CENTENARY OF THE OIV
45th WORLD CONGRESS OF VINE AND WINE
14-18 OCTOBER 2024 FRANCE-DIJON
OIV
Available on IVES Conference Series vine & wine
DOI : https://doi.org/10.58233/IMGGSIME

Development of a standardized method for metabolite analysis by NMR to assess wine authenticity

Guillaume Leleu[1], Patricia Homobono Brito de Moura[1,2], Grégory Da Costa[1,2], Cléo Berrehail[2,3], Jean-Claude Boulet[4,5], Rémi Butelle[1], Catherine Deborde[2,3#], Matthieu Dubernet[6], Laetitia Gaillard[7], Antoine Galvan[7], Karen Gaudin[1,8], Alexandra Gossé[9], Markus Herderich[10,11,12], Daniel Jacob[2,3#], Christine Le Guernevé[4], Elodie Milani Crouzet[4,5], Annick Moing[2,3], Sophie Rosset[7], Josep Valls-Fonayet[1,2], Flynn Watson[11,12], Tristan Richard[1,2]

[1] Univ. Bordeaux, Bordeaux INP, INRAE, Bordeaux Sciences Agro, OENO, UMR 1366, ISVV, F-33140 Villenave-d'Ornon, France
[2] MetaboHUB-Bordeaux, F-33140 Villenave-d'Ornon, France
[3] INRAE, Univ. Bordeaux, Biologie du Fruit et Pathologie, UMR 1332, Centre INRAE de Nouvelle Aquitaine Bordeaux, Villenave-d'Ornon, F-33140, France
[4] INRAE, Institut Agro Montpellier, Univ. Montpellier, SPO, UMR 1083, F-34060 Montpellier, France
[5] INRAE, PROBE research infrastructure, PFP polyphenols analysis facility, F-34060 Montpellier, France
[6] Laboratoire Dubernet, 35 rue de Combe du Meunier, 11100 Montredon-des-Corbières, France
[7] Service Commun des Laboratoires, 3 avenue du Dr. Albert Schweitzer, F-33600 Pessac, France
[8] Univ. Bordeaux, Laboratoire de Chimie Analytique, F-33000, Bordeaux, France
[9] Les Grands Chais De France, 1 rue de la Division Leclerc, F-67290 Petersbach, France
[10] The University of Adelaide, Waite Research Institute, Urrbrae, 5064, SA, Australia
[11] The Australian Wine Research Institute (AWRI), Urrbrae, 5064, SA, Australia
[12] Metabolomics Australia, Urrbrae, 5064, SA, Australia
[#] Actual address: INRAE, UR 1268 Biopolymères Interactions Assemblages, F-44300 Nantes, France, INRAE, PROBE Research Infrastructure, BIBS facility, Institute, F-44300 Nantes, France

Abstract. The wine sector generates a considerable amount of wealth but is facing a perpetual problem of fraud. Wine counterfeiting is one of the oldest and most common cases of food fraud worldwide. Therefore, the authenticity and traceability of wine are major concerns for both industry and consumers. To address these issues, robust and reliable analysis and control methods are necessary. Several methods have been developed, ranging from simple organoleptic tests to more advanced methodologies such as isotopic techniques or residual radioactivity measurements. However, with counterfeiting becoming increasingly sophisticated, more complex methodologies are needed for their detection. In this context, innovative tools for the analysis of metabolites and other small organic molecules can offer new perspectives for ensuring wine authenticity. In this work, we propose the development and validation of an open metabolomic approach based on ^{1}H NMR to assess wine authenticity, using standardized, flexible, and accessible analytical and data processing protocols. Additionally, combinatorial approaches using orthogonal techniques e.g. NMR, high-resolution mass spectrometry, and data fusion are presented. These efforts represent a significant advancement towards establishing a standardized and open methodology for certifying wine authenticity using ^{1}H NMR metabolomics and establish new metabolomic approaches for wine authentication and quality control.

1. Introduction

Wine is a product of great economic value. Wine fraud and counterfeiting are major challenges for the wine industry, threatening not only the quality and reputation of the products but also consumer trust. In this context, the development of reliable, robust, and reproducible methods to ensure wine authenticity is essential [1]. ^{1}H NMR-based metabolomics applied to wine has experienced considerable interest over the past decade, establishing quantitative ^{1}H NMR as an useful technique for wine analysis [2]. ^{1}H NMR spectroscopy provides a chemical profile that can be used to classify wines based on their geographical origin, grape variety, and vintages [2].

Currently, Bruker, a leading manufacturer of NMR equipment, offers a commercial service combining both targeted and untargeted approaches to ensure wine authenticity through comparison of analytical data with reference values in a proprietary database [3]. This has motivated the development of an official OIV method for quantifying six common wine metabolites: glucose, malic acid, acetic acid, fumaric acid, shikimic acid, and sorbic acid [4]. In the literature, various methods for analyzing wine samples using ^{1}H NMR have been developed [2]. However, to be effectively utilized by official control laboratories as well as accredited analytical service providers, these methods must undergo a rigorous validation process, ensuring transparency and reproducibility from sample preparation through data acquisition and processing to the development of reference databases [5]. To promote uptake of ^{1}H NMR, our consortium unites several laboratories with the aim of developing a standardized ^{1}H NMR approach with validated procedures, user-friendly, and open access data treatment, and harmonized metadata and spectra collection. In this article, we present the initial steps in the development of this standardized NMR approach.

While untargeted metabolomics approaches are highly promising theoretically, they pose significant practical challenges such as the changes in composition [5-7]. In this study, we also seek to harness the combined potential of ^{1}H NMR and high resolution mass spectrometry (HRMS) to address the issue of wine origin, evaluating a data fusion process between untargeted NMR and HRMS analyses.

2. Materials and methods

2.1. Reagents and standards

Deuterium oxide and 3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt (TSP) were purchased from Eurisotop (France). High-purity standards were used for the external standardization and spiking experiments: citric acid (≥99.5%), succinic acid (≥99.5 %), dimethyl malonate (≥99.5%), fumaric acid (≥98.0%), maleic acid (≥99.5%), benzoic acid, calcium formate (≥99.5%), gallic acid (97.5-102.5%), propionic acid (≥99.5%), all obtained from Sigma-Aldrich (Germany). Reagents used for the buffer solution, including sodium phosphate monobasic, phosphoric acid, and sodium azide, were purchased from Sigma-Aldrich.

2.2. ^{1}H NMR analysis

2.2.1. Deuterated buffer solution

A basic solution is prepared by dissolving 12.0 g of monobasic sodium phosphate in 100 mL of D_2O. An acidic solution is prepared by diluting 1.72 mL of phosphoric acid in 25 mL of D_2O. The buffer solution, with a total exceeding 100 mL, is obtained by mixing basic and acidic solutions in a 9:1 ratio. The apparent pH is then adjusted at 2.60 by adding the acidic solution as needed. Finally, 60 mg of TSP and 18 mg of sodium azide are dissolved in 100 mL of the buffer solution.

2.2.2. External standard for quantification

Various quantification standards, including benzoic acid, citric acid, succinic acid, gallic acid, propionic acid, maleic acid, and calcium formate, were evaluated prior to this study to determine the composition of the quantification standard and the quality control. In brief, stock solutions and different mixtures were analyzed, and the response of each compound was assessed.

Individual stock solutions of quantitation compounds were prepared using ultrapure water (Milli-Q, Merck-Milllipore, France) to a total volume of 50 mL. The compounds were: citric acid (7.0 g/L), dimethyl malonate (2.5 g/L), and succinic acid (5.0 g/L). The exact masses were recorded.

An external standard for quantification (Qref) was prepared by mixing 25 mL of the citric acid solution, 20 mL of the dimethylmalonic acid solution, and 5 mL of deuterated buffer. Similarly, a quality control solution (Qcon), used to validate measurements, was prepared by combining 20 mL of the citric acid solution, 5 mL of the succinic acid solution, 20 mL of the dimethylmalonic acid solution, and 5 mL of deuterated buffer.

2.2.3. Wine sample preparation

Three classes of wine were assessed in this study: red, white, and sweet wines. Due to sweet wine viscosity, these were diluted 1:1 with ultrapure water. For sample preparation, 1.4 mL of wine was centrifuged at 14,100 *g* for 5 min at room temperature. Then, 900 µL of the supernatant was mixed with 100 µL of deuterated buffer solution in a vial compatible with a semi-automatic titration unit (BTpH, Bruker BioSpin, Germany). The sample pH was adjusted to 3.10 ± 0.02 using 1M HCl or 1M NaOH solutions. Subsequently, 600 µL of the prepared sample was transferred to a 5 mm NMR tube.

For a spiking experiment, a fumaric acid stock solution (3.00 g/L) was prepared. Four solutions were prepared by diluting 1, 2, 3, or 4 mL of the stock solution to 10 mL with ultrapure water. A blank was prepared using ultrapure water only. Spiked wines were prepared using 800 µL of wine supernatant, 100 µL of dilute fumaric acid solution (or the blank), and 100 µL of deuterated buffer. The pH of the samples was then adjusted at pH 3.10 ± 0.02 using the semi-automatic titration unit.

2.2.4. ^{1}H NMR acquisition

Spectra were recorded using a 500 MHz AVANCE III NMR spectrometer (Bruker, France), equipped with a 5-mm ATMA-BBI probe and an autosampler using TopSpin IconNMR 3.0 software (Bruker). Automated tuning, matching, and shimming were performed, along with 90° pulse calibration for each sample. Three pulse sequences were acquired in automation for each wine sample: a water

suppressed 90° pulse and acquire experiment (zgpr), a water and ethanol suppressed, gradient enhanced 1D-presat-NOESY (noesygpps1d), and a water suppressed, gradient enhanced J-Resolved spectrum (jresgpprqf).

2.2.5. Spectrum treatment

An exponential line broadening of 0.3 Hz was applied prior to Fourier transformation. The chemical shift was calibrated to 0.0 ppm using TSP-d_4. Manual phase and baseline corrections were applied as necessary. To ensure spectrum quality, the full width at half maximum (FWHM) of TSP-d_4 was below 1 Hz. In the case that the FWHM was greater than 1 Hz, the sample was reacquired.

A targeted analysis of the spectra was performed using the Simple Mixture Analysis (SMA) plugin in MestReNova 14.0.0 software (Mestrelab Research, Spain). Compound identification followed previous work and single compound spiking experiments. Additionally, a tool is currently being developed to fully automate targeted spectra processing. Compound quantification was performed using the external standard Qref in combination with the PULCON method [8]. For untargeted analysis, spectral bins were extracted with the online tool NMRProcFlow v.1.4 [9].

2.3. UHPLC-HRMS analysis

2.3.1. Wine sample preparation

The wine samples for mass spectrometry analysis were diluted in methanol (3:1), centrifuged at 14,100 *g*, at room temperature, and the supernatant was subsequently analysed.

2.3.2. Mass spectrometry acquisition

Samples separated and analysed using ultra-high-pressure liquid chromatography (UHPLC, Thermo-Fisher Scientific) equipped with a Luna Omega Polar C18 column (50 x 2.1 mm, 1.8 µm) coupled to a Q-Exactive Plus Orbitrap mass spectrometer (Thermo-Fisher Scientific, PAYS). Mass spectrometry analysis was conducted in negative ionization mode with an electrospray ionization (ESI) source. Data was acquired in data-dependent acquisition mode (DDA). The resulting raw data was processed with MS-DIAL software v.4.94, using an MS^1 tolerance of 0.01, an MS^2 tolerance of 0.025, and minimum peak height 30,000.

2.4. Statistical analysis

Statistical analyses were performed using SIMCA 17 software (Sartorius, Germany). The data were centred and scaled to unit variance using SIMCA's default procedure. Principal component analysis (PCA) was conducted to visualize the dataset's distribution. For the variable importance in projection (VIP) selection, orthogonal projection to latent structures discriminant analyses (OPLS-DA) was used [10, 11].

3. Results and discussion

3.1. ^{1}H NMR method development

3.1.1. Sample preparation

Various methods have been developed for preparing wine samples for ^{1}H NMR analysis of wines [2]. In our opinion, the most robust approach is to use both a buffer solution and subsequent pH adjustment of the wines prior to analysis, as this minimizes chemical shift variations between samples [12]. For calibration and quantification, adding an internal standard is the simplest method, with TSP-d_4 frequently used as a calibration standard. For quantification purposes, TSP-d_4 or other compounds, such as calcium formate, can be employed [13]. However, potential interactions with the matrix, such as the residual presence of formic acid in wines, can compromise measurement quality. Consequently, the use of an external standard is often a preferred solution [3], where various compounds can be utilized [14]. Teipel et al. have proposed an efficient method involving a mixture of reference standards to ensure accurate quantification and quality control [14]. In our research, we have opted to adopt their approach to establish a quantification reference solution (Qref) and quality control solution (QCon).

3.1.2. ^{1}H NMR experiments

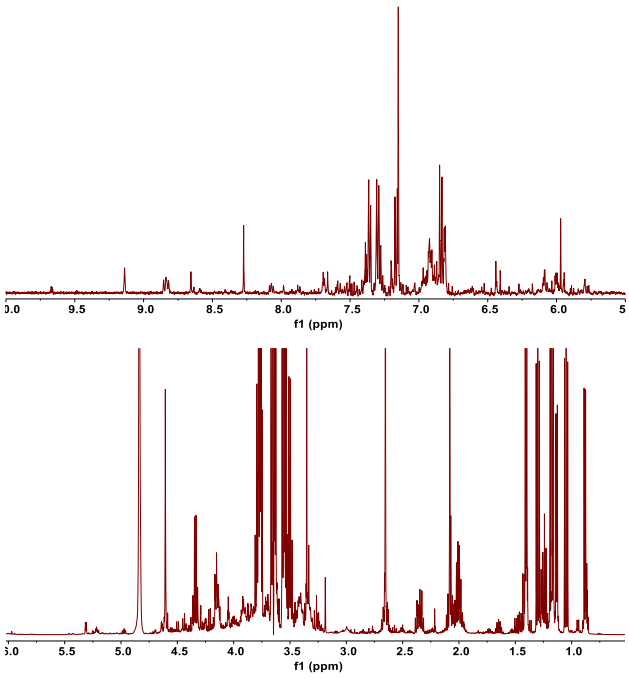

Figure 1. Typical water and ethanol suppressed ^{1}H-1D-presat-NOESY spectrum of a red wine sample.

To analyze wines, three sequences are generally used [12, 14]: (1) a water suppressed 1D ^{1}H 90° pulse and acquire experiment; (2) a water and ethanol suppressed ^{1}H-1D-presat-NOESY, (3) a water suppressed 2D ^{1}H-^{1}H J-resolved (JRES). The first experiment allows for quantification of several major compounds whose signals may be affected by the suppression of ethanol signals (e.g.,

ethanol, glycerol, etc.) [13]. The second sequence allows for the quantification of up to 40 other compounds. The final sequence provides additional information, which can resolve ambiguity in the case of overlap.

3.1.3. Compound identification

Attributing peaks seen in the ^{1}H NMR spectra to compounds is a key step in the data processing process. Attribution, in the first instance, is based on the comparison to previously published studies, and databases. To confirm the attribution of the signals, we used spiking experiments for all compounds listed in Table 1. This process enabled the unambiguous assignment of each compound; this confirmed the identification of 54 compounds in wine (Table 1).

Table 1. Compounds identified in NMR spectra of wine and the details of the spectral feature, which is used for quantitation.

Compound	δ_H (ppm)	Mult.	*J* (Hz)	*N*
Leucine	**0.96**	d	6.2	$2CH_3$
Valine	**0.99**	d	7.4	CH_3
Isoleucine	**1.00**	d	7.0	CH_3
2,3-Butanediol	**1.13**	d	6.2	CH_3
Ethanol	**1.18**	t	7.2	CH_3
Threonine	**1.33**	d	6.7	CH_3
Acetoin	**1.37**	d	7.0	CH_3
Lactic acid	**1.40**	d	7.0	CH_3
Alanine	**1.50**	d	7.2	CH_3
Isoamyl alcohol	**1.65**	m	-	CH
Cadaverine	**1.70**	m	-	$2CH_2$
Putrescine	**1.75**	m	-	$2CH_2$
Arginine	**1.91**	m	-	CH_2
Proline	**1.99**	m	-	CH_2
Ethyl acetate	**2.07**	s	-	CH_3
Acetic acid	**2.08**	s	-	CH_3
Ethanal	**2.23**	d	3.0	CH_3
Pyruvic acid	**2.35**	s	-	CH_3
γ-Aminobutyric acid	**2.50**	t	7.3	CH_2
Succinic acid	**2.65**	s	-	$2CH_2$
Malic acid	**2.89**	dd	16.3; 4.5	CH
Citric acid	**2.96**	d	15.6	CH_2
Choline	**3.19**	s	-	$3CH_3$
Myo-inositol	**3.27**	t	9.7	CH
Methanol	**3.35**	s	-	CH_3
Isobutanol	**3.37**	d	6.7	CH_2
Glycerol	**3.55**	dd	11.8; 6.5	CH_2
Mannitol	**3.86**	dd	11.9; 2.8	CH_2
Fructose	**4.02**	dd	12.8; 1.0	CH_2
Ethyl lactate	**4.21**	q	7.1	CH_2
Arabinose	**4.50**	d	7.7	CH
Tartaric acid	**4.60**	s	-	2CH
Glucose	**4.63**	d	7.9	CH
Xylose	**5.18**	d	3.7	CH
Glucose	**5.23**	d	3.6	CH
Galactose	**5.25**	d	3.7	CH
Galacturonic acid	**5.30**	d	3.7	CH
Glucuronic acid	**5.55**	d	4.0	CH
Sorbic acid	**5.78**	d	15.3	CH
Catechin	**5.99**	d	2.0	CH
Epicatechin	**6.10**	d	2.0	CH
Caffeic acid	**6.33**	d	16.0	CH
Caftaric acid	**6.42**	d	15.9	CH_3
Fumaric acid	**6.75**	s	-	2CH
Shikimic acid	**6.82**	m	-	CH
Tyrosol	**6.85**	d	8.4	2CH
Tyrosine	**6.88**	d	8.4	2CH
Vanillic acid	**6.95**	d	8.8	CH
Gallic acid	**7.16**	s	-	2CH
Phenethyl alcohol	**7.33**	m	-	5CH (3)
Syringic acid	**7.36**	s	-	2CH
Formic acid	**8.27**	s	-	CH
Histidine	**8.65**	d	1.1	CH
Trigonelline	**9.14**	s	-	CH

Mult.: multiplicity; s: singlet; d: doublet; dd: doublet of doubletrs; m: multiplet, q: quintuplet; t: triplet. *J*: coupling constant.

3.1.4. Compound quantification

The quantification step is crucial. In this work, we chose to use an external standard consisting of several compounds. The concentration of the wine sample constituents will be calculated based on the signal areas of these compounds using the PULCON method [8]. This method relies on calculating a response factor, called the PULCON factor, from the spectrum of the external standard sample (Qref). The PULCON factor is given by the following formula [14]:

$$f_{PULCON} = \frac{I \times SW \times M}{SI \times \rho \times N} \quad (1)$$

where I is the area of the selected resonance of the reference standard; SW is the spectral width; M is the reference standard molecular weight; N is the number of protons corresponding to the selected resonance; ρ is the exact mass concentration in the Qref; SI is the size of the real spectrum.

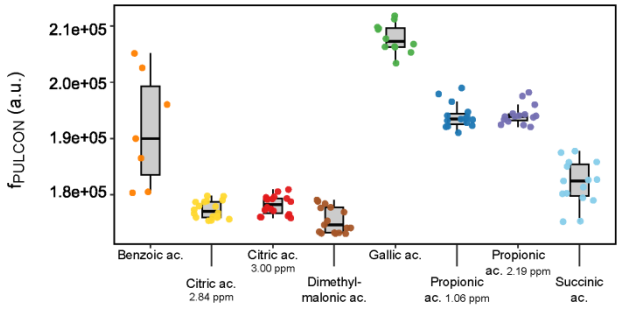

Figure 2. The PULCON factor (f_{PULCON}) calculated for benzoic acid, citric acid, dimethylmalonic acid, gallic acid, propionic acid, and succinic acid across various mixtures and concentrations.

To define the Qref, we assessed various reference standards at different concentrations: benzoic acid, citric acid, dimethylmalonic acid, gallic acid, propionic acid, and succinic acid. Ideally, the PULCON factors obtained for these different compounds should be identical. However, our findings reveal that this is not the case (Fig. 2), indicating that experimental conditions affect the NMR response.

For the standardized method we defined the QRef to consist of two standards: citric acid and dimethylmalonic acid (Table 2). An average PULCON factor was calculated from these standards, which was then used for quantification. For quality control purposes, the relative deviation between the two individuals PULCON factors should remain within 2%.

Table 2. Reference standards used in the QRef and Qcon.

Compound	δ_H (ppm)	Mult.	*J* (Hz)	*N*
Citric acid	**2.84**	d	15.7	2H
Citric acid	**3.00**	d	15.7	2H
Dimethyl-malonate	**1.42**	s	-	6H
Succinic acid	**1.66**	s	-	4H

The concentration of compounds in a wine sample is determined using the following formula [14]:

$$\rho = CF \frac{I \times SW \times M \times NS_{QRef} \times P_1}{SI \times f_{PULCON} \times NS \times f^{dil} \times P_1^{QRef}} \quad (2)$$

where I is the area of the selected resonance of the quantified compound; SW is the spectral width; M is the quantified compound molecular weight; N is the number of protons corresponding to the selected resonance; ρ is the exact mass concentration; SI is the size of the real spectrum, f^{dil} is the wine sample dilution and P_1 et P_1^{QRef} are the respective pulses of the wine and Qref samples, and CF is the correction factor. The CF is further discussed below.

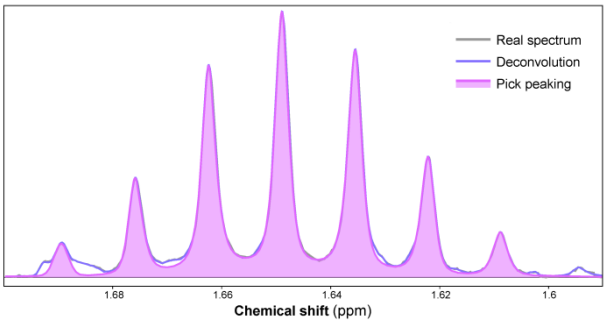

Figure 3. Deconvolution applied to the isopentanol signal in a real wine spectrum using NMRProcFlow.

Accurate determination of compound concentrations hinges on integrating the corresponding signal areas within the NMR spectrum. Due to the inherent complexity of the NMR spectrum, it is crucial to deconvolute and accurately assign all signals of interest (Fig. 3). This process can be accomplished using a range of tools and software [2]. In our project, we selected NMRprocFlow as our primary tool, benefiting from its free online accessibility and user-friendly interface. In the future, a dedicated tool will be developed specifically for the targeted quantification of wine compounds.

Finally, to apply the formula (2), it is essential to evaluate the correction factor (CF), which compensates for various issues related to the acquisition and processing of NMR spectra, such as matrix effects, sequence effects, and data processing. To establish the correction factor for each compound, we undertook spiking experiments in three classes of wine: red, white, and sweet wines. In addition, this approach also enables the assessment of the limit of detection (LOD), limit of quantitation (LOQ), and the working range [15].

Figure 4 illustrates the calculation of CF for fumaric acid for red, white, and sweet wines. For this metabolite, the CF is 1 regardless of the type of wine, demonstrating the generality of the method regardless of the wine analyzed. For other compounds, the CF is not equal to 1 and is attributed to various effects.

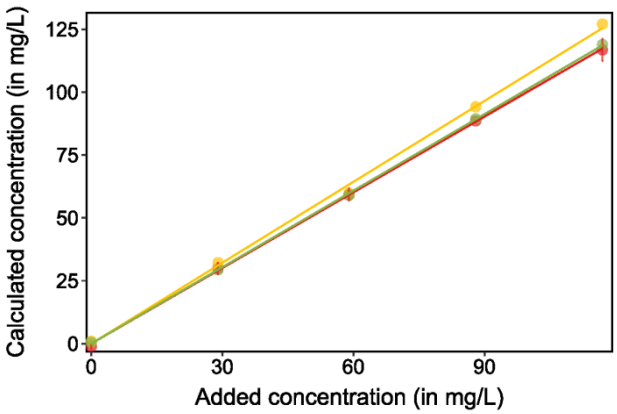

Figure 4. Determination of the correction factor for fumaric acid for different wines (Red: red wine; Green: white wine; Yellow: sweet wine).

3.1.5. Quality assurance

To ensure the accuracy of measurements, areas, PULCON factors, and concentrations of the reference standards in the Qref are automatically documented on a control chart. These values must not deviate by more than 5% from their initial values. For the spectra of wine samples, both the FWHM and the area of the TSP-d_4 signal (internal standard) are consistently measured and recorded. According to Teipel et al., the FWHM should not exceed 1 Hz.

Furthermore, for quality assurance, a control solution (Qcon) is analyzed following each measurement series. This solution includes three compounds: citric acid, dimethylmalonic acid and succinic acid (Table 2). The concentrations of these compounds in the Qcon must not vary by more than ± 5% from their known values.

3.1.6. Interlaboratory assays

Finally, to validate the method and assess its repeatability and reproducibility, an interlaboratory test will be scheduled involving different magnetic fields. This work will help validate and complete the list of compounds already quantified in the method proposed to the OIV.

3.2. Untargeted ^{1}H NMR and MS-based metabolomics analysed by multi-platform data fusion

Determining the authenticity of wines is a complex challenge that often requires the application of multiple complementary analytical techniques. In this regard, the synergistic combination of ^{1}H NMR fingerprinting with UHPLC-HRMS represents an innovative approach. As a proof of concept, we analyzed a collection of wines from three grape varieties, each originating from different areas of Languedoc-Roussillon, using both untargeted ^{1}H NMR and UHPLC-HRMS metabolomics. The data were then compared using principal component analysis (PCA). Additionally, we used a mid-level data fusion process to effectively differentiate between grape varieties and wine origins [16].

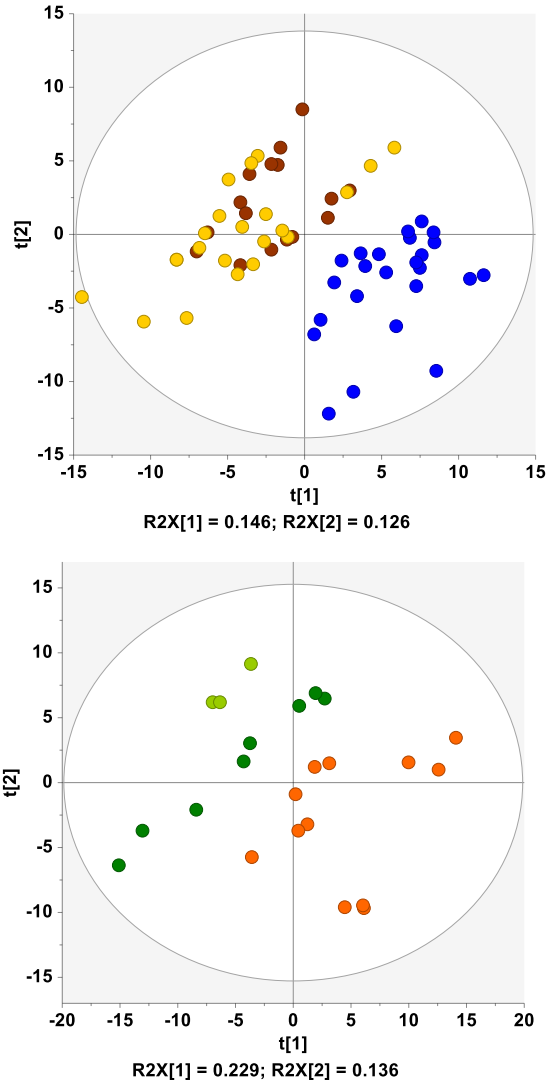

Figure 3. PCA score plot based on ^{1}H NMR fingerprinting. A: classification of wines from different varieties (Blue: Pinot; Red: Cabernet Sauvignon; Yellow: Merlot). B: classification of Merlot wines from different French regions (Orange: Aude; Dark green: Hérault, Green: Gard).

3.2.1. ^{1}H NMR untargeted analysis

The ^{1}H NMR spectra were processed using NMRprocFlow software (version 1.4). The spectra were calibrated using the TSP-d_4 signal to 0.0 ppm, globally baseline-corrected, aligned, divided into uniform spectral bins (0.04 ppm width), and integrated. The resulting data matrix was then exported to SIMCA software for multivariate analysis.

Unsupervised principal component analysis (PCA) was conducted to examine the distribution of the samples without any prior assumptions about their classification. Figure 3A illustrates the distribution of the samples based on grape variety. The PCA score plot shows a distinct classification of Pinot wines but reveals an overlap between Cabernet Sauvignon and Merlot wines. This finding is unexpected, given that NMR is typically known for its ability to distinguish between different grape varieties [2].

To better understand this result, we isolated the Merlot samples and applied a PCA on this subset. Figure 3B evidences the influence of geographical origin on the classification of the wines, with the three origins being clearly separated.

Therefore, untargeted NMR analyses capture a combination of both grape variety and geographical origin.

3.2.2 HRMS untargeted analysis

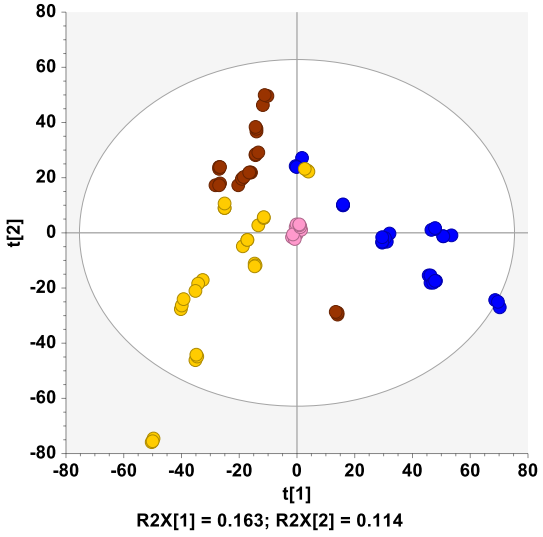

Figure 4. PCA score plot based on HRMS fingerprinting. Classification of wines from different varieties (Blue: Pinot; Red: Cabernet Sauvignon; Yellow: Merlot; Pink: Quality control).

The mass spectra obtained from UHPLC-HRMS analysis were processed using MS-DIAL software. The data were then analyzed with multivariate techniques in SIMCA. Principal component analysis (PCA) was used to explore the distribution of the wine samples. Figure 4 displays the PCA score plot results for wines from different grape varieties, showing a clear discrimination of the three varieties, although some partial overlap among the samples is observed. Untargeted analysis based on mass spectrometry seems to be more effective at differentiating grape varieties than NMR spectroscopy. Similarly, we evaluated the impact of geographical origin on the classification of Merlot wines. In this instance, PCA score plot did not provide a satisfactory classification based on geographical origin (data not shown).

These results confirm the effectiveness of untargeted high resolution mass spectrometry analyses in discriminating wines based on grape variety. However, the findings are more nuanced when it comes to classification based on geographical origin.

3.2.2. Mid-level data fusion

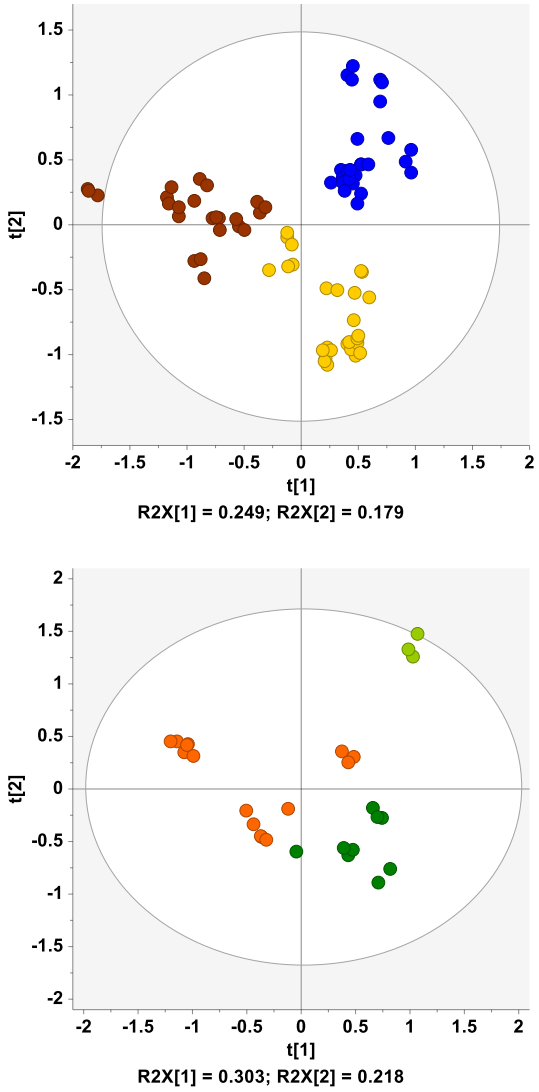

Figure 5. PCA score plot based on mid-level data fusion. A: classification of wines from different varieties (Blue: Pinot; Red: Cabernet Sauvignon; Yellow: Merlot). B: classification of Merlot wines from different French regions (Orange: Aude; Dark green: Hérault, Green: Gard).

To explore the potential synergy between untargeted ^{1}H NMR and HRMS analyses, we utilized a data fusion strategy. This approach integrates data from various analytical sources to enhance model predictability. Specifically, we used a mid-level data fusion approach which involves selecting the most discriminative variables from each data block. Orthogonal projection to latent structures discriminant analysis (OPLS-DA) was conducted on each dataset to identify discriminating features based on their variable importance in projection (VIP). Variables with VIP values greater than 1.5 from each dataset were retained. The different data blocks were then concatenated in SIMCA to appropriately account for the size of each block before further processing.

The data obtained were analyzed using unsupervised PCA. Figure 5A shows the PCA score plot for classifying wines based on grape variety. The wine samples from the three different varieties are correctly classified. Thus, data

fusion allows for the construction of a more effective model for discriminating grape varieties compared to individual approaches. PCA was applied to the dataset corresponding to Merlot wines to classify the samples based on geographical origin. A classification of the wines based on their geographical origin is also achieved (Fig. 5B).

These results underscore the power of integrating multiple analytical approaches to verify wine authenticity. The mid-level data fusion approach effectively integrates two data sources, offering a comprehensive analysis that describes both geographical origin and grape variety.

4. Conclusion

The present study outlines the development and initial validation of ^{1}H NMR and HRMS-based metabolomics approaches for assessing wine authenticity.

A standardized protocol was established for preparing deuterated buffer solution and wine samples, and key ^{1}H NMR acquisition and processing parameters for routine measurements were determined. To enable reliable quantification, various standards were tested to define an external quantification standard (Qref) and a quality control solution (Qcon) to ensure the accuracy of the quantification. The Qref consists of citric acid and dimethyl malonate, while the Qcon includes also succinic acid.

A total of 54 compounds were spiked into wines, facilitating their identification and determination of correction factors which enables their accurate quantification. Additionally, a fully automated tool for quantifying wine constituent based on deconvolution modeling of wine spectra is under development for 400 and 500 MHz spectrometers. This tool will enable consistent measurement of targeted wine compound concentrations agnostic of institution.

Further, we show that integration of ^{1}H NMR and HRMS fingerprinting data through mid-level data fusion significantly enhances the classification of wine samples, effectively separating both grape variety and geographical origin, providing a more robust and comprehensive analysis. These findings highlight the potential of data fusion strategies in improving the accuracy of wine traceability, suggesting a promising direction for future studies in the field of enology.

5. Acknowledgments

This research was made possible by the funding from ANR: WAP-NMR (ANR-21-CE21-0014-01) and MetaboHUB (ANR-11-INBS-0010). We would like to thank the Fondation of Bordeaux for its financial support (donors: Baron Philippe de Rothschild SA, Château Cheval Blanc, Château Lafite Rothschild, Le Domaine Clarence Dillon, Château Petrus). The South Australian node of Metabolomics Australia is funded by Bioplatforms Australia (BPA) under the National Collaborative Research Infrastructure Strategy (NCRIS), with co-investment from the South Australian Government and the AWRI. Finally, we thank the EU-funded vWISE project (Grant agreement ID: 872394).

6. References

1. 1. B. Médina, M.H. Salagoïty, F. Guyon, J. Gaye, P. Hubert, F. Guillaume, *New Analytical Approaches for Verifying the Origin of Food*, 149 (Woodhead Publishing, 2013)
2. 2. I. Le Mao, G. Da Costa, T. Richard, Oeno One **57**, 15 (2023)
3. 3. R. Godelmann, C. Kost, C.-D. Patz, R. Ristow, H. Wachter, J. AOAC Int. **99**, 1295 (2016)
4. 4. *Compendium of International Methods of Wine and Must Analysis*, OIV-MA-AS316-01, (2024)
5. 5. P.A. Solovyev, C. Fauhl-Hassek, J. Riedl, S. Esslinger, L. Bontempo, F. Camin, Compr. Rev. Food Sci. Food Saf. **20**, 2040 (2021)
6. 6. M. Ehlers, B. Horn, J. Raeke, C. Fauhl-Hassek, A. Hermann, J. Brockmeyer, J. Riedl, Food Control **132**, 108508 (2022)
7. 7. S. Esslinger, J. Riedl, C. Fauhl-Hassek, Food Res. Int. **60**, 189 (2014)
8. 8. Y.B. Monakhova, M. Kohl-Himmelseher, T. Kuballa, D.W. Lachenmeier, J. Pharm. Biomed. Anal. **100**, 381 (2014)
9. 9. D. Jacob, C. Deborde, M. Lefebvre, M. Maucourt, A. Moing, Metabolomics **13**, (2016)
10. 10. M. Wheelock Å, C.E. Wheelock, Molecular bioSystems **9**, 2589 (2013)
11. 11. B. Worley, R. Powers, Current Metabolomics **4**, 97 (2016)
12. 12. R. Godelmann, F. Fang, E. Humpfer, B. Schütz, M. Bansbach, H. Schäfer, M. Spraul, J. Agric. Food. Chem. **61**, 5610 (2013)
13. 13. I. Le Mao, G. Da Costa, C. Bautista, G. de Revel, T. Richard, Food Control **145**, 109423 (2023)
14. 14. J.C. Teipel, T. Hausler, K. Sommerfeld, A. Scharinger, S.G. Walch, D.W. Lachenmeier, T. Kuballa, Foods **9**, (2020)
15. 15. Y.B. Monakhova, B. Schütz, H. Schäfer, M. Spraul, T. Kuballa, H. Hahn, D.W. Lachenmeier, Accreditation and Quality Assurance **19**, 17 (2014)
16. 16. E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Aceña, O. Busto, Anal. Chim. Acta **891**, 1 (2015)