using the compressed sensing algorithm (irls) within TopSpin.[49] Forward complex linear prediction was applied with 16 coefficients and 128 predicted points.

**Signal integration**

**1D $^{1}$H spectra.** After application of the pre-processing workflow in TopSpin, 1D $^{1}$H and $^{13}$C spectra were uploaded on the local installation of the software NMRProcflow.[46] An adaptative intelligent bucketing algorithm was applied on the entire pre-processed spectra in NMRProcflow. Data matrices were downloaded from NMRProcflow in a csv format and exported to Microsoft Excel for further statistical analysis.

**2D spectra.** 2D spectra were manually bucketed in TopSpin. An integration pattern was drawn and applied to all spectra. Data matrices were manually created by gathering 2D integration data and putting it together in an Excel table for further statistical analysis.

**Statistical analysis**

Data matrices from TopSpin or NMRProcflow were successively normalized i) by the reference area (buckets between -0.0036 and 0.2847 ppm for $^{1}$H, buckets between -1.008 and -0.794 ppm for adiabatic $^{13}$C INEPT, bucket between 0.4839 and -0.2594 ppm in F1 and between 0.5420 and -0.2805 ppm in F2 for UF COSY, between 13.1459 and 7.8273 ppm in F1 and between 0.2404 and 0.0373 ppm in F2 for SYMAPS HSQC), ii) by the extract mass of the replicate and iii) by the mass used to prepare the sample for NMR analysis to ensure comparability of the samples.

Statistical analyses were performed on MetaboAnalyst 5.0 (www.metaboanalyst.ca) [50,51] by using Principal Component Analysis (PCA) (Figure 3, Supplementary Figure SI3), Orthogonal Projection to Latent Structures Discriminant Analysis (OPLS-DA) (Supplementary Figures SI4, SI5, SI6, SI7) after Pareto scaling.[52] The OPLS-DA model robustness ($R^2X$, $R^2Y$, $Q^2$ of principal components) were evaluated with permutation tests (n = 2000) with $p$-value < 5 x $10^{-4}$ (Supplementary Table SI1). Variable (chemical shifts) selection, using the OPLS-DA models, was performed through the Variable Importance in the Projection (VIP) score, and the top-100 buckets were considered (Supplementary Tables SI2 to SI13).

Comparison between NMR techniques ($^{1}$H, $^{13}$C INEPT adiabatic, UF COSY, SYMAPS HSQC) was performed manually to highlight signals corresponding to the same chemical shifts (aligned signals between 1D/2D NMR).

## Results and discussion

To assess the performance and complementarity of the above-mentioned multiple pulse and multidimensional NMR methods for metabolomics, we chose a chemical ecology case study consisting of four groups of complex marine-fungal extracts. These samples were initially designed to study the biological interaction between a marine fungus, *A. chevalieri*, and its original host, the red sea sponge *Tetilla sp.* Indeed, sea sponges host complex communities of micro-organisms and are widely exploited for their ability to provide original molecules including active pharmaceutical compounds.[39] These extracts offer a great spectral complexity, well-suited to assess the ability of NMR methods to separate overlapped spectral fingerprints. Note that the chemical characterization of these extracts, whose diversity remains largely unknown, is out of the scope of this analytical study, as well as the subsequent spectral annotation. This work has an analytical chemistry goal which consists in characterizing the ability of the NMR methods to highlight

statistically relevant spectral features, and their complementarity with conventional $^{1}$H NMR.

**$^{1}$H NMR spectral fingerprints**

Figure 1 illustrates such spectral complexity by showing the 1D $^{1}$H spectra for 4 different extracts pertaining to different sample groups. Control and Sponge samples correspond respectively to extracts from *A. chevalieri* cultivated without or with the presence of *Tetilla* sp. These two conditions were monitored at two different time points (T1 = 13 days and T2 =20 days) in order to increase the chemical complexity of extracts and to challenge our NMR methods, as it has been shown that the metabolic richness of fungal extracts is a function of the incubation time.[53] Extracts were highly concentrated (just below the solubility limit of extracts), therefore $^{1}$H NMR spectra were acquired in only 5 min at high field (700 MHz), a duration compatible with high-throughput metabolomics.

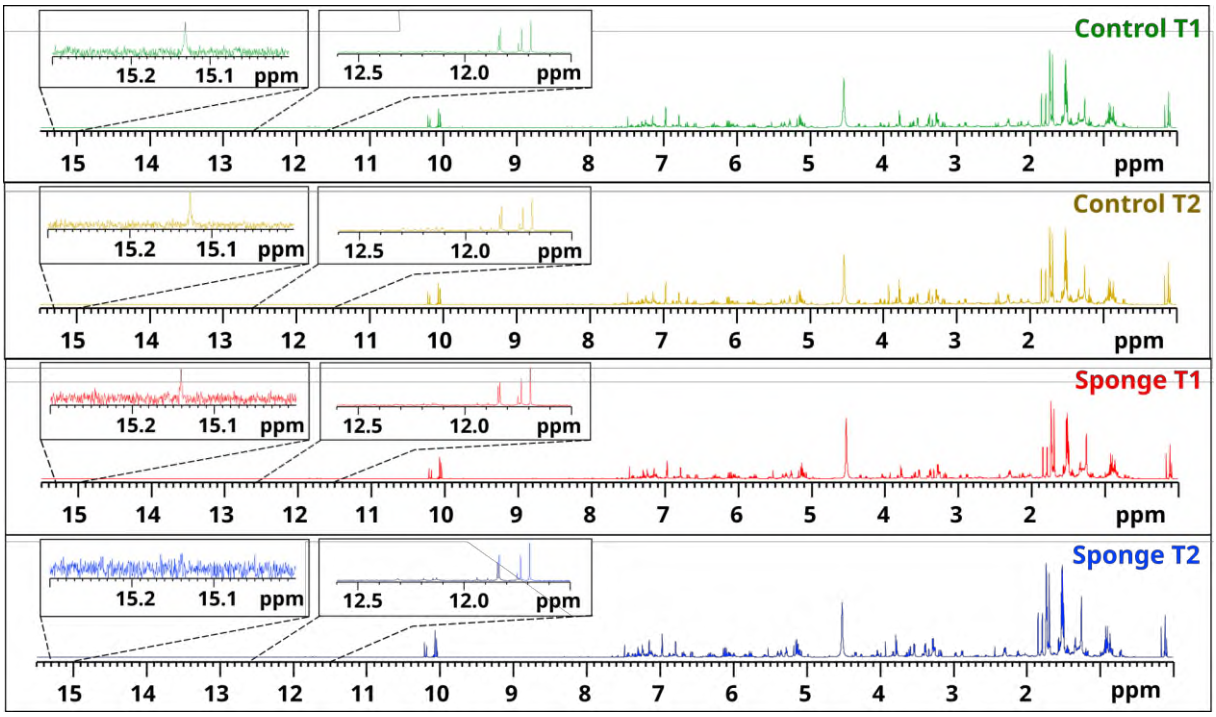


**Figure 1.** $^{1}$H NMR spectra of 4 representative extracts from *Aspergillus chevalieri* cultivated at two different time points (T1 = 13 days and T2 = 20 days), without (Control) or with (Sponge) the presence of *Tetilla* sp. Spectra were acquired in 5 min at 293 K on a 700 MHz spectrometer equipped with a cryogenically cooled probe.

The spectra presented in Figure 1 show a continuous and complex spectral fingerprint, exhibiting numerous major signals between 1 and 8 ppm, but also smaller peaks at high chemical shifts. A noticeable feature of these spectra is the diversity of signal intensities (Supplementary Table 2): the signal-to-noise ratio (SNR) is lower than 10 for 8% of the signals, while the maximum SNR reaches 24,000, illustrating the challenged posed by such samples to the dynamic range of the NMR receiver. It should be noted that temperature, pH and NMR solvents were finely tuned to obtain the sharpest possible peaks, resulting in a mixture of NMR solvents ($CDCl_3$/MeOD/$D_2O$ 60:35:5 at pH = 2). We chose not to suppress NMR signals from residual protonated solvents to avoid impacting nearby peaks of interest.

The comparison of spectra from the 4 sample groups in Figure 1 did not reveal any striking difference between the culture times, thus motivating further statistical analyses. The continuous spectral fingerprint, associated with the unavoidable chemical shift variations between samples, justified the need for adaptative intelligent bucketing procedures.[46] The bucketing of $^{1}$H spectra performed with such an intelligent bucketing approach included in the NMRProcFlow software, resulted in 901 buckets spread over the entire spectrum (Supplementary Table SI14 and Figure SI8). The majority of these buckets (> 90%) had an SNR higher than 10, highlighting the good sensitivity of the $^{1}$H approach.

**NMR methods to improve signal dispersion**

The challenging complexity encountered in 1D $^{1}$H NMR justified the need for NMR methods that could better separate overlapped signals while maintaining a reasonable experiment duration. Figure 2 shows how recent multiple pulse and/or multidimensional methods have the potential to provide such separation. The three methods evaluated in this study were chosen because they were specifically tailored, in previous studies, to yield an optimal analytical performance in terms of repeatability – a crucial feature for inter-sample comparison in metabolomics. In this study, they are benchmarked to the conventional 1D $^{1}$H NMR approach presented above.

Adiabatic $^{13}$C INEPT (Figure 2A) relies on polarization transfer from $^{1}$H to $^{13}$C to benefit from the greater spectra dispersion of $^{13}$C

spectroscopy. It provides an improved sensitivity compared to direct $^{13}C$ spectra, and an optimum repeatability thanks to optimized adiabatic pulses.[54] The larger $^{13}C$ spectral range, associated with the singlet character of all the peaks thanks to $^{1}H$ decoupling, considerably reduces peak overlap compared to $^{1}H$ NMR, as visible in Figure 2A. The repetition time between two scans is dictated by $^{1}H$ relaxation, which makes it possible to preserve a reasonable experiment time (20 min here). However, the polarization transfer mechanism involved in this pulse sequence makes quaternary carbon signals invisible on INEPT spectra, which is not necessarily critical assuming that the overwhelming majority of metabolites will involve at least one protonated carbon atom. The intelligent bucketing of the $^{13}C$ spectrum resulted in 411 buckets (Supplementary Table SI15 and Figure SI9) versus 901 for $^{1}H$, which was consistent with the greater sparsity of the $^{13}C$ spectrum. However, a significant proportion (approx. 60%) of those buckets had an SNR lower than 10, which could be related to the low sensitivity of the experiment relying on the natural abundance of $^{13}C$ nuclei (1.1%).

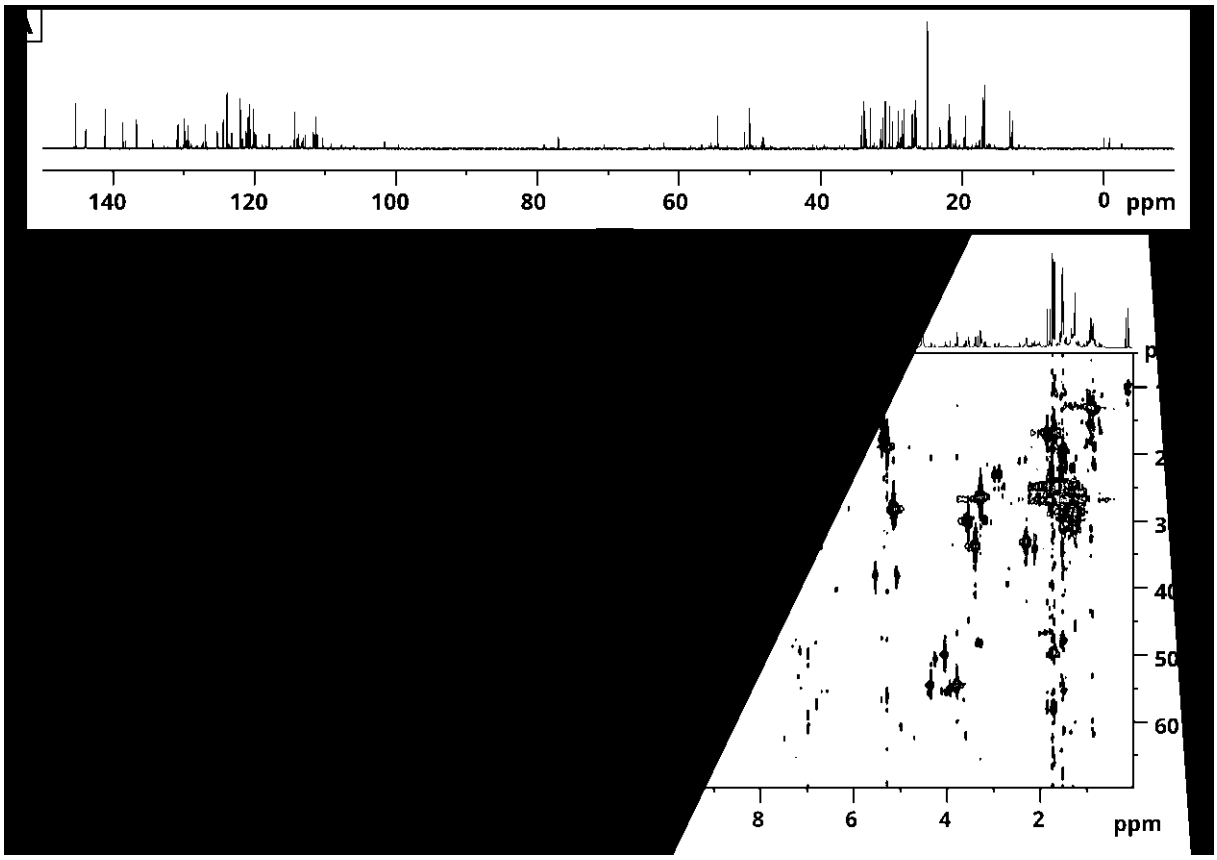


**Figure 2.** $^{13}C$ adiabatic INEPT (A), UF COSY (B) and SYMAPS HSQC (C) NMR spectra of an *Aspergillus chevalieri* extract, acquired in 20 min (A), 10 min (B) and 45 min (C) at 293 K on a 700 MHz spectrometer equipped with a cryogenically cooled probe.

Ultrafast COSY (Figure 2B) spreads $^{1}H$ resonances over two orthogonal dimensions while providing information on $^{1}H$-$^{1}H$ through-bond connectivities. UF COSY, which relies on a spatial encoding approach derived from imaging, was shown to be more repeatable than conventional COSY for the analysis of metabolic mixtures.[42] Although slightly less sensitive than conventional $^{1}H$ spectroscopy, UF COSY provides spectra devoid of $t_1$ noise contrary to conventional COSY, which makes it a valuable tool for the analysis of samples with a great diversity of metabolite concentrations. When samples are concentrated enough, UF COSY provides a rich metabolic fingerprint in a short time (10 min here), with an improved spectral dispersion. UF COSY also has specific features, such as the need to compromise between spectral width and resolution[47] – a compromise that was optimized here thanks to the use of an interleaved acquisition approach. This explains why the acquisition of COSY spectra was limited to the 0-8 ppm spectral range in this study. Attempts to record UF COSY spectra over a full spectral width showed no detectable signal in the upper chemical shift range, while deteriorating the resolution in the 0-8 ppm range. The bucketing of the UF COSY spectra was performed manually, because no tool was found adapted to the bucketing of such 2D NMR metabolomics datasets with a diversity of peak patterns and shapes, and residual artefacts associated with the interleaving procedures would have been challenging to identify for an automated algorithm. Manual bucketing resulted in 1056 buckets spread across the entire spectrum (Supplementary Figures SI10 and SI11).

SYMAPS HSQC (Figure 2C) offers an even higher level of spectral dispersion by spreading the $^{1}H$ peaks along an orthogonal $^{13}C$

dimension. It shares similarities with INEPT, such as the polarization transfer mechanism that prevents the detection of quaternary carbons and the $^{1}$H-dependent repetition times, but it also has an increased sensitivity compared to INEPT since $^{1}$H are both the excited and detected nuclei in the pulse sequence. SYMAPS HSQC relies on a conventional 2D acquisition scheme which requires a longer experiment time. However, the pulse sequence was here accelerated by relying on non-uniform sampling (NUS), spectral aliasing and variable recycling time (VRT); resulting in a total experiment time of 45 min. Variable recycling time consists in a progressive reduction of the recycling time throughout successive $t_1$ increments.[35,55] Both sensitivity and resolution were improved by a pure-shift acquisition procedure in the $F_2$ dimension, turning all $^{1}$H multiplets into singlets. As for the INEPT pulse sequence, the use of adiabatic pulses was important to reduce variability sources, hence leading to an optimized repeatability. Note that the use of spectral aliasing may require additional care in the interpretation of spectral data,[56] although we ensured that aliased spectral regions did not overlap (Supplementary Figure SI12). As for UF COSY, the bucketing was performed manually, yielding a number of buckets (441) similar to INEPT and spread over the entire spectrum (Supplementary Figures SI13 and SI14).

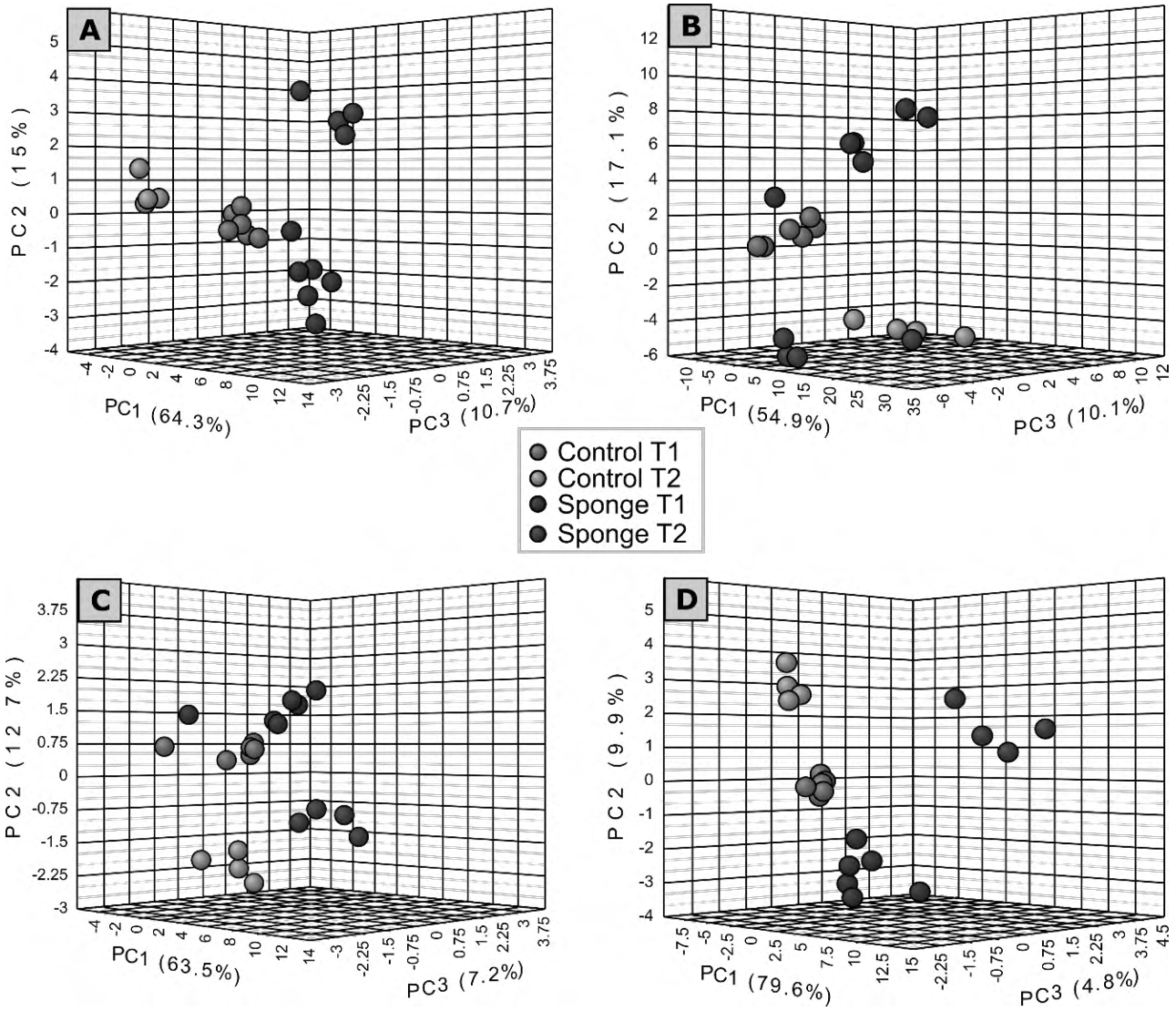


**Figure 3.** 3D PCA score plots obtained from NMR data acquired on *Aspergillus chevalieri* extracts cultivated at two different time points (T1 = 13 days and T2 = 20 days), without (Control) or with (Sponge) the presence of *Tetilla* sp. Data obtained from (A) $^{1}$H spectra, (B) adiabatic INEPT spectra, (C) UF COSY spectra and (D) SYMAPS HSQC spectra. Color code is as follows: Green: Control T1; Yellow: Control T2; Red: Sponge T1; Blue: Sponge T2.

## Unsupervised statistical analysis

To assess the classification performance of the above-mentioned NMR methods, data were first subjected to a conventional NMR metabolomics processing workflow (Supplementary Figure SI2). As mentioned above, 1D $^{1}$H and $^{13}$C INEPT spectra were subjected to an intelligent bucketing approach, while bucketing was performed manually for 2D spectra. The processing workflow deliberately did not include any spectral alignment procedure (which would have been available only for 1D data), to avoid an additional and tedious

processing step and to facilitate the comparison between the four methods. All the data matrices were subjected to a triple normalization procedure to i) the NMR internal reference compound, ii) the mass of each replicate and iii) the mass used to prepare the NMR tube. Normalized data were subjected to Pareto scaling[52] and analyzed with MetaboAnalyst.

Figure 3 shows PCA score plots obtained with the four NMR methods. The choice was made to represent 3D score plots, since 2D plots showed apparent superposition between groups, making it more difficult to compare methods (Supplementary Figure SI3). The 3D PCA score plots clearly highlight the ability of the four methods to efficiently separate the four sample groups. The sum of variances expressed on the three first components of the PCA was high in all cases (>80%). The score plots also highlighted the relatively high intra-group variability, which seemed more pronounced for the sponge groups. Moreover, the time factor (T1 *vs* T2) separated the groups slightly better than the medium factor (Control *vs* Sponge).

**Supervised statistical analysis**

To further evaluate the performance and complementarity of the four NMR methods, we chose to apply supervised statistical analysis through OPLS-DA (Orthogonal Projection to Latent Structures Discriminant Analysis) traditionally used in metabolomics studies,[57,58] focusing on the medium effect (Control *vs* Sponge). Score plots obtained for the four methods are shown in Supplementary Figures SI4 to SI7, with associated model parameters in Supplementary Table SI1. Model parameters showed that all models were reliable ($p$-value ≤ 5 %, $Q^2 \geq 0.5$ and absence of overfitting – $Q^2 < R^2Y$),[59] with SYMAPS HSQC spectra containing more significant signals as revealed by the highest explained variance ($R^2X$). The comparison of $Q^2$ values showed that models obtained from homonuclear data (UF COSY) provided a more robust classification than those built from heteronuclear data (HSQC).

To investigate the complementarity and the potential redundancy of the spectroscopic information provided by the four NMR methods, we performed a detailed study of the 100 most discriminant variables (VIPs, determined by their Variable Importance in the Projection score) contributing to the OPLS-DA models. Figure 4 provides a summary of this investigation (a detailed list of the VIPs is provided in Supplementary Tables SI2-13). Reading Figure 4 makes it possible to understand how many buckets are shared between two methods, and how many are specific of each method.

Let us first compare the results obtained in homonuclear 2D NMR (UF COSY) versus $^1$H NMR. Figure 4B (example case h) shows that 16 VIP buckets were specific to UF COSY and were not present among the $^1$H NMR 100 most discriminant buckets, nor in the 100 $^1$H VIP buckets that follow, excluding a threshold effect. A detailed analysis of the spectral data showed that for most of them, these 16 buckets corresponded to signals in heavily crowded regions that were too heavily overlapped in 1D NMR to be detected. Specific examples are highlighted in Figure 5. These results clearly highlight the benefit of using 2D NMR in metabolomics workflow, as the additional dimension makes it possible to reveal discriminating signals that were invisible in 1D NMR.

Conversely, Figure 4A also highlights that 18 of the 100 most discriminant VIP buckets in 1D $^1$H NMR did not show any correspondence in UF COSY (example case d). A detailed analysis showed that these buckets either corresponded to signals whose intensity is below the limit of detection of UF COSY, while others lied outside of the detected 0-8 ppm spectral range, as could be expected from the known limitations of UF COSY. Overall, UF COSY appears complementary to 1D $^1$H spectroscopy, with a better ability to highlight relevant signals in crowded regions, while 1D $^1$H spectroscopy is best suited to detect signals close to the limit of detection of NMR. Moreover, UF COSY shows that homonuclear 2D spectra remain quite overlapped in some regions –which can be problematic when trying to identify discriminant signals– justifying the need for heteronuclear techniques.

The results obtained for SYMAPS HSQC (Figure 4E) showed that 9 among the 100 most important VIP buckets did not have any correspondence with a 1D technique (example case o). Corresponding buckets were present in the VIP list of 1D $^1$H data, but beyond the 100 first ones (threshold effect), while they were not present in the adiabatic $^{13}$C INEPT data, owing to a too low SNR. Moreover, 33 among the 100 most important HSQC buckets showed a correspondence in $^{13}$C INEPT data only (example case p). As for UF COSY, all these buckets corresponded to buckets in regions which were too heavily overlapped to be discriminated based on 1D $^1$H data alone. Finally, 9 VIP buckets were present in both HSQC and 1D $^1$H most important buckets, but not in the $^{13}$C INEPT data (example case m). These buckets corresponded to signals with a medium-range SNR that can be detected with HSQC but not with INEPT, highlighting the better sensitivity of 2D HSQC among heteronuclear methods. Still, data analysis for adiabatic INEPT (Figure 4D) showed that 9 among the 100 most important VIPs for this method were not reflected in HSQC data (example case l). These buckets corresponded to signals in crowded regions of HSQC. In such cases, the high resolution of INEPT (which is better than the $^{13}$C resolution of HSQC in the $^{13}$C dimension) helped to better separate overlapped signals.

Finally, as for UF COSY, 1D $^1$H NMR showed ability to highlight signals that could not be detected with heteronuclear 2D NMR (example case j – Fig. 4C). In most cases, these buckets corresponded to signals with low intensity, that were below the limit of detection of HSQC at natural $^{13}$C abundance. Overall, HSQC buckets that appeared among the 100 most discriminant VIPs were those with a sufficient SNR. While less sensitive than 1D $^1$H spectroscopy, HSQC provides a greater peak dispersion at the cost of a *ca.* 100 times lower sensitivity per unit time. In cases where sensitivity is sufficient and where peak overlap is important, the excellent resolution of adiabatic INEPT can allow the detection of additional relevant signals.

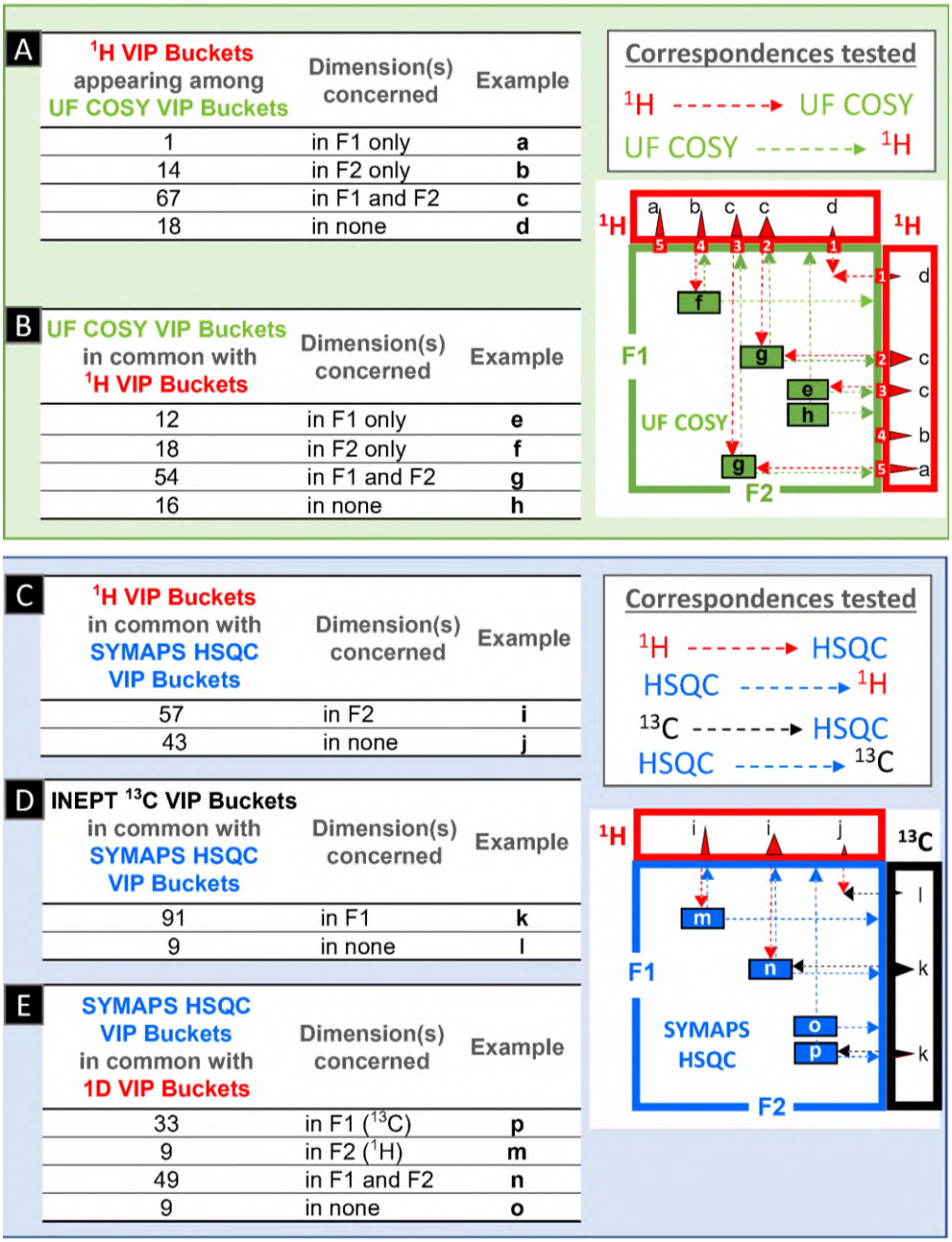


**Figure 4.** Number of VIP buckets (among the 100 most discriminant VIPs) in common between the OPLS-DA models built from the four different NMR methods ($^{1}$H, adiabatic $^{13}$C INEPT, UF COSY, SYMAPS HSQC). **(A)** VIP from $^{1}$H spectra in common with the F2 and/or F1 dimensions of UF COSY; **(B)** VIP from the F2 and/or F1 dimensions of UF COSY spectra in common with $^{1}$H spectra; **(C)** VIP from $^{1}$H spectra in common with the F2 dimension of SYMAPS HSQC; **(D)** VIP from adiabatic $^{13}$C INEPT in common with the F1 dimension of SYMAPS HSQC; **(E)** VIP from SYMAPS HSQC spectra in common with 1D $^{1}$H or $^{13}$C adiabatic INEPT spectra. Small letters in the "example" column refer to the situations reported in the spectra placed on the right side of the figure. As an example, Figure 4E should be understood as follows: among the 100 most discriminant VIP buckets of the SYMAPS HSQC model, 33 of them also appear in the 100 most discriminant VIPs of the 1D adiabatic $^{13}$C INEPT spectrum, 9 also appear in the 100 most discriminant VIPs of the 1D $^{1}$H spectrum, 49 appear in both and 9 in none (meaning that these VIPs are specific to the SYMAPS HSQC model). The correspondence between the different methods was made based on the chemical shifts of the buckets.

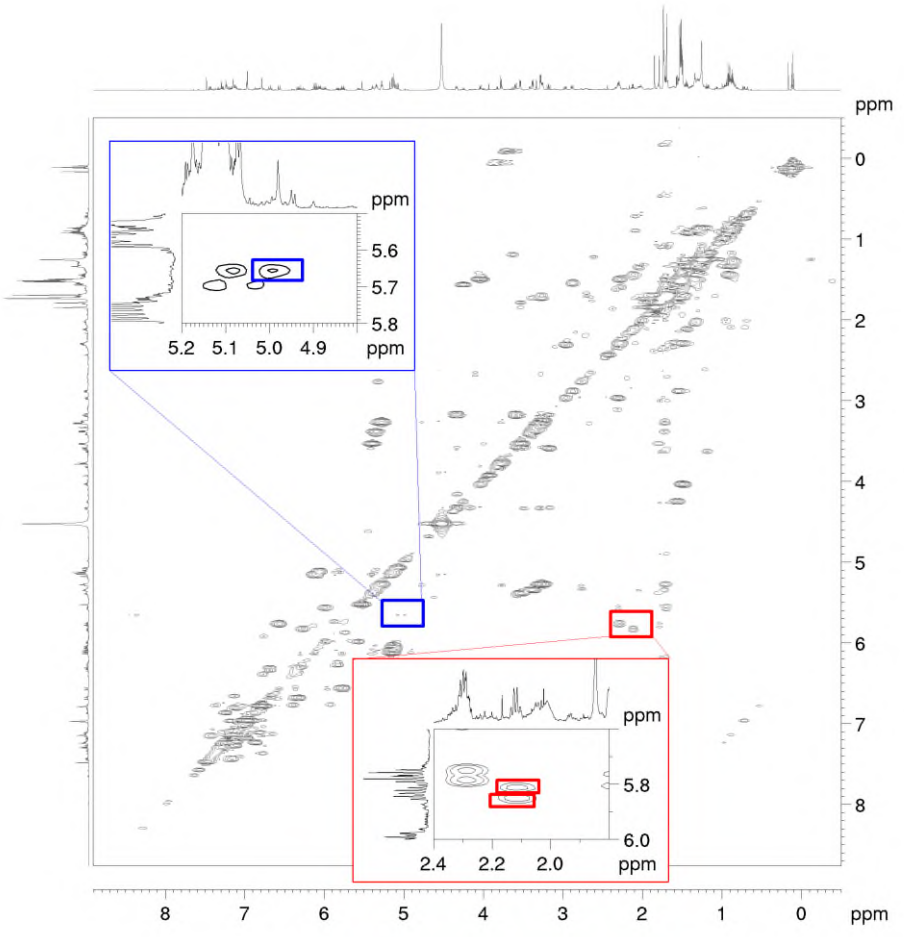


**Figure 5.** Example of discriminant buckets that are highlighted among the first 100 VIPs of UF COSY data but are not reflected in $^{1}$H data due to strong peak overlaps. Spectrum recorded in 10 min on an *Aspergillus chevalieri* extract belonging to the Sponge T1 group, on a 700 MHz spectrometer equipped with a cryogenically cooled probe.

## Conclusions

The results described in this chemical ecology case study highlight the performance and the complementarity of advanced NMR methods for untargeted metabolomics of extracts, in particular in such a complex chemical composition. First, all the three methods show a great classification ability equivalent to 1D $^{1}$H NMR with a reasonable experiment time that remains compatible with high-throughput data acquisitions. Classification results highlight that UF COSY, adiabatic INEPT and SYMAPS HSQC have reached a sufficient level of maturity and robustness to be routinely incorporated in metabolomics workflows. Second, a detailed analysis of discriminating signals upon supervised statistical analysis reveals that each method has an interesting degree of specificity. Indeed, multi-pulse and multi-dimensional methods, thanks to their ability to separate overlapped resonances, have the potential to highlight relevant signals in regions which are too crowded to be efficiently exploited in 1D $^{1}$H NMR. Conversely, these advanced methods should be seen as complementary to 1D $^{1}$H NMR, since the latter retains a major sensitivity advantage. Therefore, one could easily figure out the relevance of combining several techniques in a single untargeted NMR metabolomics workflow, to maximize the accessible metabolic information. Although part of the information is redundant between the different methods, such redundancy should be an asset in the determination of potential biomarkers from NMR data. The high-throughput character of the NMR methods that were employed is an important characteristic towards such applications, since the results described above would not have been obtained with conventional, time-consuming, 2D NMR methods.

This study also raises further practical questions for further application of multinuclear and multi-dimensional methods in metabolomics workflows. Future work will consider the implementation and use of more automated tools for the bucketing of 2D datasets.[60] Indeed, manual bucketing may complicate the routine implementation of such methods. Peak matching methods may also help improving the quality of 2D data analysis.[61] Moreover, the correlation between buckets from different datasets (which was done manually here) could benefit from approaches capable of integrating NMR datasets of different nature (1D and 2D, $^{1}$H and $^{13}$C) in an automated way. These would include multi-block statistics[62,63] to combine data matrices from different datasets in a single model, but also the use of correlation tools such as STOCSY [64] or

dereplication approaches such as MADByTE[65] to better reveal synergies between multiple datasets. Such perspectives will be investigated in future studies.

## Author contributions

**Aurore Michaud**: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Writing - original draft; **Samuel Bertrand**: Conceptualization; Data curation; Formal analysis; Writing - review & editing; **Serge Akoka**: Conceptualization; Methodology; Investigation; Writing - review & editing; **Jonathan Farjon**: Methodology; Investigation; Writing - review & editing; **Estelle Martineau**: Investigation; Writing - review & editing; **Nicolas Ruiz**: Investigation; Writing - review & editing; **Thibaut Robiou Du Pont**: Investigation; **Olivier Grovel**: Conceptualization; Funding acquisition; Methodology; Project administration; Supervision; Writing - review & editing; **Patrick Giraudeau**: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Supervision; Writing - original draft

## Conflicts of interest

There are no conflicts to declare.

## Data availability

Data supporting this article have been included as part of the Supplementary Information, which includes full details of the 100 first VIPs for each method. All spectral data used for this manuscript has been deposited online at https://doi.org/10.57745/N87KTD. Programs to implement and process ultrafast COSY experiments are available from https://madoc.univ-nantes.fr/course/view.php?id=24710&section=1

## Acknowledgements

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement SUMMIT no. 814747). The authors acknowledge the French National Infrastructure for Metabolomics and Fluxomics MetaboHUB-ANR-11-INBS-0010 (www.metabohub.fr) and the Corsaire metabolomics core facility (Biogenouest). This work includes NMR experiments carried out on the CEISAM NMR platform.

## References

1 J.-L. Wolfender, J.-M. Nuzillard, J. J. J. van der Hooft, J.-H. Renault and S. Bertrand, *Anal. Chem.*, 2019, **91**, 704–742.
2 J. Wist, *Magn. Reson. Chem.*, 2017, **55**, 22–28.
3 R. M. Borges, G. de A. Ferreira, M. M. Campos, A. M. Teixeira, F. das N. Costa, F. O. das Chagas and M. Colonna, *Phytochem. Anal.*, 2023, **34**, 385–392.
4 J.-N. Dumez, *Chem. Commun.*, 2022, **58**, 13855–13872.
5 A.-H. Emwas, R. Roy, T. R. McKay, L. Tenori, E. Saccenti, A. N. G. Gowda, D. Raftery, F. Alahmari, L. Jaremko, M. Jaremko and S. D. Wishart, *Metabolites*, 2019, **9**, 123.
6 P. G. Takis, V. Ghini, L. Tenori, P. Turano and C. Luchinat, *TrAC Trends Anal. Chem.*, 2019, **120**, 115300.
7 M. G. Fernand, C. Roullier, Y. Guitton, J. Lalande, S. Lacoste, J. Dupont, N. Ruiz, Y. F. Pouchus, C. Raheriniaina and E. Ranaivoson, *Aquaculture*, 2017, **479**, 750–758.
8 M. Halabalaki, S. Bertrand, A. Stefanou, K. Gindro, S. Kostidis, E. Mikros, L. A. Skaltsounis and J.-L. Wolfender, *Phytochem. Anal.*, 2014, **25**, 350–356.
9 P. Giraudeau, V. Silvestre and S. Akoka, *Metabolomics*, 2015, **11**, 1041–1055.
10 S. Goulitquer, M. Croyal, J. Lalande, A.-L. Royer, Y. Guitton, D. Arzur, S. Durand, C. Le Jossic-Corcos, A. Bouchereau, P. Potin, S. Akoka, J.-P. Antignac, M. Krempf, V. Ferchaud-Roucher, P. Giraudeau and L. Corcos, *Cell. Death. Dis.*, 2018, **9**, 745.
11 E. K. Nawrocka, M. Urbańczyk, K. Koziński and K. Kazimierczuk, *RSC Adv.*, 2021, **11**, 35321–35325.
12 N. Harvey, P. G. Takis, J. C. Lindon, J. V. Li and B. Jiménez, *Anal. Chem.*, 2023, **95**, 3147–3152.
13 V. Ribay, A. Dey, B. Charrier, C. Praud, J. Mandral, J.-N. Dumez, M. P. M. Letertre and P. Giraudeau, *Angew. Chem. Int. Ed.*, 2023, **62**, e202302110.
14 Z. Wang, S. Pisano, V. Ghini, P. Kadeřávek, M. Zachrdla, P. Pelupessy, M. Kazmierczak, T. Marquardsen, J.-M. Tyburn, G. Bouvignies, G. Parigi, C. Luchinat and F. Ferrage, *J. Am. Chem. Soc.*, 2021, **143**, 9393–9404.
15 C. S. Clendinen, B. Lee-McMullen, C. M. Williams, G. S. Stupp, K. Vandenborne, D. A. Hahn, G. A. Walter and A. S. Edison, *Anal. Chem.*, 2014, **86**, 9242–9250.
16 N. Merchak, T. Rizk, V. Silvestre, G. S. Remaud, J. Bejjani and S. Akoka, *Food Chem.*, 2018, **245**, 717–723.
17 A. Dey, B. Charrier, E. Martineau, C. Deborde, E. Gandriau, A. Moing, D. Jacob, D. Eshchenko, M. Schnell, R. Melzi, D. Kurzbach, M. Ceillier, Q. Chappuis, S. F. Cousin, J. G. Kempf, S. Jannin, J.-N. Dumez and P. Giraudeau, *Anal. Chem.*, 2020, **92**, 14867–14871.
18 J. M. Lopez, R. Cabrera and H. Maruenda, *Sci Rep*, 2019, **9**, 6900.
19 X. Chen, C. Caradeuc, A. Montagne, V. Baud, G. Bertho, C. Lucas-Torres and N. Giraud, *Anal. Chem.*, 2022, **94**, 14974–14984.
20 J. Marchand, E. Martineau, Y. Guitton, G. Dervilly-Pinel and P. Giraudeau, *Curr. Op. Biotechnol.*, 2017, **43**, 49–55.
21 E. Martineau, J.-N. Dumez and P. Giraudeau, *Magn. Reson. Chem.*, 2020, **58**, 390–403.
22 E. A. Mahrous and M. A. Farag, *J Adv Res*, 2015, **6**, 3–15.
23 K. Bingol, D.-W. Li, B. Zhang and R. Brüschweiler, *Anal. Chem.*, 2016, **88**, 12411–12418.
24 Q. N. Van, H. J. Issaq, Q. Jiang, Q. Li, G. M. Muschik, T. J. Waybright, H. Lou, M. Dean, J. Uitto and T. D. Veenstra, *J. Proteome Res.*, 2008, **7**, 630–639.
25 E. Martineau and P. Giraudeau, *Methods Mol. Biol.*, 2019, 365–383.
26 S. Watermann, C. Schmitt, T. Schneider and T. Hackl, *Metabolites*, 2021, **11**, 39.
27 M. Jeeves, J. Roberts and C. Ludwig, *Magn. Reson. Chem.*, 2021, **59**, 287–299.
28 J. Marchand, E. Martineau, Y. Guitton, B. Le Bizec, G. Dervilly-Pinel and P. Giraudeau, *Metabolomics*, 2018, **14**, 60.
29 B. Féraud, E. Martineau, J. Leenders, B. Govaerts, P. de Tullio and P. Giraudeau, *Metabolomics*, 2020, **16**, 42.
30 T. Jézéquel, C. Deborde, M. Maucourt, V. Zhendre, A. . Moing and P. Giraudeau, *Metabolomics*, 2015, **11**, 1231–1242.
31 P. Giraudeau, S. Massou, Y. Robin, E. Cahoreau, J.-C. Portais and S. Akoka, *Anal. Chem.*, 2011, **83**, 3112–3119.
32 N. Merchak, V. Silvestre, L. Rouger, P. Giraudeau, T. Rizk, J. Bejjani and S. Akoka, *Talanta*, 2016, **156–157**, 239–244.
33 C. Thibaudeau, G. Remaud, V. Silvestre and S. Akoka, *Anal. Chem.*, 2010, **2010**, 5582–5590.
34 N. Merchak, J. Bejjani, T. Rizk, V. Silvestre, G. S. Remaud and S. Akoka, *Anal. Methods*, 2015, **7**, 4889–4891.
35 J. Farjon, C. Milande, E. Martineau, S. Akoka and P. Giraudeau, *Anal. Chem.*, 2018, **90**, 1845–1851.
36 I. Timári, C. Wang, A. L. Hansen, G. Costa dos Santos, S. O. Yoon, L. Bruschweiler-Li and R. Brüschweiler, *Anal. Chem.*, 2019, **91**, 2304–2311.
37 M. K. Wooster, O. Voigt, D. Erpenbeck, G. Wörheide and M. L. Berumen, in *Coral Reefs of the Red Sea*, eds. C. R. Voolstra and M. L. Berumen, Springer International Publishing, Cham, 2019, pp. 91–122.

38 V. J. Paul, C. J. Freeman and V. Agarwal, *Integr. Comp. Biol.*, 2019, **59**, 765–776.
39 D. Varijakzhan, J.-Y. Loh, W.-S. Yap, K. Yusoff, R. Seboussi, S.-H. E. Lim, K.-S. Lai and C.-M. Chong, *Mar. Drugs*.
40 E. Tenailleau and S. Akoka, *J Magn Reson*, 2007, **185**, 50–58.
41 T.-L. Hwang, P. C. M. Van Zijl and M. Garwood, *J. Magn. Reson.*, 1997, **124**, 250–254.
42 A. Le Guennec, I. Tea, I. Antheaume, E. Martineau, B. Charrier, M. Pathan, S. Akoka and P. Giraudeau, *Anal. Chem.*, 2012, **84**, 10831–10837.
43 L. Rouger, B. Charrier, M. Pathan, S. Akoka and P. Giraudeau, *J. Magn. Reson.*, 2014, **238**, 87–93.
44 L. Haddad, S. Renou, G. S. Remaud, T. Rizk, J. Bejjani and S. Akoka, *Anal. Bioanal. Chem.*, 2021, **413**, 1521–1532.
45 E. Tenailleau and S. Akoka, *J Magne Reson*, 2007, **185**, 50–58.
46 D. Jacob, C. Deborde, M. Lefebvre, M. Maucourt and A. Moing, *Metabolomics*, 2017, **13**, 36.
47 C. Lhoste, B. Lorandel, C. Praud, A. Marchand, R. Mishra, A. Dey, A. Bernard, J.-N. Dumez and P. Giraudeau, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2022, **130–131**, 1–46.
48 P. Giraudeau and S. Akoka, *Magn. Reson. Chem.*, 2011, **49**, 307–313.
49 A. S. Stern and J. C. Hoch, *Magn. Reson. Chem.*, 2015, **53**, 908–912.
50 Z. Pang, G. Zhou, J. Ewald, L. Chang, O. Hacariz, N. Basu and J. Xia, *Nat. Protoc.*, 2022, **17**, 1735–1761.
51 Z. Pang, J. Chong, G. Zhou, D. A. de Lima Morais, L. Chang, M. Barrette, C. Gauthier, P.-É. Jacques, S. Li and J. Xia, *Nucleic Acids Res.*, 2021, **49**, W388–W396.
52 R. van den Berg, H. C. J. Hoefsloot, J. A. Westerhuis, A. K. Smilde and M. J. van der Werf, *BMC Genomics*, , DOI:10.1186/1471-2164-7-142.
53 C. Roullier, S. Bertrand, E. Blanchet, M. Peigné, T. Robiou du Pont, Y. Guitton, Y. F. Pouchus and O. Grovel, *Mar. Drugs*, , DOI:10.3390/md14050103.
54 U. Bussy, C. Thibaudeau, F. Thomas, J.-R. Desmurs, E. Jamin, G. Remaud, V. Silvestre and S. Akoka, *Talanta*, 2011, **85**, 1909–1914.
55 S. Macura, *J. Am. Chem. Soc.*, 2009, **131**, 9606–9607.
56 D. Jeannerat, *Magn Reson Chem*, 2003, **41**, 3–17.
57 M. Bylesjö, M. Rantalainen, O. Cloarec, J. K. Nicholson, E. Holmes and J. Trygg, *J. Chemom.*, 2006, **20**, 341–351.
58 P. S. Gromski, H. Muhamadali, D. I. Ellis, Y. Xu, E. Correa, M. L. Turner and R. Goodacre, *Anal. Chim. Acta*, 2015, **879**, 10–23.
59 L. Eriksson, J. Trygg and S. Wold, *J. Chemom.*, 2008, **22**, 594–600.
60 D.-W. Li, A. L. Hansen, C. Yuan, L. Bruschweiler-Li and R. Brüschweiler, *Nat. Commun.*, 2021, **12**, 5229.
61 H. Du, X. Gu, J. Chen, C. Bai, X. Duan and K. Hu, *Anal. Chem.*, 2023, **95**, 3195–3203.
62 V. Cariou, E. M. Qannari, D. N. Rutledge and E. Vigneau, *Sensometrics 2016 Sensometrics---Sea*, 2018, **67**, 27–34.
63 C. Palaric, S. Pilard, J.-X. Fontaine, J. Boccard, D. Mathiron, S. Rigaud, D. Cailleu, F. Mesnard, Y. Gut, T. Renaud, A. Petit, J.-Y. Beaumal and R. Molinié, *Metabolomics*, 2019, **15**, 107.
64 O. Cloarec, M.-E. Dumas, A. Craig, R. H. Barton, J. Trygg, J. Hudson, C. Blancher, D. Gauguier, J. C. Lindon, E. Holmes and J. Nicholson, *Anal. Chem.*, 2005, **77**, 1282–1289.
65 L. Flores-Bocanegra, Z. Y. Al Subeh, J. M. Egan, T. El-Elimat, H. A. Raja, J. E. Burdette, C. J. Pearce, R. G. Linington and N. H. Oberlies, *J. Nat. Prod.*, 2022, **85**, 614–624.

# SUPPORTING INFORMATION

# Exploring the complementarity of fast multipulse and multidimensional NMR methods for metabolomics: a chemical ecology case study

Aurore Michaud[1,2], Samuel Bertrand[2,3], Serge Akoka[1], Jonathan Farjon[1], Estelle Martineau[1], Nicolas Ruiz[2], Thibaut Robiou Du Pont[2], Olivier Grovel[2**], Patrick Giraudeau[1*]

[1]Nantes Université, CNRS, CEISAM UMR 6230, F-44000 Nantes, France.
*Email: patrick.giraudeau@univ-nantes.fr
[2]Nantes Université, Institut des Substances et Organismes de la Mer, ISOMer, UR 2160, F-44000 Nantes, France
**Email: olivier.grovel@univ-nantes.fr
[3]Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France

# Figures



# Tables

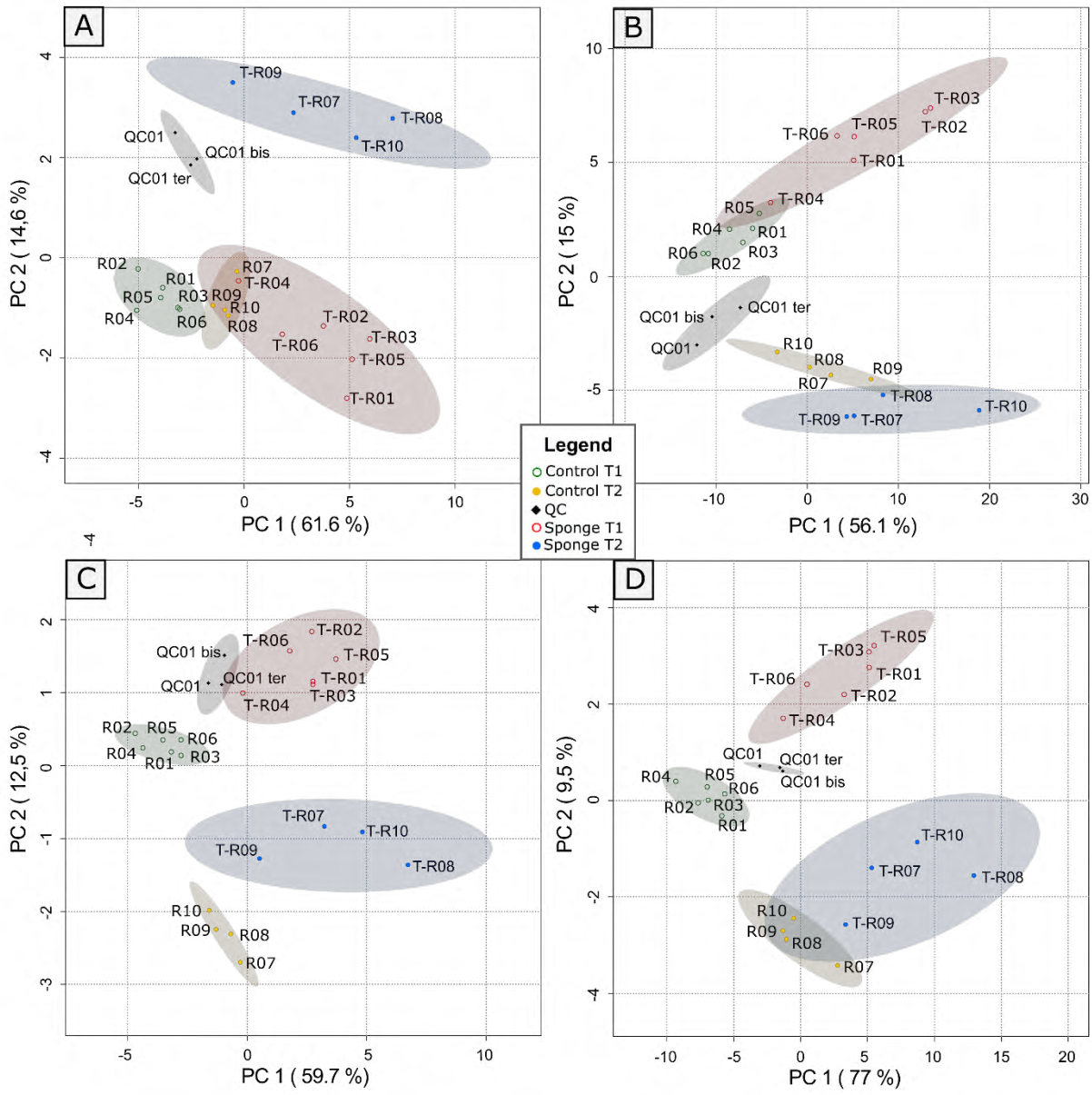


**Figure SI 1 - 2D PCA score plots with QC**: (A) $^{1}H$; (B) $^{13}C$; (C) UF COSY; (D) SYMAPS HSQC

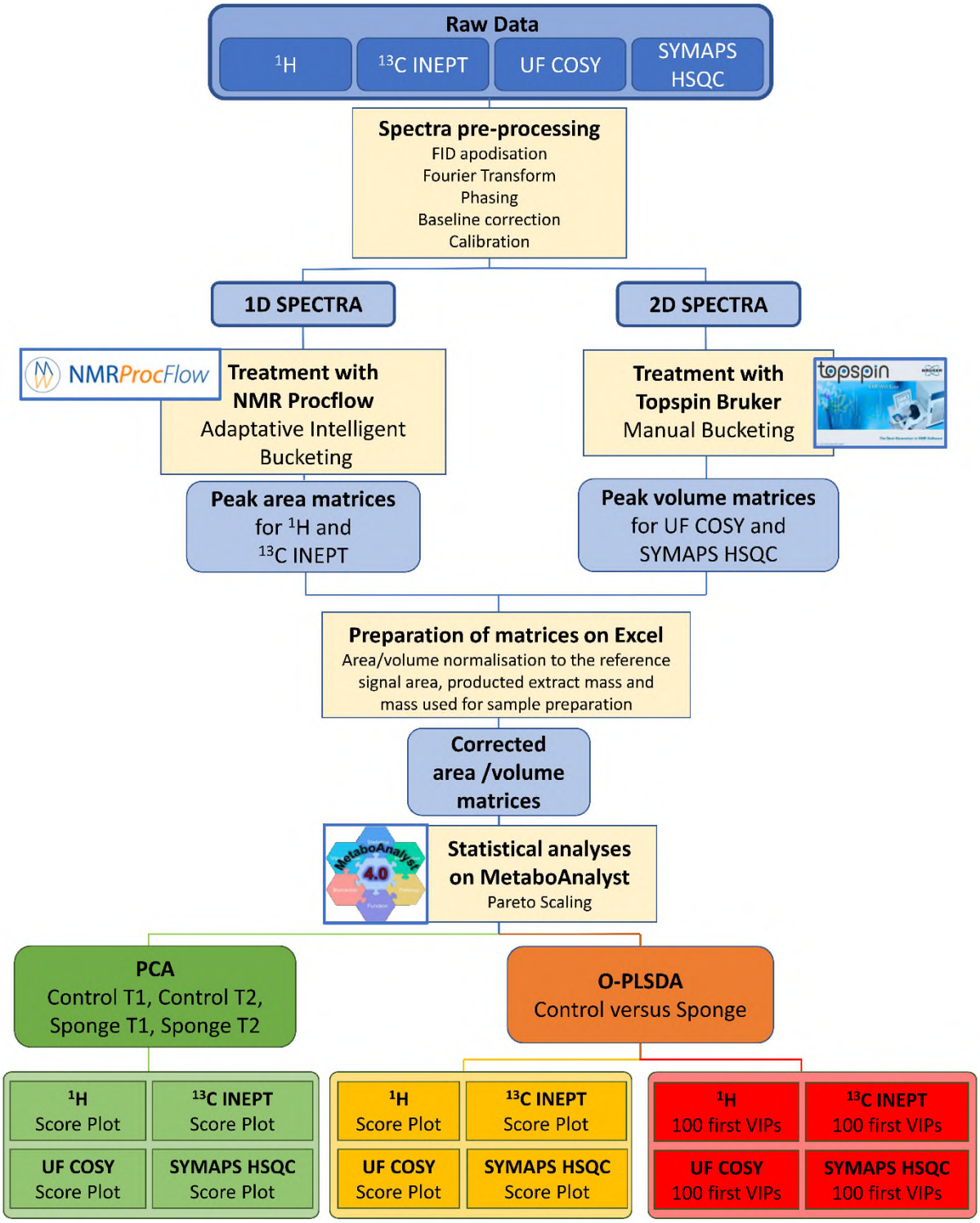


**Figure SI 2 - NMR data processing workflow**

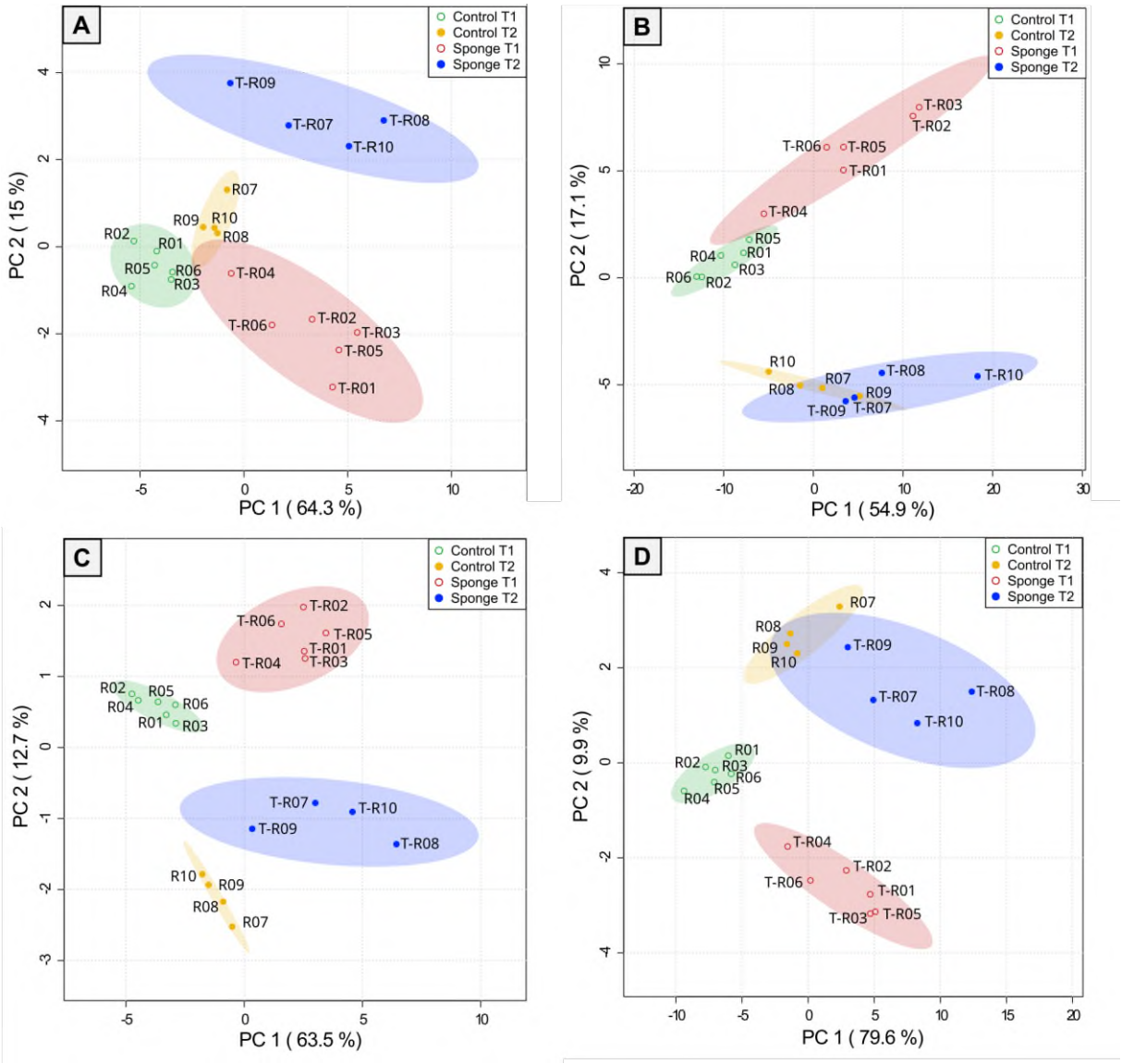


**Figure SI 3 - 2D PCA score plots:** (A) $^{1}H$; (B) $^{13}C$; (C) UF COSY; (D) SYMAPS HSQC

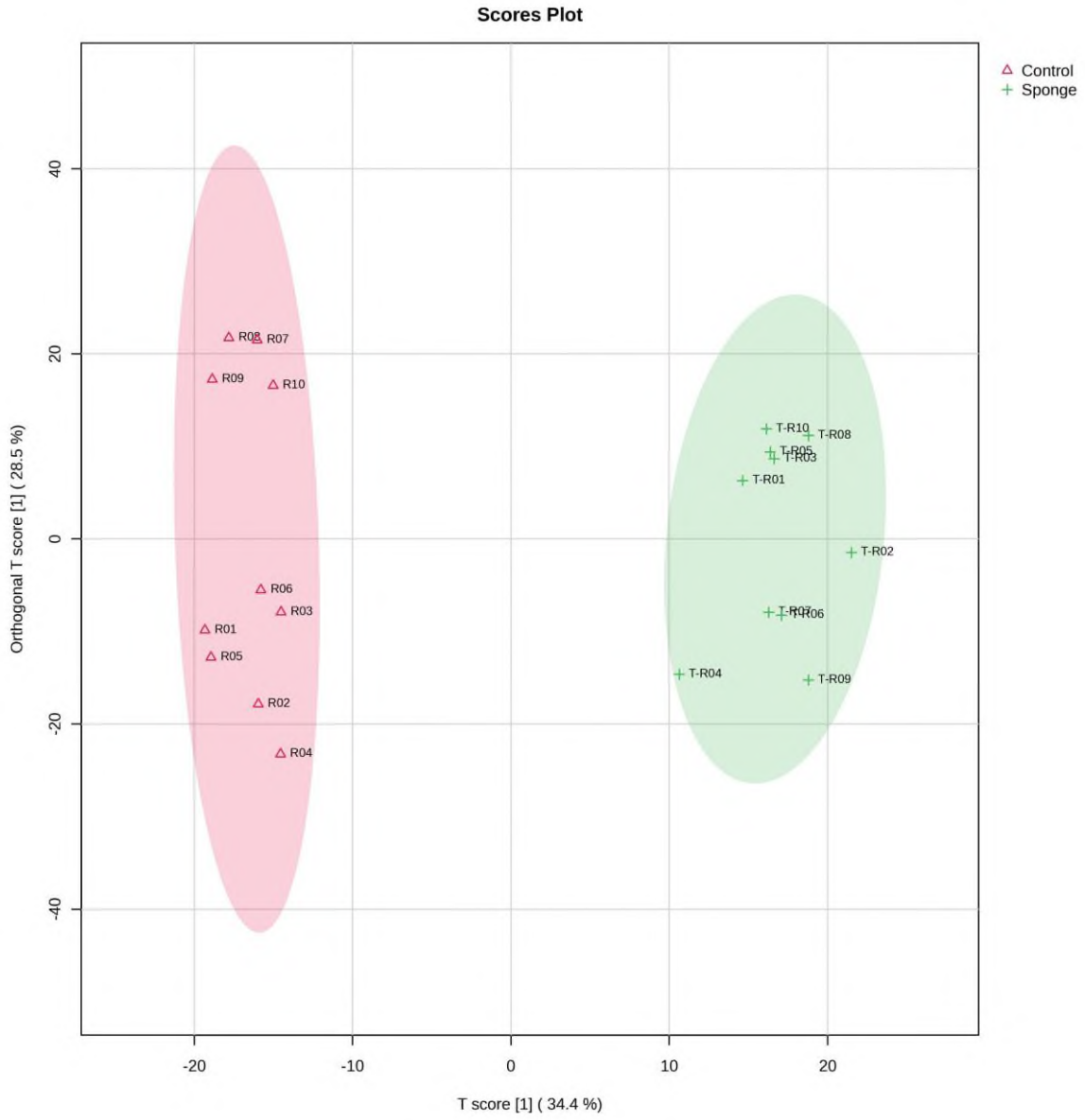


**Figure SI 4 - $^{1}$H OPLS-DA score plot:** medium effect

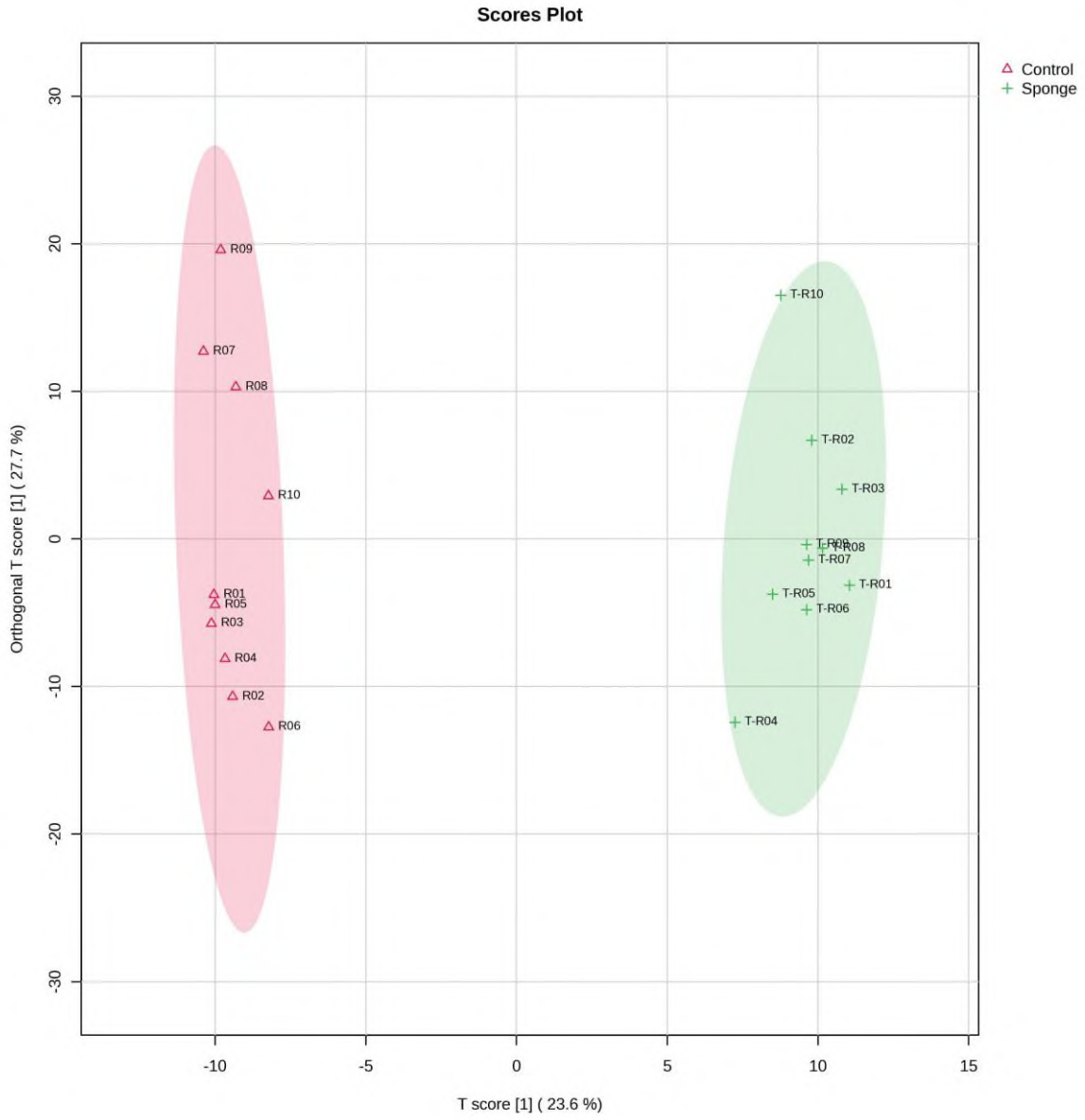


**Figure SI 5 - $^{13}$C INEPT OPLS-DA score plot:** medium effect

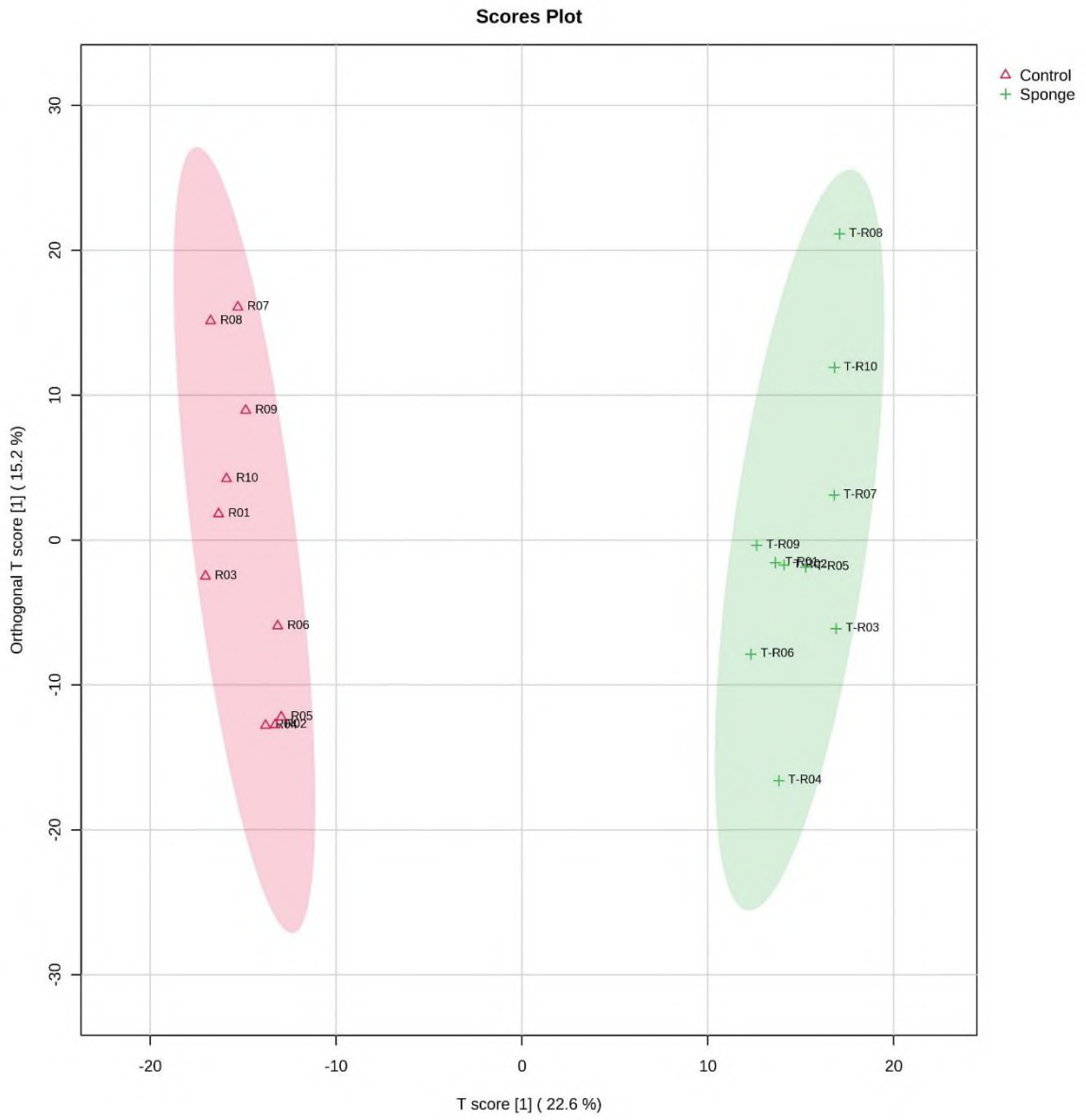


**Figure SI 6 - UF COSY OPLS-DA score plot:** medium effect

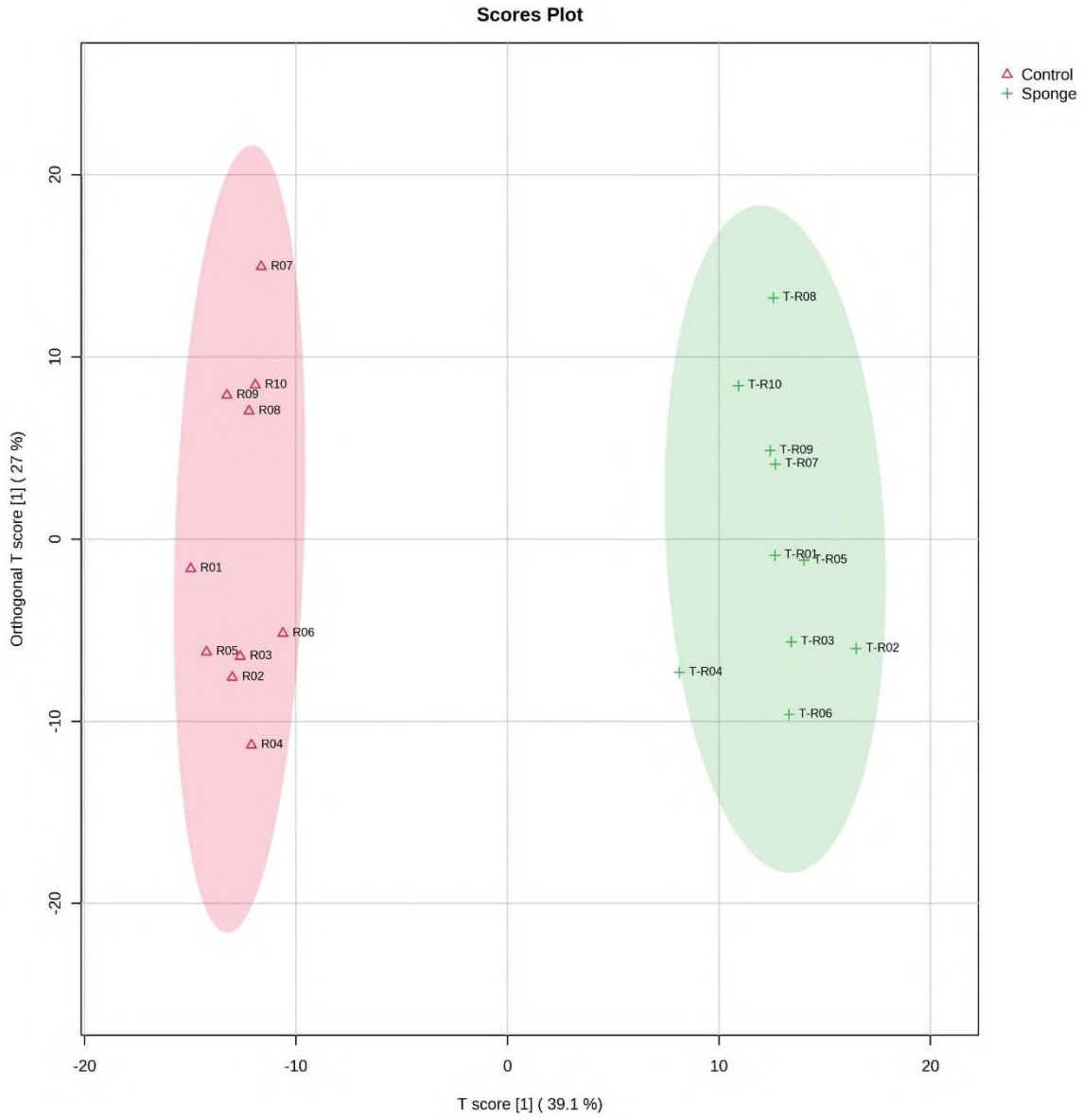


**Figure SI 7 - SYMAPS HSQC OPLS-DA score plot :** medium effect

**100 first VIP buckets**

| 15 to 16 ppm | 14 to 15 ppm | 13 to 14 ppm | 12 to 13 ppm | 11 to 12 ppm | 10 to 11 ppm | 9 to 10 ppm | 8 to 9 ppm | 7 to 8 ppm | 6 to 7 ppm | 5 to 6 ppm | 4 to 5 ppm | 3 to 4 ppm | 2 to 3 ppm | 1 to 2 ppm | 0 to 1 ppm |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 2 | 3 | 0 | 0 | 12 | 16 | 26 | 2 | 4 | 3 | 14 | 1 |

**All buckets**

| 15 to 16 ppm | 14 to 15 ppm | 13 to 14 ppm | 12 to 13 ppm | 11 to 12 ppm | 10 to 11 ppm | 9 to 10 ppm | 8 to 9 ppm | 7 to 8 ppm | 6 to 7 ppm | 5 to 6 ppm | 4 to 5 ppm | 3 to 4 ppm | 2 to 3 ppm | 1 to 2 ppm | 0 to 1 ppm |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 9 | 11 | 33 | 19 | 35 | 99 | 111 | 106 | 91 | 98 | 91 | 106 | 90 |

Min — Max

**Figure SI 8 - Distribution of the 100 first VIPs and all $^{1}$H buckets:** the upper part reports the 100 first VIP buckets highlighted by OPLS-DA analysis; the lower part reports all $^{1}$H buckets. Color code corresponds to bucket density in the spectrum range, from low amount (blue), to medium (white) and high (red).

**100 first VIP buckets**

| 140 to 150 ppm | 130 to 140 ppm | 120 to 130 ppm | 110 to 120 ppm | 100 to 110 ppm | 90 to 100 ppm | 80 to 90 ppm | 70 to 80 ppm | 60 to 70 ppm | 50 to 60 ppm | 40 to 50 ppm | 30 to 40 ppm | 20 to 30 ppm | 10 to 20 ppm | 0 to 10 ppm |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 3 | 14 | 23 | 3 | 0 | 0 | 1 | 0 | 4 | 1 | 9 | 28 | 8 | 0 |

**All buckets**

| 140 to 150 ppm | 130 to 140 ppm | 120 to 130 ppm | 110 to 120 ppm | 100 to 110 ppm | 90 to 100 ppm | 80 to 90 ppm | 70 to 80 ppm | 60 to 70 ppm | 50 to 60 ppm | 40 to 50 ppm | 30 to 40 ppm | 20 to 30 ppm | 10 to 20 ppm | 0 to 10 ppm | -10 to 0 ppm |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 18 | 19 | 82 | 62 | 7 | 1 | 0 | 3 | 6 | 23 | 20 | 39 | 83 | 40 | 0 | 5 |

Min — Max

**Figure SI 9 - Distribution of the 100 first VIP and all $^{13}$C INEPT buckets:** the upper part reports the 100 first VIP buckets highlighted by OPLS-DA analysis; the lower part reports all $^{13}$C buckets. Color code corresponds to bucket density in the spectrum range, from low amount (blue), to medium (white) and high (red).

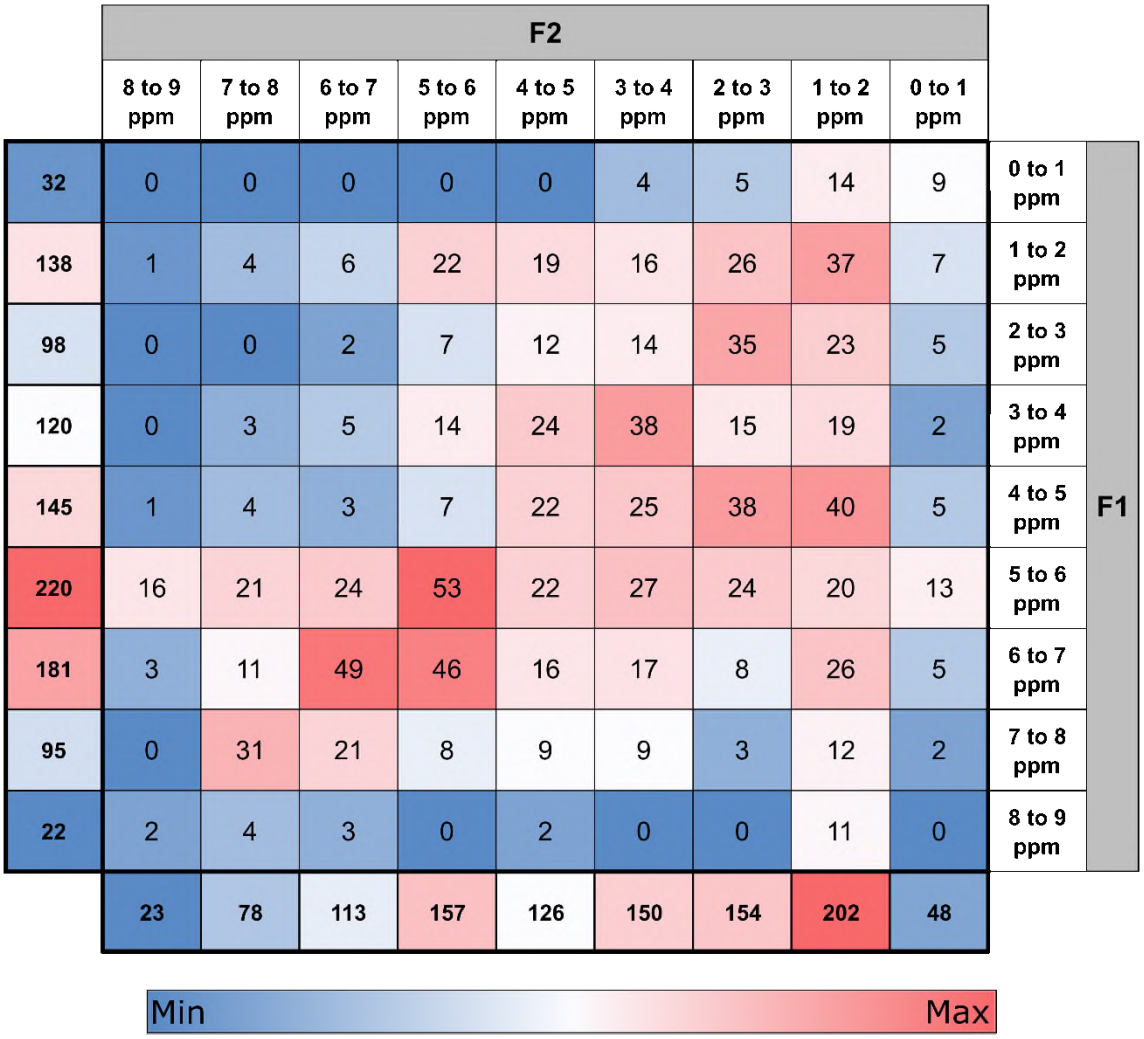


**Figure SI 10 - Spectral distribution of UF COSY buckets**. Color code corresponds to bucket density in the spectrum range, from low amount (blue), to medium (white) and high (red).

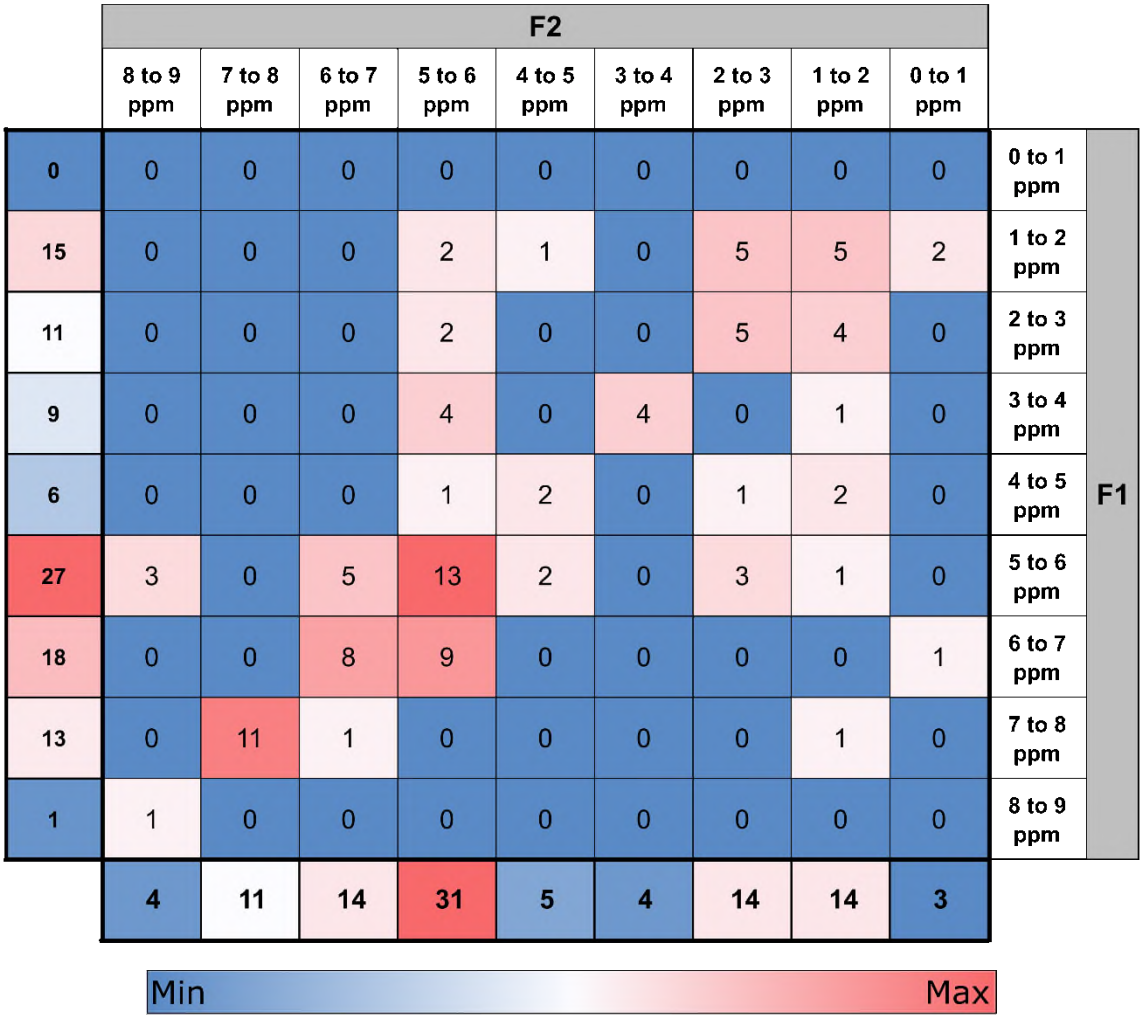


**Figure SI 11 - Spectral distribution of the first 100 VIP in UF COSY.** Color code corresponds to bucket density in the spectrum range, from low amount (blue), to medium (white) and high (red).

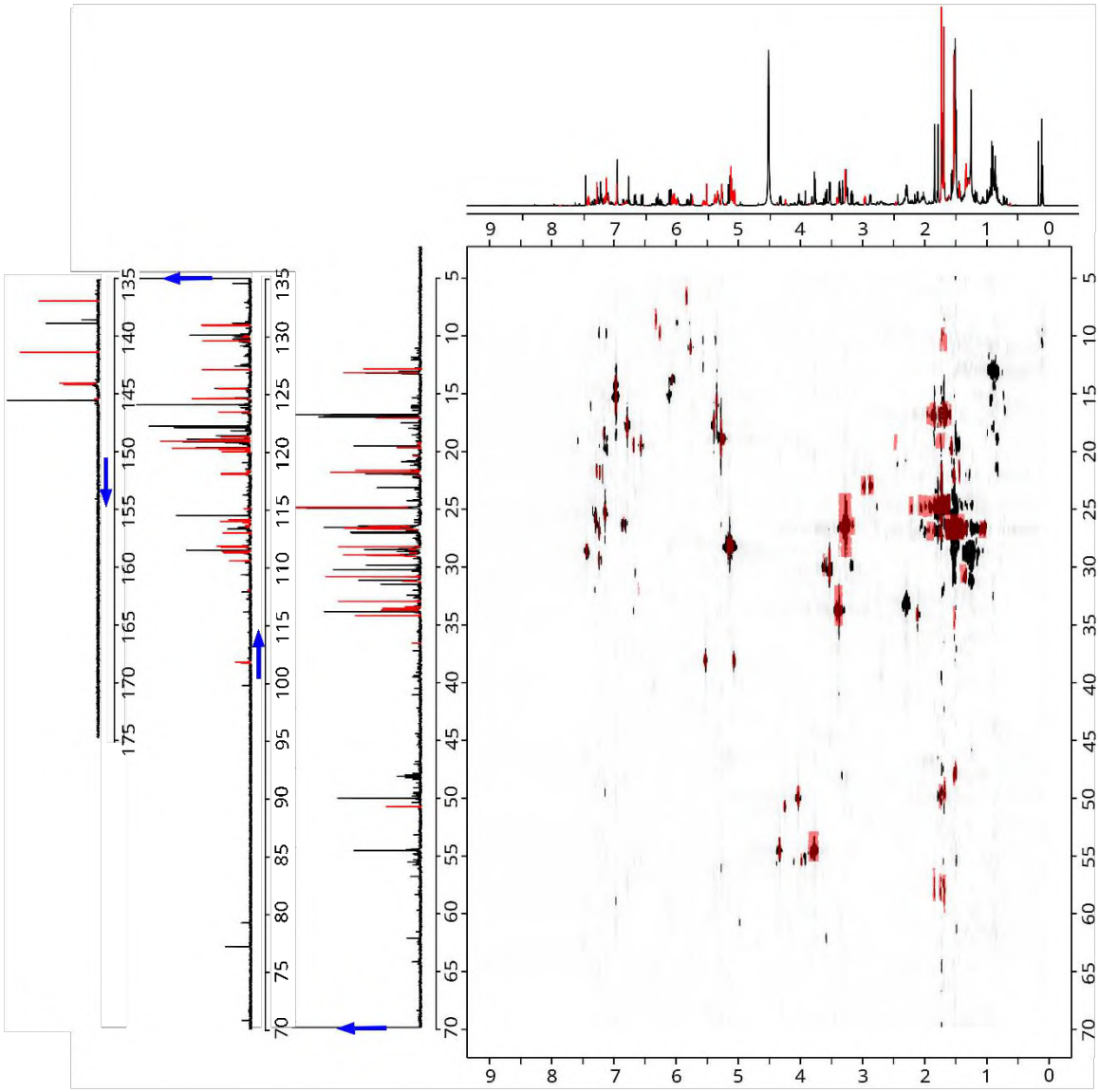


**Figure SI 12 - SYMAPS HSQC with $^{1}H$ and numerically folded $^{13}C$ INEPT spectra as projections:** on the top, 1D $^{1}H$ spectrum with their 100 first OPLS-DA VIP buckets in red, on the bottom left, 1D $^{13}C$ INEPT numerically folded spectrum with their 100 first OPLS-DA VIP buckets in red, on the bottom right, 2D SYMAPS HSQC spectrum with their 100 first OPLS-DA VIP buckets in red.

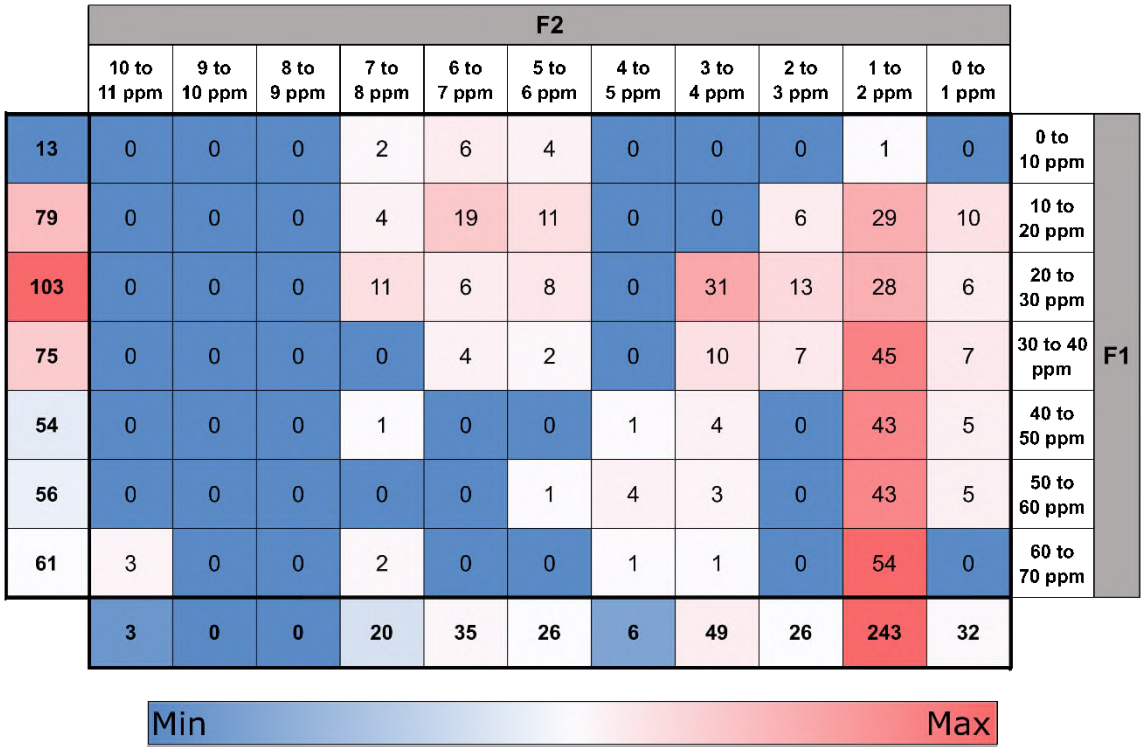


**Figure SI 13 - Spectral distribution of SYMAPS HSQC buckets.** Color code corresponds to bucket density in the spectrum range, from low amount (blue), to medium (white) and high (red).

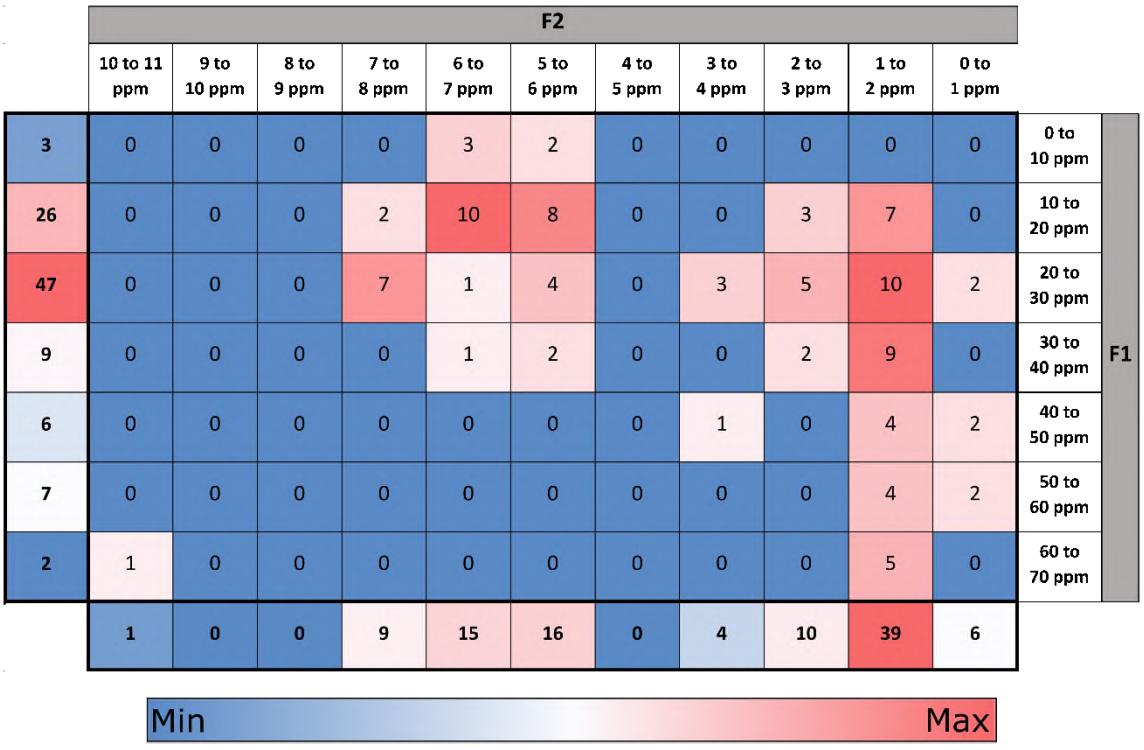


**Figure SI 14 - Spectral distribution of the first 100 VIP in SYMAPS HSQC.** Color code corresponds to bucket density in the spectrum range, from low amount (blue), to medium (white) and high (red).

**Table SI 1 - OPLS-DA model parameters for the four NMR methods** according the tested parameter (medium, time effects): "Medium" lines are the OPLS-DA results for Control (T1 + T2) samples versus Sponge (T1 + T2). Other lines show the effect time on T1 or T2 samples only. "Time" lines are the OPLS-DA results for T1 (Control + Sponge) vs T2 (Control + Sponge). For this study, we concentrate only on the results obtained on Medium lines to compare NMR methods.

| | | OPLS-DA Models Parameters | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | p1 | | | Permutation | | | | | |
| **Analysis technique** | **Studied factor** | **R2X** | **R2Y** | **Q2** | **n/2000** | **Q2** | **p** | **n/2000** | **R2Y** | **p** |
| **$^{1}$H** | **Medium** | **0.344** | **0.842** | **0.788** | **0** | **0.948** | **<5.10$^{-4}$** | **0** | **0.981** | **<5.10$^{-4}$** |
| $^{1}$H | Medium at T1 | 0.534 | 0.84 | 0.803 | 4 | 0.897 | 0.002 | 4 | 0.989 | 0.002 |
| $^{1}$H | Medium at T2 | 0.533 | 0.973 | 0.936 | 47 | 0.984 | 0.0235 | 47 | 0.998 | 0.0235 |
| **$^{1}$H** | **Time** | **0.192** | **0.838** | **0.74** | **0** | **0.954** | **<5.10$^{-4}$** | **0** | **0.983** | **<5.10$^{-4}$** |
| $^{1}$H | Time on Ctrl | 0.583 | 0.91 | 0.877 | 10 | 0.958 | 0.005 | 10 | 0.995 | 0.005 |
| $^{1}$H | Time on Sponge | 0.339 | 0.98 | 0.885 | 6 | 0.964 | 0.003 | 6 | 0.991 | 0.003 |
| **UF COSY** | **Medium** | **0.226** | **0.832** | **0.752** | **0** | **0.898** | **<5.10$^{-4}$** | **0** | **0.988** | **<5.10$^{-4}$** |
| UF COSY | Medium at T1 | 0.33 | 0.926 | 0.835 | 2 | 0.877 | 0.001 | 2 | 0.977 | 0.001 |
| UF COSY | Medium at T2 | 0.349 | 0.898 | 0.783 | 60 | 0.874 | 0.03 | 60 | 0.997 | 0.03 |
| **UF COSY** | **Time** | **0.147** | **0.844** | **0.722** | **0** | **0.909** | **<5.10$^{-4}$** | **0** | **0.984** | **<5.10$^{-4}$** |
| UF COSY | Time on Ctrl | 0.278 | 0.947 | 0.812 | 8 | 0.854 | 0.004 | 13 | 0.99 | 0.0065 |
| UF COSY | Time on Sponge | 0.247 | 0.968 | 0.773 | 6 | 0.831 | 0.003 | 32 | 0.995 | 0.016 |
| **$^{13}$C** | **Medium** | **0.236** | **0.757** | **0.644** | **0** | **0.877** | **<5.10$^{-4}$** | **12** | **0.991** | **0.006** |
| $^{13}$C | Medium at T1 | 0.357 | 0.76 | 0.676 | 7 | 0.802 | 0.0035 | 133 | 0.997 | 0.0665 |
| $^{13}$C | Medium at T2 | 0.355 | 0.93 | 0.854 | 61 | 0.957 | 0.0305 | 61 | 1 | 0.0305 |
| **$^{13}$C** | **Time** | **0.308** | **0.989** | **0.885** | **1** | **0.954** | **<5.10$^{-4}$** | **1** | **0.989** | **<5.10$^{-4}$** |
| $^{13}$C | Time on Ctrl | 0.551 | 0.895 | 0.834 | 9 | 0.938 | 0.0045 | 9 | 0.998 | 0.0045 |
| $^{13}$C | Time on Sponge | 0.369 | 0.872 | 0.797 | 6 | 0.951 | 0.003 | 6 | 0.995 | 0.003 |
| **SYMAPS HSQC** | **Medium** | **0.391** | **0.717** | **0.644** | **0** | **0.911** | **<5.10$^{-4}$** | **0** | **0.983** | **<5.10$^{-4}$** |
| SYMAPS HSQC | Medium at T1 | 0.6 | 0.871 | 0.839 | 7 | 0.885 | 0.0035 | 10 | 0.949 | 0.005 |
| SYMAPS HSQC | Medium at T2 | 0.53 | 0.834 | 0.755 | 53 | 0.942 | 0.0265 | 53 | 0.998 | 0.0265 |
| **SYMAPS HSQC** | **Time** | **0.225** | **0.582** | **0.5** | **0** | **0.962** | **<5.10$^{-4}$** | **0** | **0.995** | **<5.10$^{-4}$** |
| SYMAPS HSQC | Time on Ctrl | 0.551 | 0.895 | 0.834 | 13 | 0.938 | 0.0065 | 13 | 0.998 | 0.0065 |
| SYMAPS HSQC | Time on Sponge | 0.369 | 0.872 | 0.797 | 4 | 0.967 | 0.002 | 4 | 0.992 | 0.002 |

**Table SI 2 - 100 first $^{1}$H VIP buckets from OPLS-DA (Part 1/3)**

| $^{1}$H VIP Ranking | Bucket | V1 | V2 | Center [ppm] | Min [ppm] | Max [ppm] | Potential correspondence in F1 in UF COSY | Potential correspondence in F2 in UF COSY | Potential correspondence in F2 in HSQC |
|---|---|---|---|---|---|---|---|---|---|
| 1 | B7_2997 | 1.651 | 0.239 | 7.300 | 7.293 | 7.306 | Yes | Yes | Yes |
| 2 | B12_1959 | 1.626 | 0.182 | 12.196 | 12.187 | 12.205 | No | No | No |
| 3 | B7_4380 | 1.611 | 0.439 | 7.438 | 7.431 | 7.445 | Yes | Yes | Yes |
| 4 | B5_1073 | 1.593 | 0.467 | 5.107 | 5.103 | 5.112 | Yes | Yes | Yes |
| 5 | B6_0605 | 1.592 | 0.467 | 6.060 | 6.058 | 6.063 | Yes | Yes | Yes |
| 6 | B6_3701 | 1.591 | 0.391 | 6.370 | 6.368 | 6.372 | No | Yes | No |
| 7 | B7_1593 | 1.587 | 0.497 | 7.159 | 7.154 | 7.165 | Yes | Yes | Yes |
| 8 | B6_9771 | 1.585 | 0.479 | 6.977 | 6.973 | 6.981 | Yes | Yes | No |
| 9 | B4_3958 | 1.579 | 0.377 | 4.396 | 4.393 | 4.399 | No | No | No |
| 10 | B3_2907 | 1.577 | 0.506 | 3.291 | 3.279 | 3.302 | Yes | Yes | Yes |
| 11 | B7_5404 | 1.572 | 0.352 | 7.540 | 7.536 | 7.545 | No | No | No |
| 12 | B7_8954 | 1.569 | 0.406 | 7.895 | 7.889 | 7.902 | No | No | No |
| 13 | B6_4305 | 1.566 | 0.438 | 6.430 | 6.429 | 6.432 | No | Yes | No |
| 14 | B5_2933 | 1.561 | 0.522 | 5.293 | 5.285 | 5.301 | Yes | Yes | Yes |
| 15 | B6_6041 | 1.559 | 0.419 | 6.604 | 6.602 | 6.606 | Yes | Yes | Yes |
| 16 | B5_0751 | 1.559 | 0.515 | 5.075 | 5.071 | 5.079 | Yes | Yes | Yes |
| 17 | B6_0410 | 1.557 | 0.556 | 6.041 | 6.034 | 6.048 | Yes | Yes | Yes |
| 18 | B10_0666 | 1.555 | 0.526 | 10.067 | 10.058 | 10.075 | No | No | Yes |
| 19 | B6_8944 | 1.554 | 0.450 | 6.894 | 6.889 | 6.900 | No | Yes | No |
| 20 | B7_1698 | 1.553 | 0.560 | 7.170 | 7.165 | 7.175 | Yes | Yes | Yes |
| 21 | B3_4237 | 1.551 | 0.537 | 3.424 | 3.414 | 3.433 | Yes | Yes | No |
| 22 | B5_0607 | 1.551 | 0.512 | 5.061 | 5.050 | 5.071 | Yes | Yes | Yes |
| 23 | B7_5636 | 1.550 | 0.046 | 7.564 | 7.559 | 7.569 | No | No | No |
| 24 | B5_9981 | 1.546 | 0.488 | 5.998 | 5.995 | 6.001 | Yes | Yes | Yes |
| 25 | B7_7377 | 1.540 | 0.406 | 7.738 | 7.736 | 7.739 | No | No | No |
| 26 | B5_3480 | 1.538 | 0.587 | 5.348 | 5.342 | 5.354 | Yes | Yes | No |
| 27 | B2_5900 | 1.537 | 0.446 | 2.590 | 2.584 | 2.596 | No | Yes | No |
| 28 | B6_0304 | 1.535 | 0.550 | 6.030 | 6.027 | 6.034 | Yes | Yes | Yes |
| 29 | B5_0959 | 1.535 | 0.593 | 5.096 | 5.089 | 5.103 | Yes | Yes | Yes |
| 30 | B5_1211 | 1.532 | 0.595 | 5.121 | 5.115 | 5.127 | Yes | Yes | Yes |
| 31 | B7_1798 | 1.529 | 0.609 | 7.180 | 7.175 | 7.185 | Yes | Yes | Yes |
| 32 | B5_0812 | 1.528 | 0.449 | 5.081 | 5.079 | 5.083 | Yes | Yes | Yes |
| 33 | B7_2179 | 1.528 | 0.510 | 7.218 | 7.216 | 7.220 | Yes | Yes | No |
| 34 | B6_0852 | 1.526 | 0.592 | 6.085 | 6.082 | 6.088 | Yes | Yes | Yes |
| 35 | B5_5211 | 1.525 | 0.562 | 5.521 | 5.511 | 5.531 | Yes | Yes | Yes |
| 36 | B5_5333 | 1.524 | 0.529 | 5.533 | 5.531 | 5.535 | Yes | Yes | Yes |
| 37 | B6_0683 | 1.519 | 0.621 | 6.068 | 6.063 | 6.073 | Yes | Yes | Yes |
| 38 | B5_5422 | 1.519 | 0.596 | 5.542 | 5.540 | 5.545 | Yes | Yes | Yes |
| 39 | B6_0061 | 1.517 | 0.605 | 6.006 | 6.001 | 6.011 | Yes | Yes | Yes |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 40 | B1_5327 | 1.515 | 0.596 | 1.533 | 1.527 | 1.538 | Yes | Yes | Yes |
| 41 | B7_7332 | 1.515 | 0.548 | 7.733 | 7.730 | 7.736 | No | No | No |
| 42 | B7_7459 | 1.513 | 0.596 | 7.746 | 7.743 | 7.748 | No | No | No |

**Table SI 3 - 100 first $^{1}$H VIP buckets from OPLS-DA (Part 2/3)**

| $^{1}$H VIP Ranking | Bucket | V1 | V2 | Center [ppm] | Min [ppm] | Max [ppm] | Potential correspondence in F1 in UF COSY | Potential correspondence in F2 in UF COSY | Potential correspondence in F2 in HSQC |
|---|---|---|---|---|---|---|---|---|---|
| 43 | B6_8274 | 1.512 | 0.466 | 6.827 | 6.822 | 6.833 | No | Yes | Yes |
| 44 | B6_5057 | 1.511 | 0.439 | 6.506 | 6.504 | 6.507 | Yes | Yes | No |
| 45 | B2_4699 | 1.505 | 0.626 | 2.470 | 2.459 | 2.481 | Yes | Yes | Yes |
| 46 | B5_2790 | 1.501 | 0.607 | 5.279 | 5.273 | 5.285 | Yes | Yes | Yes |
| 47 | B1_3410 | 1.500 | 0.642 | 1.341 | 1.338 | 1.344 | Yes | Yes | Yes |
| 48 | B1_3349 | 1.500 | 0.635 | 1.335 | 1.332 | 1.338 | Yes | Yes | Yes |
| 49 | B7_1063 | 1.499 | 0.651 | 7.106 | 7.101 | 7.112 | No | Yes | Yes |
| 50 | B5_1325 | 1.499 | 0.643 | 5.132 | 5.127 | 5.138 | Yes | Yes | Yes |
| 51 | B5_5699 | 1.499 | 0.631 | 5.570 | 5.564 | 5.576 | Yes | Yes | No |
| 52 | B5_3620 | 1.498 | 0.642 | 5.362 | 5.354 | 5.370 | Yes | Yes | No |
| 53 | B5_5806 | 1.497 | 0.629 | 5.581 | 5.576 | 5.586 | Yes | Yes | No |
| 54 | B-0_3424 | 1.495 | 0.552 | -0.342 | -0.346 | -0.339 | No | No | No |
| 55 | B7_0661 | 1.491 | 0.608 | 7.066 | 7.061 | 7.072 | No | Yes | No |
| 56 | B5_4040 | 1.490 | 0.664 | 5.404 | 5.398 | 5.410 | Yes | Yes | No |
| 57 | B7_4522 | 1.487 | 0.648 | 7.452 | 7.445 | 7.459 | Yes | Yes | Yes |
| 58 | B5_2158 | 1.486 | 0.668 | 5.216 | 5.213 | 5.219 | Yes | Yes | No |
| 59 | B10_0452 | 1.486 | 0.642 | 10.045 | 10.039 | 10.051 | No | No | Yes |
| 60 | B5_5595 | 1.485 | 0.611 | 5.560 | 5.555 | 5.564 | Yes | Yes | No |
| 61 | B1_7432 | 1.485 | 0.662 | 1.743 | 1.727 | 1.759 | Yes | Yes | Yes |
| 62 | B10_1628 | 1.481 | 0.634 | 10.163 | 10.136 | 10.189 | No | No | No |
| 63 | B5_5498 | 1.479 | 0.624 | 5.550 | 5.545 | 5.555 | Yes | Yes | Yes |
| 64 | B7_4231 | 1.477 | 0.689 | 7.423 | 7.415 | 7.431 | Yes | Yes | No |
| 65 | B1_5253 | 1.471 | 0.673 | 1.525 | 1.523 | 1.527 | Yes | Yes | Yes |
| 66 | B7_8832 | 1.471 | 0.433 | 7.883 | 7.878 | 7.889 | No | No | No |
| 67 | B1_6921 | 1.470 | 0.663 | 1.692 | 1.685 | 1.700 | Yes | Yes | Yes |
| 68 | B2_9812 | 1.469 | 0.673 | 2.981 | 2.975 | 2.987 | Yes | Yes | Yes |
| 69 | B5_0862 | 1.468 | 0.670 | 5.086 | 5.083 | 5.089 | Yes | Yes | Yes |
| 70 | B6_0546 | 1.468 | 0.703 | 6.055 | 6.051 | 6.058 | Yes | Yes | Yes |
| 71 | B1_3268 | 1.467 | 0.677 | 1.327 | 1.322 | 1.332 | Yes | Yes | Yes |
| 72 | B6_8366 | 1.465 | 0.485 | 6.837 | 6.833 | 6.840 | No | No | Yes |
| 73 | B6_0793 | 1.461 | 0.698 | 6.079 | 6.076 | 6.082 | Yes | Yes | Yes |
| 74 | B6_9982 | 1.460 | 0.667 | 6.998 | 6.989 | 7.008 | Yes | Yes | No |
| 75 | B2_9691 | 1.459 | 0.677 | 2.969 | 2.963 | 2.975 | Yes | Yes | Yes |
| 76 | B-0_3297 | 1.459 | 0.531 | -0.330 | -0.339 | -0.320 | No | No | No |
| 77 | B5_7574 | 1.458 | 0.674 | 5.757 | 5.752 | 5.762 | Yes | Yes | Yes |
| 78 | B3_4054 | 1.456 | 0.663 | 3.405 | 3.397 | 3.414 | Yes | Yes | No |
| 79 | B4_2482 | 1.455 | 0.696 | 4.248 | 4.243 | 4.254 | Yes | Yes | No |
| 80 | B6_8452 | 1.455 | 0.523 | 6.845 | 6.840 | 6.850 | No | No | Yes |
| 81 | B5_5068 | 1.454 | 0.646 | 5.507 | 5.503 | 5.511 | Yes | Yes | No |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 82 | B1_4388 | 1.453 | 0.703 | 1.439 | 1.435 | 1.442 | Yes | Yes | Yes |
| 83 | B11_8978 | 1.453 | 0.391 | 11.898 | 11.886 | 11.909 | No | No | No |
| 84 | B11_9453 | 1.453 | 0.398 | 11.945 | 11.935 | 11.955 | No | No | No |

**Table SI 4 - 100 first $^{1}$H VIP buckets from OPLS-DA (Part 3/3)**

| $^{1}$H VIP Ranking | Bucket | V1 | V2 | Center [ppm] | Min [ppm] | Max [ppm] | Potential correspondence in F1 in UF COSY | Potential correspondence in F2 in UF COSY | Potential correspondence in F2 in HSQC |
|---|---|---|---|---|---|---|---|---|---|
| 85 | B5_5376 | 1.452 | 0.619 | 5.538 | 5.535 | 5.540 | Yes | Yes | Yes |
| 86 | B0_6249 | 1.452 | 0.677 | 0.625 | 0.614 | 0.636 | No | No | No |
| 87 | B5_3913 | 1.451 | 0.727 | 5.391 | 5.385 | 5.398 | Yes | Yes | No |
| 88 | B6_6852 | 1.451 | 0.595 | 6.685 | 6.682 | 6.689 | Yes | Yes | Yes |
| 89 | B1_6594 | 1.450 | 0.670 | 1.659 | 1.653 | 1.666 | Yes | Yes | Yes |
| 90 | B1_3470 | 1.448 | 0.716 | 1.347 | 1.344 | 1.350 | Yes | Yes | Yes |
| 91 | B3_8614 | 1.448 | 0.617 | 3.861 | 3.857 | 3.865 | No | Yes | No |
| 92 | B1_7036 | 1.447 | 0.722 | 1.704 | 1.700 | 1.708 | Yes | Yes | Yes |
| 93 | B1_2969 | 1.445 | 0.711 | 1.297 | 1.292 | 1.302 | No | Yes | Yes |
| 94 | B1_6109 | 1.445 | 0.734 | 1.611 | 1.605 | 1.617 | No | Yes | Yes |
| 95 | B7_1467 | 1.443 | 0.724 | 7.147 | 7.140 | 7.154 | Yes | Yes | No |
| 96 | B6_8849 | 1.443 | 0.640 | 6.885 | 6.881 | 6.889 | No | No | No |
| 97 | B5_5910 | 1.442 | 0.709 | 5.591 | 5.586 | 5.596 | Yes | Yes | No |
| 98 | B4_7440 | 1.441 | 0.650 | 4.744 | 4.740 | 4.748 | No | No | No |
| 99 | B1_2871 | 1.439 | 0.722 | 1.287 | 1.282 | 1.292 | No | Yes | Yes |
| 100 | B5_4208 | 1.439 | 0.737 | 5.421 | 5.417 | 5.425 | Yes | Yes | No |

**Table SI 5 - 100 first $^{13}C$ VIP buckets from OPLS-DA (Part 1/3)**

| $^{13}C$ VIP Ranking | Bucket | V1 | V2 | Center [ppm] | Min [ppm] | Max [ppm] | Potential Correspondence in F1 |
|---|---|---|---|---|---|---|---|
| 1 | B113_7829 | 1.821 | 0.568 | 113.783 | 113.748 | 113.818 | Yes |
| 2 | B119_9937 | 1.806 | 0.654 | 119.994 | 119.962 | 120.026 | Yes |
| 3 | B143_7794 | 1.804 | 0.672 | 143.779 | 143.757 | 143.801 | Yes |
| 4 | B107_8090 | 1.795 | 0.614 | 107.809 | 107.782 | 107.836 | Yes |
| 5 | B111_5956 | 1.792 | 0.585 | 111.596 | 111.572 | 111.620 | Yes |
| 6 | B28_9028 | 1.789 | 0.547 | 28.903 | 28.877 | 28.929 | Yes |
| 7 | B129_6858 | 1.763 | 0.637 | 129.686 | 129.670 | 129.701 | Yes |
| 8 | B117_8638 | 1.762 | 0.717 | 117.864 | 117.825 | 117.903 | Yes |
| 9 | B111_1731 | 1.761 | 0.715 | 111.173 | 111.144 | 111.203 | Yes |
| 10 | B28_8019 | 1.749 | 0.674 | 28.802 | 28.782 | 28.822 | Yes |
| 11 | B129_8164 | 1.746 | 0.642 | 129.816 | 129.789 | 129.844 | Yes |
| 12 | B118_1131 | 1.741 | 0.419 | 118.113 | 118.090 | 118.136 | Yes |
| 13 | B113_7300 | 1.723 | 0.279 | 113.730 | 113.712 | 113.748 | Yes |
| 14 | B143_8644 | 1.710 | 0.734 | 143.864 | 143.854 | 143.875 | Yes |
| 15 | B27_5577 | 1.704 | 0.603 | 27.558 | 27.541 | 27.575 | Yes |
| 16 | B117_9304 | 1.699 | 0.754 | 117.930 | 117.903 | 117.958 | Yes |
| 17 | B20_3724 | 1.699 | 0.608 | 20.372 | 20.356 | 20.389 | Yes |
| 18 | B101_6241 | 1.698 | 0.681 | 101.624 | 101.576 | 101.672 | Yes |
| 19 | B113_6743 | 1.697 | 0.583 | 113.674 | 113.637 | 113.712 | Yes |
| 20 | B13_2898 | 1.692 | 0.750 | 13.290 | 13.246 | 13.334 | Yes |
| 21 | B26_6971 | 1.681 | 0.781 | 26.697 | 26.679 | 26.716 | Yes |
| 22 | B28_8418 | 1.672 | 0.740 | 28.842 | 28.822 | 28.862 | Yes |
| 23 | B36_5855 | 1.671 | 0.661 | 36.585 | 36.567 | 36.604 | Yes |
| 24 | B112_7543 | 1.667 | 0.772 | 112.754 | 112.686 | 112.823 | Yes |
| 25 | B33_5858 | 1.665 | 0.829 | 33.586 | 33.562 | 33.609 | Yes |
| 26 | B70_5963 | 1.662 | 0.685 | 70.596 | 70.578 | 70.614 | No |
| 27 | B111_6413 | 1.657 | 0.823 | 111.641 | 111.620 | 111.663 | Yes |
| 28 | B32_9499 | 1.650 | 0.820 | 32.950 | 32.893 | 33.007 | Yes |
| 29 | B28_5689 | 1.640 | 0.766 | 28.569 | 28.540 | 28.598 | Yes |
| 30 | B113_4295 | 1.638 | 0.356 | 113.430 | 113.395 | 113.464 | Yes |
| 31 | B30_8279 | 1.635 | 0.857 | 30.828 | 30.783 | 30.872 | Yes |
| 32 | B31_2518 | 1.628 | 0.813 | 31.252 | 31.195 | 31.309 | Yes |
| 33 | B123_2084 | 1.625 | 0.805 | 123.208 | 123.173 | 123.244 | Yes |
| 34 | B19_2220 | 1.623 | 0.617 | 19.222 | 19.206 | 19.239 | Yes |
| 35 | B26_6268 | 1.622 | 0.805 | 26.627 | 26.608 | 26.646 | Yes |
| 36 | B34_1764 | 1.605 | 0.843 | 34.176 | 34.122 | 34.231 | Yes |
| 37 | B26_7396 | 1.605 | 0.823 | 26.740 | 26.716 | 26.763 | Yes |
| 38 | B28_6188 | 1.584 | 0.860 | 28.619 | 28.598 | 28.639 | Yes |
| 39 | B129_3957 | 1.583 | 0.885 | 129.396 | 129.353 | 129.438 | Yes |
| 40 | B113_8888 | 1.583 | 0.633 | 113.889 | 113.860 | 113.917 | Yes |
| 41 | B120_8366 | 1.582 | 0.878 | 120.837 | 120.803 | 120.870 | Yes |
| 42 | B136_6586 | 1.581 | 0.863 | 136.659 | 136.601 | 136.717 | Yes |

**Table SI 6 - 100 first $^{13}C$ VIP buckets from OPLS-DA (Part 2/3)**

| $^{13}C$ VIP Ranking | Bucket | V1 | V2 | Center [ppm] | Min [ppm] | Max [ppm] | Potential Correspondence in F1 |
|---|---|---|---|---|---|---|---|
| 43 | B141_0771 | 1.572 | 0.905 | 141.077 | 141.021 | 141.134 | Yes |
| 44 | B18_0579 | 1.566 | 0.707 | 18.058 | 18.017 | 18.099 | Yes |
| 45 | B28_4757 | 1.566 | 0.888 | 28.476 | 28.461 | 28.490 | Yes |
| 46 | B126_9110 | 1.566 | 0.825 | 126.911 | 126.863 | 126.959 | No |
| 47 | B120_1204 | 1.562 | 0.899 | 120.120 | 120.088 | 120.152 | Yes |
| 48 | B13_1803 | 1.561 | 0.909 | 13.180 | 13.155 | 13.206 | No |
| 49 | B119_7879 | 1.561 | 0.867 | 119.788 | 119.730 | 119.846 | Yes |
| 50 | B118_1616 | 1.560 | 0.588 | 118.162 | 118.136 | 118.187 | Yes |
| 51 | B59_6992 | 1.559 | 0.289 | 59.699 | 59.677 | 59.721 | Yes |
| 52 | B124_3739 | 1.555 | 0.891 | 124.374 | 124.324 | 124.424 | No |
| 53 | B28_2444 | 1.552 | 0.914 | 28.244 | 28.208 | 28.280 | Yes |
| 54 | B28_9483 | 1.550 | 0.804 | 28.948 | 28.929 | 28.968 | Yes |
| 55 | B19_3796 | 1.527 | 0.715 | 19.380 | 19.354 | 19.405 | Yes |
| 56 | B21_9970 | 1.526 | 0.889 | 21.997 | 21.967 | 22.027 | Yes |
| 57 | B21_7945 | 1.525 | 0.928 | 21.794 | 21.758 | 21.831 | Yes |
| 58 | B145_1500 | 1.522 | 0.520 | 145.150 | 145.109 | 145.191 | No |
| 59 | B110_3815 | 1.522 | 0.878 | 110.381 | 110.352 | 110.411 | Yes |
| 60 | B125_2562 | 1.519 | 0.781 | 125.256 | 125.231 | 125.282 | Yes |
| 61 | B113_9521 | 1.510 | 0.756 | 113.952 | 113.917 | 113.987 | Yes |
| 62 | B28_6892 | 1.509 | 0.857 | 28.689 | 28.658 | 28.721 | Yes |
| 63 | B24_8313 | 1.508 | 0.920 | 24.831 | 24.808 | 24.855 | Yes |
| 64 | B111_5450 | 1.505 | 0.965 | 111.545 | 111.518 | 111.572 | Yes |
| 65 | B12_8547 | 1.491 | 0.926 | 12.855 | 12.809 | 12.901 | No |
| 66 | B130_7097 | 1.489 | 0.888 | 130.710 | 130.674 | 130.746 | Yes |
| 67 | B143_8842 | 1.488 | 0.974 | 143.884 | 143.875 | 143.894 | Yes |
| 68 | B111_0725 | 1.479 | 0.958 | 111.073 | 111.043 | 111.102 | Yes |
| 69 | B121_0437 | 1.472 | 0.930 | 121.044 | 120.999 | 121.088 | Yes |
| 70 | B111_1227 | 1.470 | 0.786 | 111.123 | 111.102 | 111.144 | Yes |
| 71 | B113_3233 | 1.467 | 0.737 | 113.323 | 113.291 | 113.355 | Yes |
| 72 | B50_7373 | 1.466 | 0.947 | 50.737 | 50.675 | 50.799 | Yes |
| 73 | B33_4582 | 1.464 | 0.991 | 33.458 | 33.432 | 33.485 | Yes |
| 74 | B44_4838 | 1.457 | 0.263 | 44.484 | 44.467 | 44.501 | No |
| 75 | B33_7496 | 1.456 | 0.968 | 33.750 | 33.719 | 33.780 | Yes |
| 76 | B101_5378 | 1.446 | 0.600 | 101.538 | 101.500 | 101.576 | Yes |
| 77 | B26_4658 | 1.446 | 0.869 | 26.466 | 26.455 | 26.477 | Yes |
| 78 | B22_0804 | 1.440 | 0.735 | 22.080 | 22.056 | 22.105 | Yes |
| 79 | B29_0139 | 1.435 | 1.001 | 29.014 | 28.968 | 29.060 | Yes |
| 80 | B119_8680 | 1.429 | 0.687 | 119.868 | 119.846 | 119.890 | Yes |
| 81 | B130_7794 | 1.426 | 0.956 | 130.779 | 130.746 | 130.813 | Yes |
| 82 | B124_4408 | 1.424 | 1.018 | 124.441 | 124.424 | 124.458 | No |
| 83 | B19_6859 | 1.421 | 0.998 | 19.686 | 19.622 | 19.750 | Yes |
| 84 | B17_1152 | 1.420 | 0.957 | 17.115 | 17.085 | 17.146 | Yes |

**Table SI 7 - 100 first $^{13}$C VIP buckets from OPLS-DA (Part 3/3)**

| $^{13}$C VIP Ranking | Bucket | V1 | V2 | Center [ppm] | Min [ppm] | Max [ppm] | Potential Correspondence in F1 |
|---|---|---|---|---|---|---|---|
| 85 | B21_6387 | 1.418 | 1.023 | 21.639 | 21.613 | 21.665 | Yes |
| 86 | B143_8394 | 1.415 | 0.959 | 143.839 | 143.825 | 143.854 | Yes |
| 87 | B28_6485 | 1.413 | 0.969 | 28.649 | 28.639 | 28.658 | Yes |
| 88 | B120_7122 | 1.399 | 1.012 | 120.712 | 120.679 | 120.745 | Yes |
| 89 | B121_1191 | 1.388 | 0.946 | 121.119 | 121.088 | 121.150 | Yes |
| 90 | B29_0763 | 1.371 | 1.042 | 29.076 | 29.060 | 29.093 | Yes |
| 91 | B114_8669 | 1.370 | 0.623 | 114.867 | 114.833 | 114.900 | Yes |
| 92 | B58_1908 | 1.366 | 0.444 | 58.191 | 58.171 | 58.211 | Yes |
| 93 | B26_6623 | 1.348 | 1.046 | 26.662 | 26.646 | 26.679 | Yes |
| 94 | B125_2064 | 1.334 | 0.663 | 125.206 | 125.182 | 125.231 | Yes |
| 95 | B27_2848 | 1.326 | 0.961 | 27.285 | 27.249 | 27.321 | Yes |
| 96 | B38_7969 | 1.322 | 0.618 | 38.797 | 38.751 | 38.843 | Yes |
| 97 | B26_5584 | 1.316 | 1.088 | 26.558 | 26.521 | 26.596 | Yes |
| 98 | B55_6512 | 1.312 | 0.886 | 55.651 | 55.628 | 55.674 | No |
| 99 | B26_4991 | 1.304 | 1.060 | 26.499 | 26.477 | 26.521 | Yes |
| 100 | B29_2522 | 1.290 | 1.001 | 29.252 | 29.196 | 29.308 | Yes |

**Table SI 8 - 100 first UF COSY VIP buckets from OPLS-DA (Part 1/3)**

| UF COSY VIP Ranking | Nom | V1 | V2 | Center F1 [ppm] | Start $^{1}$H F1 [ppm] | End $^{1}$H F1 [ppm] | Center F2 [ppm] | Start $^{1}$H F2 [ppm] | End $^{1}$H F2 [ppm] | Potential correspondence in F1 | Potential correspondence in F2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | B_1013 | 2.022 | 0.448 | 7.447 | 7.455 | 7.434 | 7.456 | 7.541 | 7.370 | Yes | Yes |
| **2** | B_0998 | 1.987 | 0.550 | 7.145 | 7.184 | 7.114 | 7.434 | 7.526 | 7.334 | Yes | Yes |
| **3** | B_1007 | 1.960 | 0.580 | 7.437 | 7.477 | 7.400 | 7.176 | 7.266 | 7.093 | Yes | Yes |
| **4** | B_0016 | 1.944 | 0.728 | 1.338 | 1.368 | 1.307 | 1.335 | 1.405 | 1.261 | Yes | Yes |
| **5** | B_1004 | 1.940 | 0.609 | 7.279 | 7.302 | 7.256 | 7.084 | 7.148 | 7.028 | Yes | Yes |
| **6** | B_0461 | 1.934 | 0.666 | 5.061 | 5.095 | 5.035 | 5.088 | 5.176 | 4.996 | Yes | Yes |
| **7** | B_0981 | 1.933 | 0.658 | 7.016 | 7.035 | 6.999 | 7.014 | 7.119 | 6.902 | Yes | Yes |
| **8** | B_0627 | 1.919 | 0.700 | 1.705 | 1.733 | 1.675 | 5.644 | 5.793 | 5.493 | Yes | Yes |
| **9** | B_0469 | 1.918 | 0.788 | 5.764 | 5.810 | 5.715 | 6.625 | 6.804 | 6.439 | Yes | Yes |
| **10** | B_1003 | 1.899 | 0.738 | 7.224 | 7.249 | 7.201 | 7.084 | 7.151 | 7.014 | Yes | Yes |
| **11** | B_0542 | 1.897 | 0.690 | 6.071 | 6.087 | 6.052 | 6.086 | 6.172 | 6.002 | Yes | Yes |
| **12** | B_0468 | 1.894 | 0.848 | 5.561 | 5.609 | 5.514 | 5.973 | 6.126 | 5.818 | Yes | Yes |
| **13** | B_0486 | 1.891 | 0.422 | 5.779 | 5.793 | 5.760 | 6.372 | 6.425 | 6.323 | Yes | Yes |
| **14** | B_0526 | 1.888 | 0.780 | 6.051 | 6.065 | 6.037 | 5.132 | 5.261 | 4.993 | Yes | Yes |
| **15** | B_0956 | 1.885 | 0.671 | 6.591 | 6.602 | 6.572 | 6.367 | 6.443 | 6.291 | Yes | Yes |
| **16** | B_0224 | 1.883 | 0.398 | 1.081 | 1.107 | 1.054 | 2.532 | 2.617 | 2.448 | No | Yes |
| **17** | B_0454 | 1.872 | 0.795 | 5.521 | 5.547 | 5.496 | 5.536 | 5.671 | 5.396 | Yes | Yes |
| **18** | B_0148 | 1.872 | 0.917 | 3.383 | 3.445 | 3.323 | 5.358 | 5.490 | 5.218 | Yes | Yes |
| **19** | B_0459 | 1.863 | 0.847 | 5.259 | 5.313 | 5.206 | 5.304 | 5.415 | 5.189 | Yes | Yes |
| **20** | B_0214 | 1.856 | 0.867 | 1.556 | 1.589 | 1.529 | 4.269 | 4.377 | 4.152 | Yes | Yes |
| **21** | B_0127 | 1.855 | 0.662 | 2.284 | 2.313 | 2.259 | 2.090 | 2.163 | 2.022 | No | No |
| **22** | B_0980 | 1.854 | 0.868 | 6.967 | 6.992 | 6.938 | 6.992 | 7.104 | 6.880 | Yes | Yes |
| **23** | B_0473 | 1.854 | 0.943 | 5.140 | 5.197 | 5.079 | 6.108 | 6.228 | 5.986 | Yes | Yes |
| **24** | B_0477 | 1.853 | 0.816 | 5.086 | 5.099 | 5.073 | 5.822 | 5.871 | 5.773 | Yes | No |
| **25** | B_0633 | 1.853 | 0.384 | 2.284 | 2.320 | 2.252 | 5.968 | 6.024 | 5.912 | No | Yes |
| **26** | B_0223 | 1.851 | 0.352 | 1.150 | 1.166 | 1.127 | 2.575 | 2.667 | 2.485 | No | Yes |
| **27** | B_0147 | 1.847 | 0.934 | 3.249 | 3.319 | 3.184 | 5.283 | 5.424 | 5.140 | Yes | Yes |
| **28** | B_0962 | 1.847 | 0.919 | 6.556 | 6.601 | 6.503 | 5.768 | 5.894 | 5.643 | Yes | Yes |
| **29** | B_0463 | 1.846 | 0.728 | 5.091 | 5.118 | 5.065 | 5.369 | 5.452 | 5.284 | Yes | Yes |
| **30** | B_0502 | 1.844 | 0.965 | 5.987 | 6.017 | 5.956 | 5.558 | 5.658 | 5.465 | Yes | Yes |
| **31** | B_1006 | 1.842 | 0.599 | 7.284 | 7.302 | 7.267 | 7.176 | 7.211 | 7.149 | Yes | Yes |
| **32** | B_0991 | 1.839 | 0.725 | 7.180 | 7.201 | 7.156 | 6.976 | 7.046 | 6.901 | Yes | Yes |
| **33** | B_0033 | 1.836 | 0.892 | 2.289 | 2.317 | 2.258 | 1.496 | 1.572 | 1.429 | No | Yes |
| **34** | B_0263 | 1.833 | 0.967 | 4.249 | 4.294 | 4.200 | 1.540 | 1.636 | 1.446 | Yes | Yes |
| **35** | B_0593 | 1.833 | 0.983 | 4.244 | 4.288 | 4.198 | 4.258 | 4.344 | 4.175 | Yes | Yes |
| **36** | B_0226 | 1.825 | 0.387 | 1.046 | 1.082 | 1.008 | 2.235 | 2.324 | 2.154 | No | No |
| **37** | B_0017 | 1.825 | 0.935 | 1.338 | 1.370 | 1.307 | 0.898 | 1.009 | 0.789 | Yes | No |
| **38** | B_0023 | 1.812 | 1.001 | 1.437 | 1.478 | 1.398 | 0.909 | 0.989 | 0.838 | Yes | No |
| **39** | B_0082 | 1.810 | 0.958 | 2.314 | 2.360 | 2.266 | 2.947 | 3.035 | 2.865 | No | Yes |
| **40** | B_0476 | 1.802 | 0.676 | 5.130 | 5.145 | 5.109 | 5.800 | 5.841 | 5.763 | Yes | No |
| **41** | B_1023 | 1.801 | 0.747 | 8.244 | 8.261 | 8.225 | 8.233 | 8.292 | 8.172 | No | No |
| **42** | B_0549 | 1.800 | 0.945 | 6.284 | 6.306 | 6.255 | 5.854 | 5.972 | 5.743 | No | Yes |

**Table SI 9- 100 first UF COSY VIP buckets from OPLS-DA (Part 2/3)**

| UF COSY VIP Ranking | Nom | V1 | V2 | Center F1 [ppm] | Start $^1$H F1 [ppm] | End $^1$H F1 [ppm] | Center F2 [ppm] | Start $^1$H F2 [ppm] | End $^1$H F2 [ppm] | Potential correspondence in F1 | Potential correspondence in F2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **43** | B_0081 | 1.799 | 0.979 | 2.962 | 3.002 | 2.925 | 2.305 | 2.400 | 2.210 | Yes | No |
| **44** | B_0528 | 1.796 | 0.760 | 5.987 | 5.992 | 5.979 | 5.110 | 5.225 | 4.996 | No | Yes |
| **45** | B_0458 | 1.792 | 1.022 | 5.378 | 5.435 | 5.316 | 5.385 | 5.497 | 5.281 | Yes | Yes |
| **46** | B_0442 | 1.790 | 0.782 | 5.650 | 5.673 | 5.630 | 5.094 | 5.147 | 5.042 | No | Yes |
| **47** | B_0026 | 1.786 | 0.990 | 1.734 | 1.758 | 1.713 | 1.723 | 1.861 | 1.588 | Yes | Yes |
| **48** | B_0781 | 1.785 | 0.975 | 5.784 | 5.808 | 5.762 | 2.300 | 2.396 | 2.199 | Yes | No |
| **49** | B_0103 | 1.784 | 0.854 | 2.962 | 3.001 | 2.927 | 2.473 | 2.539 | 2.406 | Yes | Yes |
| **50** | B_0748 | 1.784 | 0.486 | 7.289 | 7.311 | 7.271 | 1.049 | 1.122 | 0.974 | Yes | No |
| **51** | B_0965 | 1.779 | 0.479 | 6.596 | 6.619 | 6.573 | 6.620 | 6.731 | 6.504 | Yes | Yes |
| **52** | B_0783 | 1.777 | 0.572 | 5.804 | 5.826 | 5.788 | 2.090 | 2.180 | 1.997 | No | No |
| **53** | B_0585 | 1.776 | 0.823 | 4.536 | 4.641 | 4.429 | 4.549 | 4.679 | 4.413 | No | No |
| **54** | B_0957 | 1.770 | 1.047 | 6.680 | 6.721 | 6.637 | 6.340 | 6.482 | 6.197 | No | Yes |
| **55** | B_0028 | 1.768 | 0.994 | 1.452 | 1.507 | 1.395 | 2.095 | 2.182 | 2.012 | Yes | No |
| **56** | B_0699 | 1.768 | 0.722 | 6.957 | 6.985 | 6.937 | 0.725 | 0.804 | 0.652 | Yes | Yes |
| **57** | B_0241 | 1.768 | 1.016 | 3.259 | 3.309 | 3.208 | 1.718 | 1.839 | 1.599 | Yes | Yes |
| **58** | B_0175 | 1.765 | 0.671 | 3.279 | 3.296 | 3.255 | 3.783 | 3.862 | 3.701 | Yes | Yes |
| **59** | B_0630 | 1.764 | 0.747 | 1.700 | 1.748 | 1.654 | 5.299 | 5.416 | 5.181 | Yes | Yes |
| **60** | B_0771 | 1.764 | 1.018 | 5.269 | 5.310 | 5.224 | 1.718 | 1.833 | 1.595 | Yes | Yes |
| **61** | B_0472 | 1.757 | 0.709 | 5.794 | 5.815 | 5.770 | 5.962 | 6.022 | 5.900 | No | Yes |
| **62** | B_0894 | 1.753 | 0.588 | 5.650 | 5.678 | 5.627 | 8.346 | 8.415 | 8.277 | No | No |
| **63** | B_0539 | 1.750 | 0.726 | 6.036 | 6.050 | 6.020 | 5.833 | 5.900 | 5.763 | Yes | No |
| **64** | B_0516 | 1.748 | 0.738 | 6.041 | 6.052 | 6.028 | 5.385 | 5.468 | 5.298 | Yes | Yes |
| **65** | B_0029 | 1.746 | 1.000 | 1.482 | 1.529 | 1.442 | 2.305 | 2.392 | 2.209 | Yes | No |
| **66** | B_0138 | 1.745 | 0.641 | 2.531 | 2.564 | 2.498 | 1.076 | 1.141 | 1.009 | No | No |
| **67** | B_1005 | 1.744 | 0.731 | 7.294 | 7.332 | 7.264 | 7.332 | 7.441 | 7.218 | Yes | Yes |
| **68** | B_0200 | 1.735 | 0.956 | 3.279 | 3.303 | 3.256 | 3.271 | 3.392 | 3.149 | Yes | Yes |
| **69** | B_0128 | 1.734 | 0.841 | 2.121 | 2.148 | 2.093 | 1.928 | 1.986 | 1.872 | No | No |
| **70** | B_0481 | 1.733 | 0.495 | 5.051 | 5.072 | 5.033 | 6.038 | 6.133 | 5.946 | Yes | Yes |
| **71** | B_0151 | 1.730 | 0.918 | 3.596 | 3.612 | 3.574 | 5.407 | 5.474 | 5.346 | No | Yes |
| **72** | B_0525 | 1.728 | 1.137 | 6.091 | 6.112 | 6.068 | 5.148 | 5.261 | 5.029 | Yes | Yes |
| **73** | B_0538 | 1.727 | 0.447 | 6.036 | 6.046 | 6.017 | 6.021 | 6.126 | 5.917 | Yes | Yes |
| **74** | B_0953 | 1.727 | 0.924 | 6.333 | 6.352 | 6.318 | 6.162 | 6.229 | 6.097 | No | No |
| **75** | B_0201 | 1.725 | 0.526 | 3.323 | 3.342 | 3.306 | 3.325 | 3.463 | 3.182 | No | Yes |
| **76** | B_0555 | 1.721 | 0.757 | 6.393 | 6.411 | 6.378 | 6.307 | 6.343 | 6.267 | No | No |
| **77** | B_0598 | 1.720 | 1.062 | 4.620 | 4.641 | 4.598 | 5.461 | 5.540 | 5.383 | No | Yes |
| **78** | B_0004 | 1.720 | 1.150 | 1.502 | 1.560 | 1.447 | 1.523 | 1.658 | 1.389 | Yes | Yes |
| **79** | B_0084 | 1.719 | 0.712 | 2.477 | 2.501 | 2.454 | 2.467 | 2.559 | 2.378 | Yes | Yes |
| **80** | B_0518 | 1.716 | 1.005 | 6.140 | 6.156 | 6.117 | 5.412 | 5.510 | 5.311 | No | Yes |
| **81** | B_0443 | 1.709 | 0.767 | 5.650 | 5.677 | 5.630 | 4.975 | 5.035 | 4.908 | No | No |
| **82** | B_0487 | 1.703 | 1.099 | 5.843 | 5.884 | 5.802 | 6.307 | 6.408 | 6.208 | No | Yes |
| **83** | B_0898 | 1.701 | 0.794 | 5.873 | 5.918 | 5.823 | 8.556 | 8.624 | 8.490 | No | No |
| **84** | B_0434 | 1.697 | 1.071 | 5.432 | 5.458 | 5.409 | 4.630 | 4.727 | 4.534 | Yes | No |

**Table SI 10 - 100 first UF COSY VIP buckets from OPLS-DA (Part 3/3)**

| UF COSY VIP Ranking | Nom | V1 | V2 | Center F1 [ppm] | Start $^1$H F1 [ppm] | End $^1$H F1 [ppm] | Center F2 [ppm] | Start $^1$H F2 [ppm] | End $^1$H F2 [ppm] | Potential correspondence in F1 | Potential correspondence in F2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **85** | B_0268 | 1.691 | 0.632 | 4.422 | 4.444 | 4.400 | 2.063 | 2.127 | 1.995 | Yes | No |
| **86** | B_0203 | 1.691 | 0.641 | 3.427 | 3.443 | 3.404 | 3.406 | 3.507 | 3.306 | Yes | Yes |
| **87** | B_0503 | 1.688 | 0.849 | 6.036 | 6.050 | 6.019 | 5.558 | 5.622 | 5.488 | Yes | Yes |
| **88** | B_0635 | 1.687 | 0.901 | 2.304 | 2.328 | 2.280 | 5.601 | 5.689 | 5.512 | No | Yes |
| **89** | B_0982 | 1.687 | 0.422 | 7.036 | 7.049 | 7.026 | 7.165 | 7.197 | 7.129 | No | Yes |
| **90** | B_0895 | 1.685 | 0.816 | 5.705 | 5.722 | 5.681 | 8.394 | 8.444 | 8.338 | No | No |
| **91** | B_1002 | 1.682 | 1.136 | 7.234 | 7.262 | 7.212 | 7.257 | 7.357 | 7.156 | Yes | Yes |
| **92** | B_0500 | 1.682 | 0.974 | 5.888 | 5.916 | 5.859 | 5.887 | 5.943 | 5.828 | No | No |
| **93** | B_0043 | 1.681 | 0.809 | 1.507 | 1.537 | 1.475 | 1.324 | 1.368 | 1.279 | Yes | Yes |
| **94** | B_0269 | 1.680 | 0.982 | 4.229 | 4.261 | 4.200 | 1.189 | 1.240 | 1.136 | Yes | No |
| **95** | B_0035 | 1.679 | 1.163 | 2.116 | 2.147 | 2.081 | 1.448 | 1.518 | 1.372 | No | Yes |
| **96** | B_0784 | 1.678 | 1.038 | 5.853 | 5.874 | 5.829 | 2.117 | 2.238 | 2.002 | No | No |
| **97** | B_0018 | 1.677 | 1.004 | 1.338 | 1.356 | 1.316 | 1.486 | 1.533 | 1.437 | Yes | Yes |
| **98** | B_1019 | 1.676 | 0.472 | 7.482 | 7.498 | 7.466 | 7.364 | 7.386 | 7.348 | No | No |
| **99** | B_0149 | 1.676 | 1.231 | 3.536 | 3.568 | 3.504 | 5.401 | 5.546 | 5.257 | No | Yes |
| **100** | B_0546 | 1.671 | 0.939 | 6.091 | 6.114 | 6.074 | 5.881 | 5.923 | 5.835 | Yes | No |

**Table SI 11 - 100 first SYMAPS HSQC VIP buckets from OPLS-DA (Part 1/3)**

| HSQC VIP Ranking | Bucket | V1 | V2 | Center F1 [ppm] | Start $^{13}$C F1 [ppm] | End $^{13}$C F1 [ppm] | Center F2 [ppm] | Start $^{1}$H F2 [ppm] | End $^{1}$H F2 [ppm] | Potential correspondence in $^{13}$C folded | Potential correspondence in $^{1}$H |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | B338 | 1.527 | 0.355 | 19.360 | 19.762 | 18.978 | 6.605 | 6.607 | 6.601 | Yes | Yes |
| 2 | B352 | 1.491 | 0.424 | 25.928 | 26.622 | 25.282 | 7.301 | 7.307 | 7.294 | Yes | Yes |
| 3 | B411 | 1.488 | 0.429 | 21.708 | 22.239 | 21.215 | 7.276 | 7.285 | 7.267 | Yes | Yes |
| 4 | B355 | 1.485 | 0.414 | 18.504 | 18.978 | 17.998 | 7.161 | 7.168 | 7.154 | Yes | Yes |
| 5 | B290 | 1.481 | 0.439 | 38.111 | 38.737 | 37.520 | 5.073 | 5.092 | 5.055 | Yes | Yes |
| 6 | B291 | 1.478 | 0.432 | 38.111 | 38.694 | 37.477 | 5.531 | 5.551 | 5.511 | Yes | Yes |
| 7 | B353 | 1.475 | 0.398 | 25.864 | 26.524 | 25.217 | 7.287 | 7.293 | 7.282 | Yes | Yes |
| 8 | B413 | 1.469 | 0.361 | 21.772 | 22.300 | 21.281 | 7.185 | 7.190 | 7.179 | Yes | Yes |
| 9 | B35 | 1.458 | 0.434 | 19.202 | 19.885 | 18.517 | 2.472 | 2.495 | 2.448 | Yes | Yes |
| 10 | B327 | 1.458 | 0.598 | 8.541 | 9.440 | 7.611 | 6.326 | 6.339 | 6.313 | Yes | No |
| 11 | B20 | 1.454 | 0.622 | 33.130 | 34.352 | 31.932 | 2.304 | 2.362 | 2.246 | Yes | No |
| 12 | B3 | 1.454 | 0.571 | 16.759 | 17.844 | 15.684 | 1.679 | 1.731 | 1.628 | Yes | Yes |
| 13 | B326 | 1.454 | 0.568 | 9.715 | 10.387 | 9.015 | 6.263 | 6.276 | 6.250 | Yes | No |
| 14 | B350 | 1.453 | 0.443 | 29.259 | 29.921 | 28.647 | 7.245 | 7.253 | 7.237 | Yes | No |
| 15 | B65 | 1.453 | 0.528 | 16.790 | 17.709 | 15.846 | 1.598 | 1.619 | 1.576 | Yes | Yes |
| 16 | B313 | 1.447 | 0.520 | 8.858 | 9.375 | 8.297 | 6.001 | 6.008 | 5.994 | Yes | Yes |
| 17 | B314 | 1.446 | 0.601 | 8.890 | 9.407 | 8.362 | 5.979 | 5.988 | 5.970 | Yes | No |
| 18 | B357 | 1.444 | 0.603 | 28.561 | 29.137 | 27.961 | 7.440 | 7.456 | 7.424 | Yes | Yes |
| 19 | B296 | 1.440 | 0.551 | 28.022 | 28.774 | 27.319 | 5.129 | 5.135 | 5.123 | Yes | Yes |
| 20 | B200 | 1.438 | 0.302 | 36.525 | 37.157 | 35.923 | 1.353 | 1.360 | 1.346 | Yes | Yes |
| 21 | B312 | 1.437 | 0.557 | 13.840 | 14.503 | 13.196 | 6.019 | 6.031 | 6.008 | Yes | Yes |
| 22 | B21 | 1.430 | 0.633 | 34.050 | 34.668 | 33.458 | 2.120 | 2.136 | 2.105 | Yes | No |
| 23 | B6 | 1.426 | 0.603 | 19.075 | 19.745 | 18.415 | 1.752 | 1.823 | 1.681 | Yes | Yes |
| 24 | B117 | 1.425 | 0.594 | 34.685 | 35.088 | 34.243 | 1.475 | 1.484 | 1.465 | No | No |
| 25 | B421 | 1.424 | 0.201 | 53.118 | 53.881 | 52.340 | 0.831 | 0.836 | 0.826 | No | No |
| 26 | B331 | 1.421 | 0.636 | 19.424 | 20.121 | 18.717 | 6.685 | 6.696 | 6.674 | Yes | Yes |
| 27 | B337 | 1.421 | 0.631 | 19.424 | 19.860 | 18.978 | 6.590 | 6.595 | 6.583 | Yes | Yes |
| 28 | B138 | 1.417 | 0.565 | 50.929 | 51.543 | 50.297 | 1.690 | 1.698 | 1.683 | Yes | Yes |
| 29 | B142 | 1.417 | 0.651 | 57.750 | 58.960 | 56.529 | 1.752 | 1.770 | 1.733 | Yes | Yes |
| 30 | B339 | 1.413 | 0.588 | 31.893 | 32.501 | 31.260 | 6.606 | 6.616 | 6.596 | Yes | Yes |
| 31 | B38 | 1.412 | 0.705 | 24.659 | 25.567 | 23.725 | 2.070 | 2.106 | 2.033 | Yes | No |
| 32 | B51 | 1.407 | 0.688 | 16.886 | 17.628 | 16.143 | 1.947 | 1.971 | 1.922 | Yes | No |
| 33 | B214 | 1.400 | 0.471 | 66.190 | 66.720 | 65.680 | 1.536 | 1.538 | 1.533 | No | Yes |
| 34 | B311 | 1.397 | 0.689 | 13.776 | 14.307 | 13.262 | 6.043 | 6.053 | 6.032 | Yes | Yes |
| 35 | B134 | 1.394 | 0.653 | 49.882 | 51.045 | 48.739 | 1.754 | 1.772 | 1.735 | Yes | Yes |
| 36 | B58 | 1.390 | 0.661 | 21.740 | 22.757 | 20.678 | 1.445 | 1.460 | 1.431 | Yes | Yes |
| 37 | B15 | 1.386 | 0.721 | 30.560 | 31.353 | 29.775 | 1.378 | 1.433 | 1.323 | Yes | Yes |
| 38 | B304 | 1.384 | 0.711 | 13.808 | 14.283 | 13.363 | 6.066 | 6.078 | 6.055 | Yes | Yes |
| 39 | B317 | 1.383 | 0.714 | 11.111 | 11.563 | 10.649 | 5.805 | 5.811 | 5.799 | Yes | No |
| 40 | B23 | 1.382 | 0.732 | 22.977 | 23.778 | 22.147 | 2.987 | 3.021 | 2.952 | No | Yes |
| 41 | B423 | 1.382 | 0.609 | 59.146 | 59.904 | 58.409 | 0.913 | 0.918 | 0.908 | Yes | No |
| 42 | B201 | 1.379 | 0.683 | 47.947 | 48.335 | 47.564 | 1.341 | 1.346 | 1.335 | No | Yes |

**Table SI 12 - 100 first SYMAPS HSQC buckets from OPLS-DA (Part 2/3)**

| HSQC VIP Ranking | Bucket | V1 | V2 | Center F1 [ppm] | Start $^{13}$C F1 [ppm] | End $^{13}$C F1 [ppm] | Center F2 [ppm] | Start $^{1}$H F2 [ppm] | End $^{1}$H F2 [ppm] | Potential correspondence in $^{13}$C folded | Potential correspondence in $^{1}$H |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 43 | B14 | 1.373 | 0.747 | 28.720 | 29.512 | 27.934 | 1.353 | 1.406 | 1.299 | Yes | Yes |
| 44 | B365 | 1.372 | 0.757 | 66.317 | 67.517 | 65.100 | 10.053 | 10.115 | 9.989 | No | Yes |
| 45 | B318 | 1.370 | 0.707 | 11.175 | 11.563 | 10.812 | 5.793 | 5.798 | 5.788 | Yes | No |
| 46 | B433 | 1.369 | 0.612 | 20.661 | 21.431 | 19.890 | 0.947 | 0.955 | 0.938 | Yes | No |
| 47 | B300 | 1.366 | 0.757 | 18.789 | 20.400 | 17.193 | 5.280 | 5.299 | 5.261 | Yes | Yes |
| 48 | B224 | 1.364 | 0.742 | 37.572 | 38.082 | 37.041 | 1.275 | 1.286 | 1.265 | Yes | Yes |
| 49 | B310 | 1.364 | 0.755 | 13.808 | 14.438 | 13.196 | 6.089 | 6.097 | 6.081 | Yes | Yes |
| 50 | B330 | 1.361 | 0.764 | 19.392 | 20.056 | 18.717 | 6.567 | 6.582 | 6.551 | Yes | No |
| 51 | B334 | 1.361 | 0.723 | 19.551 | 19.991 | 19.076 | 6.526 | 6.532 | 6.519 | Yes | No |
| 52 | B42 | 1.360 | 0.723 | 24.722 | 25.514 | 23.936 | 2.221 | 2.252 | 2.189 | Yes | No |
| 53 | B315 | 1.360 | 0.726 | 6.542 | 7.349 | 5.749 | 5.833 | 5.846 | 5.821 | Yes | No |
| 54 | B111 | 1.359 | 0.729 | 38.587 | 39.097 | 38.105 | 1.508 | 1.513 | 1.503 | Yes | No |
| 55 | B380 | 1.357 | 0.694 | 25.293 | 26.046 | 24.544 | 2.973 | 2.983 | 2.962 | Yes | Yes |
| 56 | B238 | 1.355 | 0.643 | 39.793 | 40.510 | 39.084 | 1.368 | 1.377 | 1.359 | No | No |
| 57 | B120 | 1.354 | 0.716 | 22.121 | 22.768 | 21.445 | 1.534 | 1.560 | 1.508 | Yes | Yes |
| 58 | B275 | 1.354 | 0.320 | 30.592 | 31.311 | 29.923 | 1.733 | 1.737 | 1.729 | Yes | Yes |
| 59 | B141 | 1.351 | 0.751 | 58.258 | 59.895 | 56.592 | 1.688 | 1.712 | 1.665 | Yes | Yes |
| 60 | B54 | 1.348 | 0.784 | 16.981 | 17.979 | 16.008 | 1.760 | 1.787 | 1.733 | Yes | Yes |
| 61 | B299 | 1.348 | 0.761 | 28.212 | 28.507 | 27.913 | 5.078 | 5.088 | 5.069 | Yes | Yes |
| 62 | B356 | 1.345 | 0.772 | 19.836 | 20.285 | 19.337 | 7.108 | 7.113 | 7.103 | Yes | Yes |
| 63 | B303 | 1.344 | 0.784 | 11.016 | 11.640 | 10.393 | 5.768 | 5.779 | 5.758 | Yes | No |
| 64 | B60 | 1.339 | 0.780 | 24.722 | 25.349 | 24.107 | 1.321 | 1.333 | 1.309 | Yes | Yes |
| 65 | B414 | 1.338 | 0.631 | 43.378 | 43.982 | 42.768 | 0.902 | 0.907 | 0.896 | No | No |
| 66 | B267 | 1.336 | 0.770 | 53.784 | 54.360 | 53.199 | 1.705 | 1.713 | 1.696 | No | Yes |
| 67 | B298 | 1.326 | 0.817 | 28.180 | 28.804 | 27.557 | 5.101 | 5.111 | 5.090 | Yes | Yes |
| 68 | B39 | 1.325 | 0.691 | 26.975 | 27.724 | 26.198 | 2.086 | 2.120 | 2.052 | Yes | No |
| 69 | B46 | 1.325 | 0.764 | 16.790 | 17.466 | 16.143 | 2.031 | 2.055 | 2.008 | Yes | No |
| 70 | B13 | 1.323 | 0.849 | 24.595 | 25.429 | 23.746 | 1.735 | 1.875 | 1.594 | Yes | Yes |
| 71 | B370 | 1.323 | 0.667 | 48.105 | 49.010 | 47.231 | 3.271 | 3.281 | 3.261 | No | Yes |
| 72 | B336 | 1.321 | 0.627 | 19.614 | 20.056 | 19.141 | 6.544 | 6.549 | 6.538 | Yes | No |
| 73 | B405 | 1.318 | 0.745 | 26.911 | 27.430 | 26.402 | 3.569 | 3.577 | 3.561 | Yes | No |
| 74 | B297 | 1.316 | 0.800 | 28.117 | 28.745 | 27.468 | 5.117 | 5.122 | 5.112 | Yes | Yes |
| 75 | B109 | 1.315 | 0.765 | 38.555 | 39.001 | 38.067 | 1.495 | 1.502 | 1.488 | Yes | No |
| 76 | B320 | 1.313 | 0.684 | 11.016 | 11.465 | 10.551 | 5.729 | 5.734 | 5.724 | Yes | No |
| 77 | B129 | 1.312 | 0.617 | 61.145 | 61.889 | 60.393 | 1.478 | 1.484 | 1.471 | No | No |
| 78 | B319 | 1.311 | 0.608 | 11.111 | 11.498 | 10.747 | 5.745 | 5.749 | 5.741 | Yes | Yes |
| 79 | B412 | 1.311 | 0.822 | 21.835 | 22.370 | 21.298 | 7.231 | 7.237 | 7.225 | Yes | No |
| 80 | B345 | 1.310 | 0.781 | 26.277 | 26.883 | 25.642 | 6.842 | 6.859 | 6.825 | Yes | Yes |
| 81 | B393 | 1.308 | 0.804 | 26.753 | 27.272 | 26.244 | 3.753 | 3.765 | 3.741 | Yes | No |
| 82 | B43 | 1.305 | 0.625 | 16.790 | 17.360 | 16.255 | 2.177 | 2.197 | 2.157 | Yes | No |
| 83 | B78 | 1.305 | 0.833 | 10.318 | 11.358 | 9.281 | 1.697 | 1.750 | 1.644 | Yes | Yes |
| 84 | B37 | 1.300 | 0.874 | 24.722 | 25.672 | 23.778 | 1.996 | 2.031 | 1.961 | Yes | No |

**Table SI 13 - 100 first SYMAPS HSQC VIP buckets from OPLS-DA (Part 3/3)**

| HSQC VIP Ranking | Bucket | V1 | V2 | Center F1 [ppm] | Start $^{13}$C F1 [ppm] | End $^{13}$C F1 [ppm] | Center F2 [ppm] | Start $^{1}$H F2 [ppm] | End $^{1}$H F2 [ppm] | Potential correspondence in $^{13}$C folded | Potential correspondence in $^{1}$H |
|---|---|---|---|---|---|---|---|---|---|---|---|
| OP85 | B188 | 1.300 | 0.819 | 62.224 | 62.789 | 61.710 | 1.245 | 1.255 | 1.235 | No | No |
| 86 | B63 | 1.300 | 0.824 | 22.152 | 23.000 | 21.299 | 1.342 | 1.358 | 1.326 | Yes | Yes |
| 87 | B410 | 1.297 | 0.876 | 18.916 | 19.446 | 18.379 | 5.237 | 5.259 | 5.214 | Yes | No |
| 88 | B61 | 1.293 | 0.880 | 24.786 | 25.457 | 24.134 | 1.258 | 1.290 | 1.225 | Yes | Yes |
| 89 | B123 | 1.289 | 0.880 | 47.661 | 48.552 | 46.744 | 1.500 | 1.512 | 1.489 | No | No |
| 90 | B280 | 1.288 | 0.827 | 29.608 | 30.052 | 29.213 | 1.641 | 1.658 | 1.624 | Yes | Yes |
| 91 | B435 | 1.279 | 0.849 | 26.689 | 27.220 | 26.147 | 0.957 | 0.968 | 0.945 | Yes | No |
| 92 | B102 | 1.278 | 0.607 | 35.287 | 35.720 | 34.861 | 1.741 | 1.750 | 1.731 | No | Yes |
| 93 | B259 | 1.278 | 0.624 | 66.571 | 66.974 | 66.172 | 1.728 | 1.732 | 1.724 | No | Yes |
| 94 | B199 | 1.277 | 0.837 | 45.884 | 46.485 | 45.290 | 1.250 | 1.261 | 1.239 | No | No |
| 95 | B53 | 1.277 | 0.893 | 16.759 | 17.736 | 15.819 | 1.894 | 1.917 | 1.870 | Yes | No |
| 96 | B32 | 1.276 | 0.819 | 26.626 | 27.303 | 25.935 | 3.457 | 3.485 | 3.430 | Yes | No |
| 97 | B281 | 1.275 | 0.900 | 28.434 | 28.890 | 27.986 | 1.638 | 1.660 | 1.616 | Yes | Yes |
| 98 | B409 | 1.273 | 0.853 | 18.853 | 19.327 | 18.379 | 5.312 | 5.326 | 5.299 | Yes | No |
| 99 | B418 | 1.272 | 0.847 | 43.346 | 43.842 | 42.815 | 0.926 | 0.929 | 0.922 | No | No |
| 100 | B258 | 1.271 | 0.764 | 66.634 | 67.227 | 66.088 | 1.735 | 1.737 | 1.732 | No | Yes |

**Table SI 14 - Global data for all $^{1}$H NMR buckets**: Minimum (MIN), maximum (MAX) and mean SNR were measured for each sample group. To quantify the proportion of data under the limit of detection and quantification, number of buckets with a Signal to Noise Ratio (SNR) under 3 and 10 have been counted.

| | Blank_Ctrl | Blank_Sponge | Control_T1 | Control_T2 | QC | Sponge_T1 | Sponge_T2 |
|---|---|---|---|---|---|---|---|
| **MIN** | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| **MAX** | 11471 | 17332 | 20872 | 17698 | 24118 | 21244 | 21870 |
| **MEAN** | 117 | 195 | 768 | 706 | 775 | 769 | 802 |
| Number of buckets with a SNR<3 | 256 | 167 | 10 | 16 | 12 | 15 | 12 |
| **% SNR<3** | 28% | 19% | 1% | 2% | 1% | 2% | 1% |
| Number of buckets with a SNR<10 | 400 | 340 | 63 | 65 | 63 | 78 | 65 |
| **% SNR<10** | 44% | 38% | 7% | 7% | 7% | 9% | 7% |
| Total buckets | 901 | 901 | 901 | 900 | 900 | 901 | 901 |

**Table SI 15 - Global data for all $^{13}C$ NMR buckets:** Minimum (MIN), maximum (MAX) and mean SNR were measured for each sample group. To quantify the proportion of data under the limit of detection and quantification, number of buckets with a Signal to Noise Ratio (SNR) under 3 and 10 have been counted.

| | **Blank_Ctrl** | **Blank_Sponge** | **Control_T1** | **Control_T2** | **QC** | **Sponge_T1** | **Sponge_T2** |
|---|---|---|---|---|---|---|---|
| **MIN** | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| **MAX** | 27 | 24 | 253 | 228 | 304 | 272 | 270 |
| **MEAN** | 2 | 2 | 17 | 17 | 20 | 19 | 20 |
| Number of buckets with a SNR<3 | 380 | 338 | 79 | 72 | 77 | 71 | 79 |
| **% SNR <3** | 92% | 82% | 19% | 18% | 19% | 17% | 19% |
| Number of buckets with a SNR<10 | 396 | 395 | 267 | 262 | 247 | 263 | 258 |
| **% SNR <10** | 96% | 96% | 65% | 64% | 60% | 64% | 63% |
| Total buckets | 411 | 411 | 411 | 411 | 411 | 411 | 411 |