Fractionation platform for target identification using off-line directed two-dimensional chromatography, mass spectrometry and nuclear magnetic resonance

Tom van der Laan [a], Hyung Elfrink [a], Fatemeh Azadi-Chegeni [b],
Anne-Charlotte Dubbelman [a], Amy C. Harms [a], Doris M. Jacobs [c], Ulrich Braumann [d],
Aldrik H. Velders [b], John van Duynhoven [c, e], Thomas Hankemeier [a, *]

[a] Analytical Biosciences and Metabolomics, Division of Systems Biomedicine and Pharmacology, Leiden Academic Center for Drug Research, Leiden University, 2333 CC, Leiden, the Netherlands
[b] Agrotechnology and Food Sciences, Laboratory of BioNanoTechnology, Wageningen University, 6700 EK, Wageningen, the Netherlands
[c] Unilever Global Food Innovation Centre, 6708 WH Wageningen, the Netherlands
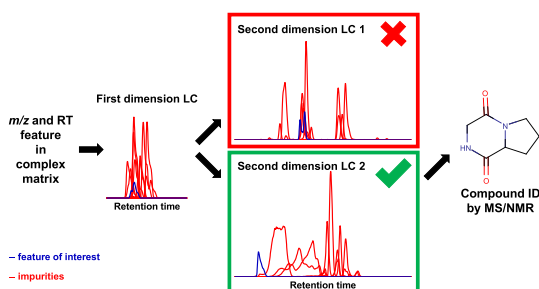[d] Applied NMR Development and Hyphenation, BD-DE, Bruker Biospin, 76287, Rheinstetten, Germany
[e] Laboratory of Biophysics, Wageningen University, 6708 WE, Wageningen, the Netherlands

HIGHLIGHTS

- Innovative directed two-dimensional chromatography which is tailored to the analyte and sample matrix.
- Fractionation methodology that efficiently purifies and upconcentrates unknown compounds from complex matrices.
- Identification platform including complementary techniques mass spectrometry and nuclear magnetic resonance spectroscopy.
- Identification of features down to tens of micromolar concentrations by nuclear magnetic resonance spectroscopy.

GRAPHICAL ABSTRACT

ARTICLE INFO

ABSTRACT

The unambiguous identification of unknown compounds is of utmost importance in the field of metabolomics. However, current identification workflows often suffer from error-sensitive methodologies, which may lead to incorrect structure annotations of small molecules. Therefore, we have developed a comprehensive identification workflow including two highly complementary techniques, i.e. liquid chromatography (LC) combined with mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR), and used it to identify five taste-related retention time and m/z features in soy sauce. An off-line directed two-dimensional separation was performed in order to purify the features prior to the identification. Fractions collected during the first dimension separation (reversed phase low pH) were evaluated for the presence of remaining impurities next to the features of interest. Based on the separation between the feature and impurities, the most orthogonal second dimension chromatography (hydrophilic interaction chromatography or reversed phase high pH) was selected for further purification. Unknown compounds down to tens of micromolar concentrations were tentatively annotated by

* Corresponding author.
  E-mail address: hankemeier@lacdr.leidenuniv.nl (T. Hankemeier).

Metabolomics
Food

MS and structurally confirmed by MS and NMR. The mass (0.4–4.2 μg) and purity of the isolated compounds were sufficient for the acquisition of one and two-dimensional NMR spectra. The use of a directed two-dimensional chromatography allowed for a fractionation that was tailored to each feature and remaining impurities. This makes the fractionation more widely applicable to different sample matrices than one-dimensional or fixed two-dimensional chromatography. Five proline-based 2,5-diketopiperazines were successfully identified in soy sauce. These cyclic dipeptides might contribute to taste by giving a bitter flavour or indirectly enhancing umami flavour.

1. Introduction

Metabolomics has gained a lot of interest in numerous disciplines like life sciences, environmental sciences and the food industry [1–3]. Despite the tremendous developments in metabolite analysis, metabolomics still encounters several analytical challenges. Metabolite identification is considered as one of the major bottlenecks in present-day metabolomics because of its labor-intensive and error-prone methodologies [4]. Structure elucidation of unknown compounds is essential in order to translate analytical data into useful information. In 2007, the Metabolomics Standards Initiative (MSI) defined four levels of identification confidence [5]. Level 1 represents a positive identification of an unknown compound. Level 2 and 3 are the putatively annotation of compounds and compound classes, respectively. Unidentified and unclassified compounds that can be differentiated based on spectral data are classified as level 4. Although these guidelines are dated, they are still used as the standard for metabolite identification. Currently, the Metabolite Identification Task Group of the Metabolomics Society is revising the reporting standards. In this work, we aim for level 1 identification with regards to the MSI guidelines and level C with regards to the newly proposed reporting standards.

Generally, metabolite identification annotates metabolites that have been characterized before, which is referred to as to as non-novel identification. Non-novel identifications are mainly performed by co-characterization with a synthetic standard. In order to reach level 1 identification, a minimum of two physical and/or chemical properties (e.g. fragmentation pattern, chromatographic retention, NMR spectra) of the unknown metabolite and the synthetic standard have to overlap. Two commonly used analytical techniques that are used for the analysis of such properties are mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy. Combined MS and NMR identification strategies have drawn an increased interest because they are highly complementary which increases the identification confidence of unknown metabolites [6,7].

High-resolution mass spectrometers report an accurate mass and intensity ratio of the isotopes that can be sufficient to suggest the elemental composition of a certain m/z value [8]. The list of possible elemental compositions can be further constrained by using MS/MS data in which a fragment ion has to contain no other elements than the precursor ion and a precursor ion has to contain all elements of the fragment ion [9]. MS/MS scans also capture structural information of an m/z value and can be compared with spectral libraries to search for identification hits. Once a hit has been found, the metabolite can be purchased/synthesized and measured using the same MS methodology. The retention time and MS/MS spectra of the synthesized compound can be compared with an unknown metabolite and lead to level 1 identification. In case no synthesized compound is available, confident identifications can be achieved by combining structural information

obtained from MS with NMR [10,11].

NMR spectroscopy can also be used for level 1 identification by comparing typical spin patterns of a synthesized compound with an unknown metabolite. Main challenges for NMR-based identifications are, however, low sensitivity and signal overlap [12]. Many metabolites consist of similar molecular structural motifs which often result in similar peaks in the NMR spectra [6]. Signal overlap is especially problematic for low abundance compounds since they are prone to be overshadowed by signals of high abundance compounds. In addition, the concentration of low abundance compounds are often below the detection limits because of the limited sensitivity of NMR [13]. Signal overlap and insufficient sensitivity can result in missing signals of an unknown compounds and can, therefore, obstruct identification or structure confirmation. Chromatographic separation can be used to drastically improve the power of NMR by isolating target compounds from complex matrices into cleaner fractions which can be up-concentrated for improved sensitivity [14–16]. The purification can even further be improved by including a second orthogonal chromatographic separation [17]. In addition, 2D NMR spectroscopy can be exploited for identification purposes and complemented with 1D NMR quantification methods.

In this study, we demonstrate a platform for the identification of retention time and m/z features in complex samples. In order to unravel the structure of these features, a tentative mass spectrometry-based identification is performed. Thereafter, NMR and MS analyses are used to confirm the structure proposed by MS. Since complex samples consist of a complex matrix and the abundance of certain compounds can be low, we have developed a comprehensive fractionation approach in order to decrease the sample complexity and to increase the concentration prior to the NMR analysis. The fractionation consisted of an off-line directed two-dimensional chromatography, which was tailored to the unknown feature and sample impurities. Fractions collected during the first dimension were evaluated for the presence of impurities. The stationary phase that resulted in the best separation between the unknown feature and the impurities was selected for the second dimension fractionation. We have applied our platform to the identification of five taste-related features in soy sauce. This identification platform provides a general approach for metabolite identification and can be tailored to specific features and samples types. The complementary structure confirmation by MS and NMR ensures a feature identification with high certainty.

2. Materials and methods

2.1. Chemicals and product

Acetonitrile Ultra LC-MS was purchased from Actu-All (Oss, The Netherlands). Ammonium hydroxide (28–30 wt% solution of ammonia in water) and formic acid (98%+) were purchased from Acros Organics (Bleiswijk, The Netherlands). Ammonium formate

(≥99.995%) and cyclo (Pro-Pro) were acquired from Sigma-Aldrich (Zwijndrecht, The Netherlands). Cyclo (Pro-Phe), cyclo (Pro-Gly) and cyclo (Pro-Thr) were purchased from Bachem AG (Bubendorf, Switzerland). Cyclo (Pro-Leu) was obtained from Santa Cruz Biotechnology (Heidelberg, Germany). Deuterated water (D$_2$O, 99.9%) was purchased from Euriso-top (Saint-Aubin, France). All soy sauce products used in this study were commercially available. The soy sauce that was used for the identification and structure confirmation of the unknown compounds was obtained from a local supermarket in Korea.

2.2. Fractionation chromatography and analytical chromatography

The identification platform included in total three off-line fractionation chromatography methods and one analytical chromatography method. All fractionation chromatography methods used an injection volume, flow rate and column temperature of 100 μL, 2 mL/min and 30 °C, respectively.

The reversed phase low pH (*RP low* pH) method used a C18 Atlantis T3, 4.6 × 150 mm, 3 μm particle size column. Mobile phase A and B consisted of 0.1% formic acid (pH 2.7) in water and acetonitrile, respectively. The column oven was set at 30 °C. The gradient started at 0% B and increased linearly to 15% B in 7 min. Subsequently, the gradient increased to 55% B in 3 min and to 100% B in an additional 0.5 min. The gradient was kept at 100% B for 2 min to flush the column. The gradient decreased to 0% B in 0.1 min and was kept at this value for 2.4 min to equilibrate the column. The total gradient time was 15 min.

The reversed phase high pH (*RP high* pH) method used a C18 Kinetex EVO, 4.6 × 150 mm, 5 μm particle size column. The gradient profile was identical to the *RP low* pH gradient, but mobile phase A and B consisted of 2 mM ammonium formate (pH 9) in water and 95/5 (v/v) acetonitrile/water, respectively.

The hydrophilic interaction chromatography (*HILIC*) method used a Sequant ZIC-*HILIC*, 4.6 × 100 mm, 5 μm particle size column. Mobile phase A consisted of 10 mM ammonium formate and 0.075/90/10 (v/v/v) formic acid/acetonitrile/water and mobile phase B of 10 mM ammonium formate and 0.075/10/90 (v/v/v) formic acid/acetonitrile/water. A pH of 3.2 was measured for 10 mM ammonium formate and 0.075% formic acid in water. The gradient started at 0% B for 1.2 min. The gradient increased linearly to 75% B in 7.96 min and was kept at this value for 6.04 min to flush the column. The gradient decreased to 0% B in 0.2 min and stayed at this value for 4.8 min to equilibrate the column. The total gradient time was 20.2 min.

The analytical chromatography *RP low* pH method used a C18 Acquity UPLC HSS T3, 2.1 × 100 mm, 1.8 μm particle size column. The injection volume and flow rate were 1 μL and 0.4 mL/min, respectively. The gradient profile was identical to the fractionation *RP low* pH method.

For the fractionation, liquid chromatography (LC) separations were performed on a Waters Acquity UPLC system (Waters, Etten-Leur, The Netherlands) and fractions were collected in 0.35 min time windows using a Waters Fraction Manager (Waters, Etten-Leur, The Netherlands). The analytical chromatography and three fractionation chromatography methods that were coupled to MS were performed on a Shimadzu Nexera UHPLC (Darmstadt, Germany).

2.3. MS analysis

MS analyses were performed on a Sciex X500R QToF (Darmstadt, Germany). The fractionation chromatography methods were coupled to MS via a 1:20 flow split in order to locate the features in the acquired fractions, assess the orthogonality of the second dimension chromatography and to identify the features. The analytical *RP low* pH method was directly coupled to the MS and was used for structure confirmation and quantification by MS.

The exact mass and isotope pattern were analyzed using Sciex OS v1.5. MS/MS scans were acquired in product ion scan mode and matched to three commonly used spectral libraries: NIST 2017, the MassBank of North America (MoNA) and mzCloud. NIST 2017 was searched in Sciex OS 1.5 by a candidate search. MoNA (https://mona.fiehnlab.ucdavis.edu) and mzCloud (https://www.mzcloud.org) were searched via the web interface. The mass tolerance of the spectral search was set at 0.01 Da. When this mass tolerance did not result in a confident hit, the mass tolerance was increased to 0.3 Da. Features that did not result in a confident library hit were evaluated further by searching their elemental composition in the Dictionary of Natural products (DNP, dnp.chemnetbase.com). The standards of the proposed structures were used for structure confirmation by comparing the retention time and MS/MS spectra of the standard and feature.

The original concentration of the identified compounds in soy sauce were determined by means of standard addition. Known concentrations of the standards were spiked to soy sauce at seven different levels (C1–C7). C7 was the highest calibration concentration and contained all the standards at relevant concentrations. The subsequent calibration levels were all 1:1 dilutions of the previous calibration point. Calibration samples were prepared by mixing 10 μL soy sauce, 10 μL calibration standards and 80 μL H$_2$O. C0 represented the non-spiked soy sauce and was prepared by replacing the calibration standards by H$_2$O. In order to construct a calibration line, the peak area obtained by analysis of the calibration samples was plotted against the concentration of the calibration standards. The original concentration of the identified compounds was determined by dividing the intercept of the calibration line by the slope of the calibration line.

2.4. NMR analysis

1D and 2D NMR spectra were recorded on 600 MHz NMR spectrometers equipped with cryoprobes suitable for sample tube diameters of 1.7 (TCI MicroCryoProbe) and 5 mm (TCI CryoProbe). In both cases samples were measured in 1.7 mm (30 μL) sample tubes. The pulse sequences for the 2D heteronuclear single quantum coherence (HSQC), total correlation spectroscopy (TOCSY), double quantum filtered homonuclear correlation spectroscopy (DQF COSY) and heteronuclear multiple-bond correlation spectroscopy (HMBC) experiments were taken from the Bruker library. 1D ^{1}H NMR spectra were deconvoluted with the CHENOMX software.

The concentration of the identified compounds in the final fraction and the total proton concentration in non-fractionated soy sauce and the final fraction were determined using the Pulse Length-based Concentration determination (PULCON) methodology [18] on a Bruker 600 MHz NMR spectrometer equipped with an Avance III HD console and a 5 mm TCI CryoProbe probehead. PULCON is an external method in which the absolute intensities in two different one-dimensional NMR spectra are correlated based on the principle of reciprocity [19–21]. Based on this principle, the 90° pulse length is inversely proportional to the NMR signal strength for a sample in a given r. f. coil. Therefore, if the concentration of a reference (ref) sample is known and the 90° pulse is precisely calibrated, the unknown concentration (U) of each analyte can be obtained using the following formula:

$$C_U = k C_{ref} \frac{A_U T_U \theta_{90}^U n_{ref}}{A_{ref} T_{ref} \theta_{90}^{ref} n_U}$$

A indicates the integral over the resonances, *T,* the sample temperature in Kelvin, θ_{90}, the 90° pulse length, and *n,* the number of transients used for the measurement of the reference and the unknown sample. The correction factor *k* stands for any differences between the experiments such as incomplete relaxation or different receiver gains which results in variation in signal intensities of the reference and the analyte. The equation is valid when the experiments are performed with the same NMR probe, which is properly tuned and matched to the same amplifier. Trimethylsilylpropanoic acid (TSP) at a concentration of 11.6 mM was used as an external reference. For quantification measurements, each 1D ^{1}H NMR experiment was recorded with a 27-s relaxation delay and 128 number of scans at 300 K. For each fraction, 90° pulse length was calibrated individually. For quantification purposes, all spectra were manually phase and baseline corrected and integral regions were set manually. The quantification of the identified compounds in the final fraction was based on manual integration of assigned and non-overlapping signals. The proton purity of the final fractions and non-fractionated soy sauce were defined as the ratio between the proton concentration of the identified compounds and the total proton concentration excluding the water peak. The concentrations of the identified compounds in non-fractionated soy sauce were determined by mass spectrometry as described in the previous section. The other concentrations were determined by PULCON.

$$\text{Proton purity } (soy\ sauce\ /\ final\ fraction)$$

$$= \frac{\left[H^1\right] \text{identified compound}}{\left[H^1\right] \text{total}}$$

2.5. Identification workflow

The identification workflow is depicted in Fig. 1. The workflow started with *m/z* and retention time features in a biological matrix that correlated with a certain effect. (*1*) The complex matrix was fractionated using the first dimension chromatography (*RP low* pH). Ten injections of 100 μL were fractionated resulting in a total fraction volume of 7 mL. (*2*) 10 μL of the first dimension fractions was analyzed using the first dimension fractionation coupled to mass spectrometry. The fraction containing the highest abundance of the feature was selected and evaluated for remaining impurities. (*3*) 50 μL of the selected first dimension fractions was analyzed using the second dimension chromatography (*HILIC* and *RP high* pH) coupled to mass spectrometry. The orthogonality of the second dimension was evaluated based on the separation between the feature and remaining impurities. The chromatography phase that resulted in the most efficient isolation of the features was selected for further purification. (*4*) The selected first dimension fractions were freeze-dried and reconstituted in 300 μL of mobile phase A (*RP high* pH) or 50/50 (v/v) acetonitrile/water (*HILIC*). Two injections 100 μL were fractionated using the selected second dimension chromatography resulting in a total fraction volume of 1.4 mL (*5*) 10 μL of the second dimension fractions was analyzed using the second dimension fractionation coupled to mass spectrometry. The features were identified by mass spectrometry and feature-containing fractions were selected. (*6*) The proposed structure was confirmed by MS by comparing the standard with the feature in the full complex matrix using the analytical chromatography method coupled to mass spectrometry. (*7*) The selected second dimension fractions were evaporated and reconstituted in 40 μL D$_2$O for NMR analysis reaching a maximum concentration

factor of 17. The proposed structure was confirmed by NMR by comparing the standard with the purified feature after the second chromatography dimension.

2.6. Complex food application

In an inhouse study performed at Unilever, we have found five retention time and *m/z* features that were strong predictors for the sensory scores of fermented soybean flavour (Figure S1 in the supplementary information). These features were measured by reversed phase LC-MS wherein in total 622 features were found (method can be found in Table S1 in the supplemenrtary information). The soy sauce with the strongest fermented soybean flavour according to the partial least squares (PLS) model was selected for the identification workflow. This particluar soy sauce demonstrated the highest abundance of the five taste-related features.

3. Results and discussion

In this study, we aimed for structure identification of retention time and *m/z* features followed by structure confirmation by two complementary techniques: mass spectrometry (MS) and nuclear magnetic resonance (NMR). Because of the high complexity of complex samples and the low apparent concentration of some unknown compounds, a thorough fractionation procedure is often required in order to yield clean and concentrated fractions for NMR analysis. A directed two-dimensional chromatography was used to tailor the fractionation procedure per individual feature. The contaminants that are present after the first dimension chromatography were used to select the most orthogonal chromatography for the second dimension. This resulted in an efficient isolation procedure that is adjustable to different features and sample matrices.

3.1. Sample fractionation

We have applied the identification platform that is depicted in Fig. 1 to isolate five features in soy sauce. The fractionation procedure is demonstrated by an example (feature A) in Fig. 2. After the first dimension chromatography, the mass spectrum of the feature-containing fraction was compared with a blank fraction. The comparison with a blank fraction demonstrated which peaks resulted from the mobile phase and which peaks resulted from the fraction itself. *M/z* peaks present in the blank were attributed to the mobile phase and other *m/z* peaks were selected as undesired contaminants from soy sauce. Thereafter, the fractions were injected into the second dimension chromatography methods, i.e. fractionation reversed phase high pH (*RP high* pH) and fractionation *HILIC*. The *m/z* values of the feature of interest as well as the contaminants were extracted in order to assess the orthogonality of the second dimension. The method that resulted in the cleanest separation of the feature from the contaminants was selected for further fractionation.

An overview of the identified features is provided in Table 1 and the orthogonality assessment is shown in Figure S2 of the supplementary information. As Figure S2 shows, the fractions obtained after the first dimension still contained highly abundant contaminants (represented by the red MS signals). Therefore, the purity of these fractions could still be improved by adding a second dimension chromatography step. For feature A, *HILIC* was selected for further purification as feature A clearly demonstrated the least overlap with the contaminants in this method. *RP high* pH was selected as the second dimension for feature B and E because of similar reasons. *HILIC* and *RP high* pH both demonstrated a powerful cleanup efficiency for feature C, as both methods resulted
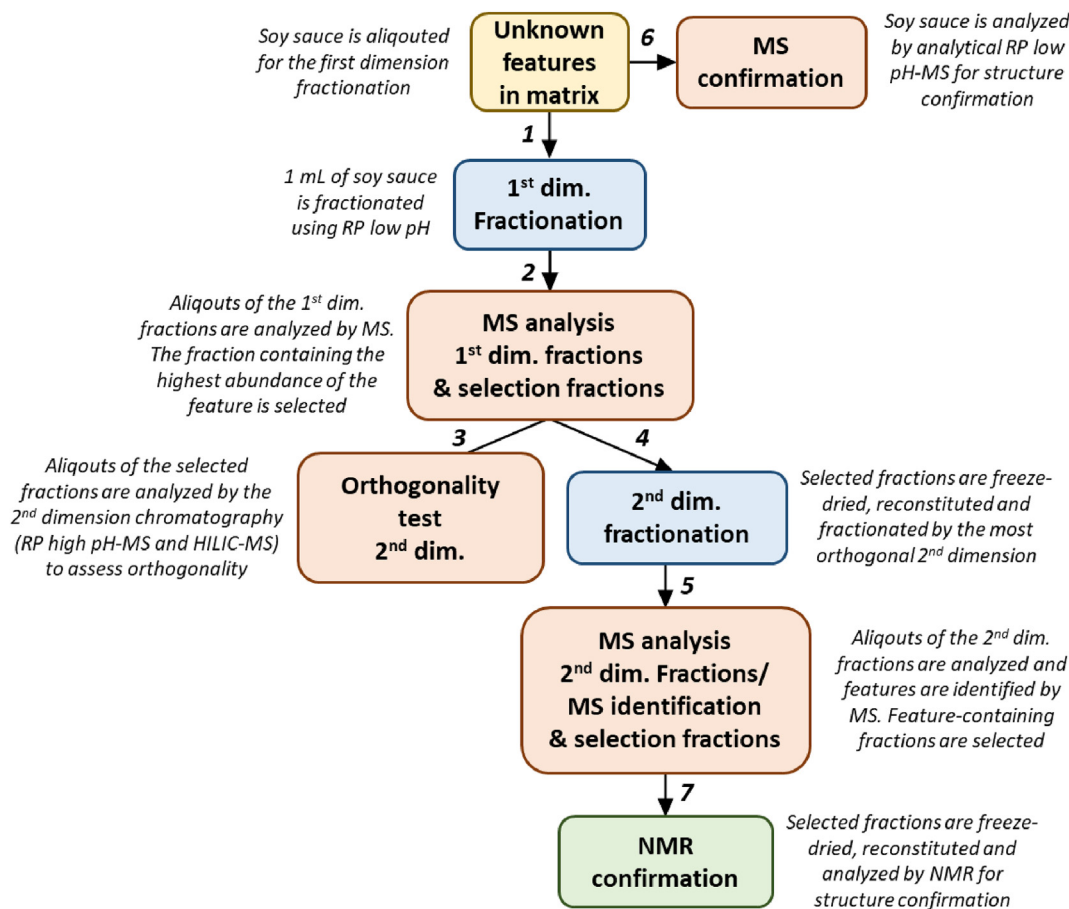
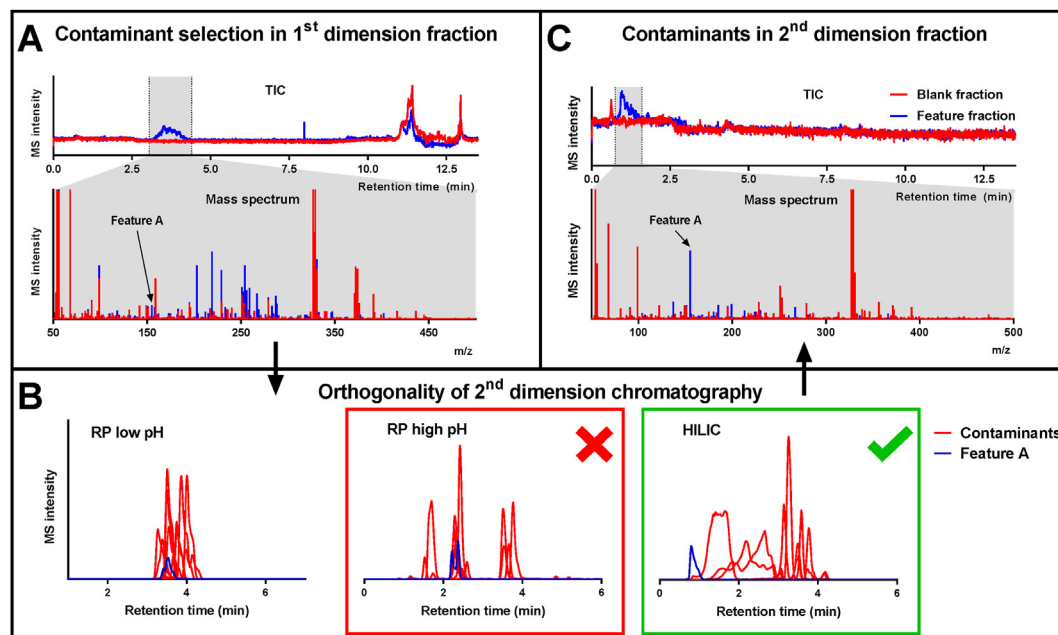**Fig. 1.** The MS/NMR-based identification workflow.

Fig. 2. The fractionation workflow of feature A. (A) The mass spectrum (ranging over the grey area) of the feature-containing and blank fraction measured in the first dimension. *M/z* values that are present in the fraction next to the feature are selected as contaminants. (B) The orthogonality of the second dimension chromatography methods is assessed by extracting the feature and contaminant *m/z* values. (C) Contamination of the feature-containing fraction is assessed after the second dimension chromatography.

Table 1
Overview of the identified features.

Feature ID	A	B	C	D	E
Mass (*m/z*)	155.082	195.113	199.108	211.14	245.13
Soy conc. (mM)	0.02	0.2	0.02	0.08	0.02
2nd dimension	HILIC	RP high pH	RP high pH	RP high pH	RP high pH
Library	DNP	DNP	MoNA	MoNA	MoNA
Cyclo	Pro-Gly	Pro-Pro	Pro-Thr	Pro-Leu	Pro-Phe

in little to no overlap with the contaminants. In this case, the selection of the second dimension was based on peak shape, which was substantially better in the *RP high* pH method. Smaller peak widths are beneficial for fractionation as it allows for a higher yield of the feature per fraction. Although fractions could have been collected over a longer time, we tried to minimize the collection time per fraction as much as possible. The chance of contaminant introduction into the fraction increases with the collection time. Even if the chromatogram looks clean around the feature of interest, undetected contaminants (neutral and/or negatively charged compounds) can still be present and possibly interfere with the NMR analysis. The second dimension fractionation of feature D demonstrated that not only contaminants of other masses but also isomeric contaminants should be taken into account. For this feature, *RP high* pH was able to separate three peaks for the feature mass whereas with *HILIC* they eluted as one. Isomeric contaminants may interfere with NMR structure confirmation to a greater extent than other contaminants because of the high chemical similarity.

The use of one-dimensional chromatography for compound fractionation is well studied and common practice in structural elucidation using NMR [22,23]. However, as demonstrated by our results, one-dimensional chromatography fractionation still leaves highly abundant contaminants in the purified fractions. Second dimension chromatography ideally provides orthogonality to the first dimension resulting in cleaner fractions. The majority of our results demonstrated that the combination of *RP low* pH and *RP high* pH resulted in the most efficient fractionation. This is in accordance to a study of Gilar et al., in which was demonstrated that the highest peak capacity was obtained by two RP chromatography methods that used significantly different pH conditions [24]. However, the combination of *RP low* pH and *RP high* pH resulted in very little orthogonality for feature A, which highly benefitted from the combination of *RP low* pH and *HILIC*. Although orthogonality between two chromatography sorbents types may have been proven, orthogonality is still highly dependent on the chemistry of the targeted compounds [25]. This is emphasized in our work because different first dimension fractions required different second dimension chromatography methods.

3.2. MS identification and structure confirmation

The MS/MS spectra of the soy features were uploaded to three commonly used spectral libraries. The confidence of the library hits was assessed by two criteria. Firstly, the elemental composition of the library precursor ion had to be similar to the elemental composition suggested by the exact mass and isotope pattern of the soy feature. Secondly, the product ion *m/z* values found in the library hit had to be present in the soy feature MS/MS spectra. A library hit was labeled as confident when it met the previously described criteria whereas a library hit was labeled as poor when it met one or none of the criteria. The search in the NIST 2017 library did not result in any library hit. The search in the mzCloud library resulted in poor library hits. The search in the MoNA library resulted in three confident hits, i.e. cyclo (Pro-Thr), cyclo (Pro-Leu) and cyclo (Pro-Phe) for feature C, D and E, respectively.

The experimental MS/MS spectra of two features (A and B) resulted in either no library hit or poor library hits. Therefore, their elemental composition was searched in the dictionary of natural products, which resulted in a list of potential candidates. Thereafter, all experimental MS/MS spectra were compared and evaluated for the presence of known substructures, which is common practice in metabolite identification [26]. All experimental MS/MS spectra contained the fragment 70.07 *m/z*, which is the immonium ion for proline [27]. The three features that could be matched using the MoNA spectral library indeed contained a proline amino acid in addition to another amino acid in a 2,5-diketopiperazine structure. Since the MS/MS spectra of the other two features also contained the characteristic fragment 70.07 *m/z*, it was hypothesized that these features also contained a proline amino acid. The molecular formula search in the dictionary of natural products revealed proline-containing structures for the two remaining features.

The five suggested dipeptides were purchased, measured and compared with the soy features. Figure S3 in the supplementary information demonstrates that the MS/MS spectra of the standards and soy features overlap. The highest abundant product ions of the standards were all found in the MS/MS spectra of the soy features. Some additional product ions that were found in the soy MS/MS spectra were most likely caused by impurities. Since soy sauce is a complex matrix, it can be expected that some coeluting precursor ions with a similar nominal mass pass through the Q1 filter in addition to the studied feature. The structures of the identified features are depicted in Fig. 3.

3.3. NMR structure confirmation and quantification of final fractions

For all the purified soy features and dipeptide standards, two-dimensional homo- (DQCOSY, TOCSY) and heteronuclear (HSQC, HMBC) NMR spectra were recorded. An overview of the ^{1}H and ^{13}C NMR signal assignments is given in Table 2. The soy features were of sufficient purity and concentration to recognize the typical ^{1}H TOCSY NMR spin patterns of the amino acids present in the suggested dipeptides [28]. For several dipeptides the ^{1}H NMR spectral assignments could be complemented with assignments of ^{13}C NMR resonances. These assignments were all in the expected regions. The ^{1}H NMR signals of the standards were identical to the ones identified in the fractions (see Table S2 in the supplementary information).

The overlaid 2D ^{1}H$-^1$H COSY spectra of feature B (blue) and cyclo (Pro-Pro) (red) are presented in Fig. 4. The data illustrates that the concentration of the features in the complex mixture is sufficient for the recording of 2D spectra with a small volume Micro-CryoProbe optimized for mass sensitivity. It also shows that the resonances of the soy feature and the dipeptide standard are identical, thus confirming the spectral assignments for cyclo (Pro-Pro) in Table 2. Moreover, the majority of the correlations in the spectra of the fraction were arising from cyclo (Pro-Pro) indicating the high purity of the fraction. For the other features, spectral assignments were also performed by the 2D ^{1}H$-^1$H COSY spectra of the corresponding dipeptide (see supplementary information Figure S4).

Fig. 5 presents the 1D spectra of feature B. The non-overlapping resonances of cyclo (Pro-Pro) between 2.25 and 2.37 ppm, 3.41–3.60 ppm and 4.40–4.48 ppm were used for quantification. The number of underlying protons was accounted for and thus a concentration of 0.6 mM was determined, as it is stated in Fig. 5. In a similar manner, the other purified soy features were quantified. The results are summarized in Table 2.

Compared with soy sauce (Table 1), the concentrations of the dipeptides in the final fractions (Table 2) increased by a factor of six,
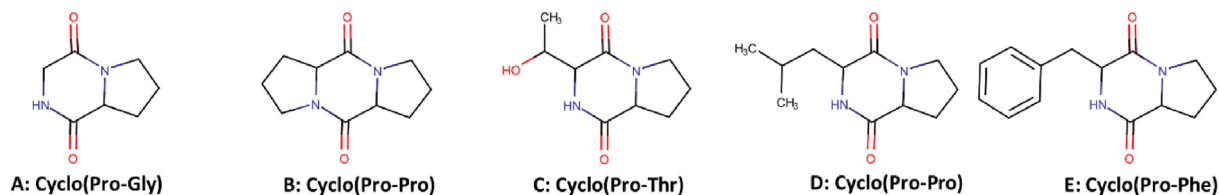
A: Cyclo(Pro-Gly)   **B: Cyclo(Pro-Pro)**   **C: Cyclo(Pro-Thr)**   **D: Cyclo(Pro-Pro)**   **E: Cyclo(Pro-Phe)**

Fig. 3. Structural formulas of identified features. No chirality is indicated and it is assumed that only L-amino acids are present.

Table 2
The concentration and the mass of the dipeptides in the final fractions, the ratio between the final fractions and non-fractionated soy sauce with regards to concentration (dipeptide concentration final fraction/soy sauce) and proton purity (dipeptide proton purity final fraction/soy sauce) and ^{1}H/^{13}C NMR assignments of soy features by spectral comparison with the dipeptide standards.

Cyclo		Pro-Gly		Pro-Pro		Pro-Thr		Pro-Leu		Pro-Phe	
Final fraction	Conc. (mM)	0.3		0.6[a]		0.05		0.3		0.1	
	Mass (μg)	1.7		4.7[a]		0.4		2.5		1.2	
Final fraction	Conc.	15		3		3		4		5	
Soy sauce	Proton purity (10^2)	13		5		8		1		16	
Nuclei analysis		^{13}C	^{1}H	^{13}C	^{1}H	^{13}C	^{1}H	^{13}C	^{1}H	^{13}C	^{1}H
Other Amino acid	Cα	56.6	3.89			63	4.15	56.6	4.29	59.2	4.59
			4.18								
	C1					68.8	4.45	40.6	1.65	40.2	3.13
									1.83		3.28
	C2										
	C3							26.8	1.81		
	C4										
	Me$_1$					21.2	1.31	23.9	0.92		
	Me$_2$							24.9	0.94		
	Φ1									132.9	7.22
	Φ2									131.5	7.39
	Φ3									130.2	7.32
Proline	(C=O) C1	–	4.33	63.5	4.43	61.6	4.30	–	4.34	–	4.06
	C2	–	1.96	30	2.00	30.9	1.91	–	1.97	–	0.79
			2.34		2.31		2.35		2.33		2.01
	C3	–	1.96	25.5	2.00	24.5	1.99	24.4	1.97	–	1.76
			2.07		2.09		2.08	–	2.09		
	NH–C4	–	3.55	48.0	3.56	48.1	3.66		3.56	–	3.34
					3.47		3.51				3.5

[a] The signals of cyclo (Pro-Pro) appeared twice as high with respect to the other dipeptides because of the symmetry. This was taken into account during the quantification.
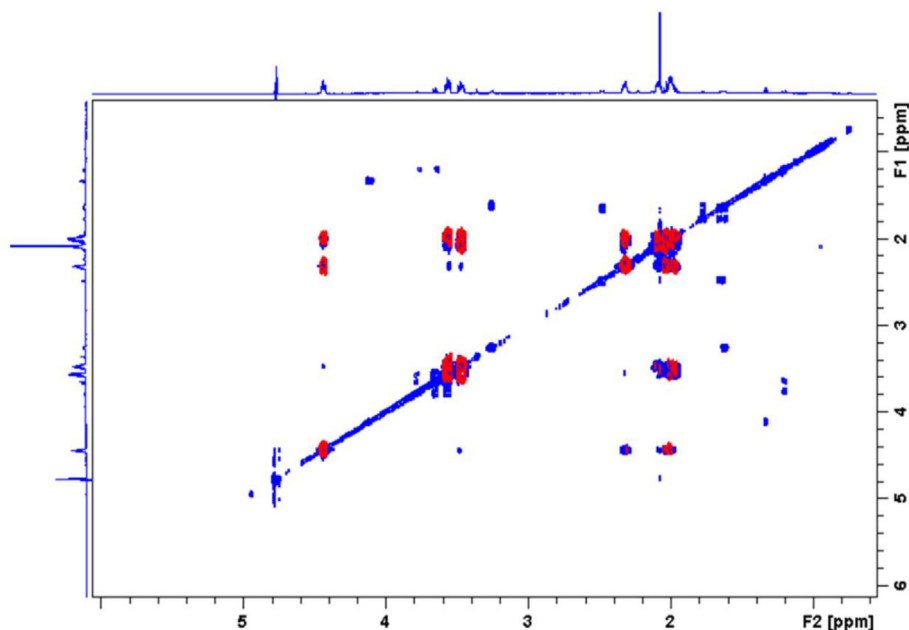

Fig. 4. 2D ^{1}H–^{1}H COSY spectra of the final fraction containing cyclo (Pro-Pro) in blue overlaid with the cyclo (Pro-Pro) standard in red.
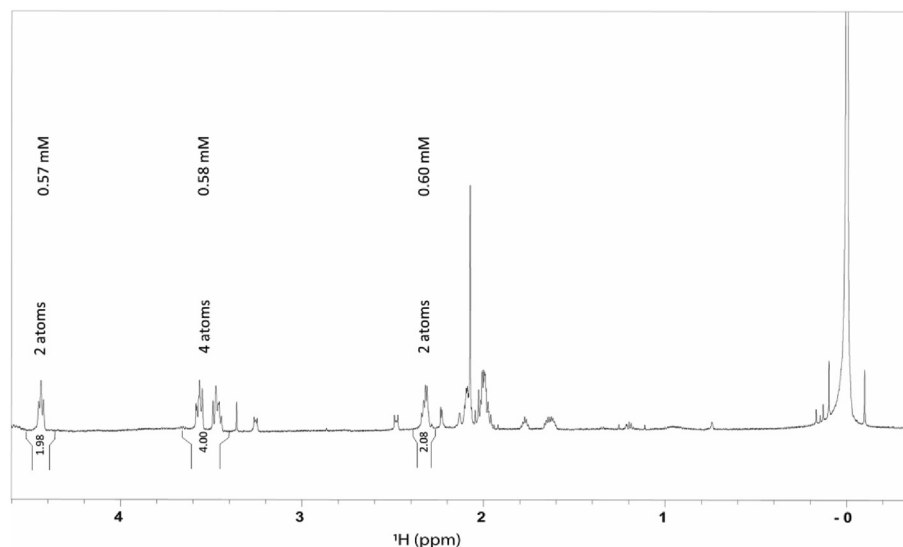
Fig. 5. 1D ^{1}H NMR spectrum of the final fraction containing cyclo (Pro-Pro). Integrated resonances, number of associated protons and calculated concentrations are indicated. The measured mass is 4.7 μg (based on a sample volume of 40 μL and a molar mass of 194.23 g/mol).

on average. Moreover, this increase in concentration was accompanied by a substantial removal of the background signals. In comparison with soy sauce, the proton purity improved by a factor of three orders of magnitude, on average, in the final fractions. The cleanup efficiency of our platform is also emphasized by the decreased complexity of the NMR spectra of the final fractions in comparison with the NMR spectrum of non-fractionated soy sauce (see supplementary information Figure S5). This enabled the recording of high quality 2D NMR spectra in a MicroCryoProbe (1.7 mm sample tubes), optimized for mass sensitivity in small (30 μL) volumes. When the 1.7 mm sample tubes were measured in a conventional 5 mm Cryoprobe, the sensitivity was sufficient for recording high quality 1D NMR spectra which allowed for accurate quantification.

Despite the increase in concentration, the dipeptides were not fully recovered. From the maximum concentration factor of 17, 15–88% was actually achieved. This can be explained by the fraction collection procedure in which the collection vial was switched every 0.35 min. When a dipeptide was eluting during the switching time, this dipeptide was collected into multiple fractions. Eventually, the fraction containing the highest abundance of the dipeptide was selected for further purification aiming for the highest analyte/impurity ratio. This resulted in the loss of the dipeptide that ended up in a non-selected fraction. The purification of Cyclo (Pro-Gly) did not suffer substantial losses as the concentration factor of 15 was close to the maximum factor. This dipeptide was mostly fractionated into one fraction instead of multiple. Although the dipeptides were not full recovered, the fractionation still allowed for a higher concentration in the final fraction as compared to the original concentration in the investigated soy sauce. This emphasizes the power of off-line fractionation prior to NMR analysis, because it enables the up-concentration of purified fractions after the chromatographic separation. In on-line LC-NMR applications, compounds would be diluted because of the mobile phase flow and diffusion caused by the considerable amount of post-column tubing [29]. This problem can be overcome by the use of an SPE trapping prior to the NMR analysis. However, an LC-SPE-NMR platform often results in a tedious analytical setup and method optimization and lacks efficiency for highly polar compounds [30]. Besides, pure and concentrated fractions, which are quantified by NMR, can be used to construct calibration curves using LC-MS and quantify unknown

features in the original sample when authentic standards are unavailable [31,32]. This makes our identification platform suitable for de novo identification with subsequent quantification.

3.4. 2,5-Diketopiperazines and taste

Soy sauce is a popular seasoning that is used particularly in Chinese and Southeast Asian cuisines. It mainly delivers a salty and umami taste, but also adds a characteristic flavour. Amino acids, sugars, organic acids and minerals are considered as main taste components of soy sauce [33,34]. Since soy sauce is a fermented product, studies have focused on protein degradation and thus several taste-active amino acids and peptides have already been identified [35]. However, there are still many unknown molecules that provide the typcial flavour of soy sauce.

We have identified five taste-related cyclic dipeptides, 2,5-diketopiperazines (2,5-DKPs), that derive from proline condensation reactions. 2,5-DKPs can be produced through proteolysis of microorganisms during fermentation or cyclization of linear peptides due to thermal processing [36]. Especially proline-based 2,5-DKPs are produced in both heated and fermented foods. Since soy sauce production involves both heating and fermentation, it can be expected that these particular cyclic dipeptides are present in soy sauce. 2,5-DKPs are often found in food and beverages and have shown to be key inducers of bitter taste [37]. In addition, hydrophobic peptides that contain proline have been associated with bitter sensations, because proline-peptides favour binding to the bitter taste receptor [38]. Although 2,5-DKPs have mostly been linked to bitter taste, these compounds have also been linked to umami taste to a lesser extent [37]. Moreover, Zhu et al. showed that bitter-tasting linear peptides were able to significantly enhance umami taste in the presence of soy sauce and monosodium glutamate [39]. Since cyclic dipeptides generally have a stronger bitter taste in comparison with linear peptides [37], it can be hypothesized that the synergistic effect on umami taste is also stronger. This might explain the correlation between the concentration of proline-based 2,5-DKPs and the taste experience of soy sauce. However, more research is needed to prove this effect. In summary, proline-based 2,5-DKPs may affect bitterness and indirectly enhance umami perception in soy sauce.

4. Conclusions

The unambiguous identification of unknown compounds remains one of the most challenging aspects of metabolomics. The combination of two highly complementary techniques, however, can drastically increase the confidence of structure annotations. In addition, when the spectral matching of MS data does not result in a library hit (e.g. compound has not been identified before or does not fragment), NMR is crucial for de-novo structure elucidation. Therefore, we have developed a workflow for metabolite identification combining NMR and MS. A comprehensive fractionation method was developed to purify five taste-related unknowns from soy sauce to allow for the NMR analysis of m/z and retention time features from a complex sample. The directed two-dimensional fractionation demonstrated that different second dimension chromatography types were needed in order to isolate the unknown compound from other matrix components. The use of a one-dimensional or a fixed two-dimensional fractionation would have resulted in a less efficient purification. The purified fractions were clean and concentrated enough to allow for metabolite identification by MS and structure confirmation by MS and NMR. Although the current study presents the identification of unknown compounds in soy sauce, the developed method is not limited to this particular sample type. The evaluation of impurities and unknown compounds between the first and second dimension ensures that this methodology can account for differences in sample matrices. Each first dimension fraction will benefit from the most orthogonal second dimension chromatography in order to remove remaining impurities. If needed, it is also possible to extend the number of second dimension chromatography types to push the versatility of the fractionation even more. In our study, however, the use of only two different second dimension chromatography types was already sufficient to successfully identify five unknown compounds. All taste-related soy features were identified as proline-based 2,5-diketopiperazines (2,5-DKPs). These compounds are generally known to affect bitterness of food products. Moreover, bitter tasting peptides have been shown to enhance umami perception in soy sauce as well. More research should be conducted to prove that this is also the case for proline-based 2,5-diketopiperazines.

Author contributions statement

Tom van der Laan designed the fractionation workflow, performed the MS experiments and wrote the manuscript. Hyung Elfrink initiated the fractionation method development. Doris Jacobs was responsible for the soy sauce model case. Fatemeh Azadi-Chegeni, Ulrich Braumann and Aldrik H. Velders performed the NMR experiments. Fatemeh Azadi-Chegeni, Doris Jacobs and John van Duynhoven analyzed the NMR data. Anne-Charlotte Dubbelman and Amy Harms contributed to the manuscript. John van Duynhoven and Thomas Hankemeier designed and oversaw the project.

CRediT authorship contribution statement

Tom van der Laan: Methodology, Investigation, Visualization, Writing - original draft. **Hyung Elfrink:** Methodology. **Fatemeh Azadi-Chegeni:** Investigation, Writing - original draft. **Anne-Charlotte Dubbelman:** Writing - review & editing. **Amy C. Harms:** Writing - review & editing, Supervision. **Doris M. Jacobs:** Conceptualization, Funding acquisition, Writing - review & editing. **Ulrich Braumann:** Conceptualization, Funding acquisition. **Aldrik H. Velders:** Supervision, Conceptualization, Funding acquisition, Writing - review & editing. **John van Duynhoven:** Conceptualization, Project administration, Funding acquisition, Writing - review & editing. **Thomas Hankemeier:** Supervision, Conceptualization, Funding acquisition, Writing - review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: John van Duynhoven and Doris Jacobs are employed by a company that manufactures and markets food products. Ulrich Braumann is employed by a company that markets and manufactures analytical equipment. The other authors declare no conflict of interest.

Acknowledgement

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2020.10.054.

Additional information

Supplementary information: this publication is accompanied by supplementary information.

References

[1] R.D. Beger, et al., Metabolomics enables precision medicine: "A white paper, community perspective.", Metabolomics 12 (2016).

[2] C.Y. Lin, M.R. Viant, R.S. Tjeerdema, Metabolomics: methodologies and applications in the environmental sciences, J. Pestic. Sci. 31 (2006) 245—251.

[3] S. Kim, J. Kim, E.J. Yun, K.H. Kim, Food metabolomics: from farm to human, Curr. Opin. Biotechnol. 37 (2016) 16—23.

[4] W.B. Dunn, et al., Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics, Metabolomics 9 (2013) 44—66.

[5] L.W. Sumner, et al., Proposed minimum reporting standards for chemical analysis: chemical analysis working Group (CAWG) metabolomics standards initiative (MSI), Metabolomics 3 (2007) 211—221.

[6] Z. Pan, D. Raftery, Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics, Anal. Bioanal. Chem. 387 (2007) 525—527.

[7] N. Castillo-Munoz, et al., Flavonol 3- O -glycosides series of Vitis vinifera cv, J. Agric. Food Chem. 57 (2009) 209—219.

[8] E. Rathahao-Paris, S. Alves, C. Junot, J.C. Tabet, High resolution mass spectrometry for structural identification of metabolites in metabolomics, Metabolomics 12 (2016) 1—15.

[9] M. Rojas-Chertó, et al., Elemental composition determination based on MS n, Bioinformatics 27 (2011) 2376—2383.

[10] J.J.J. van der Hooft, Rankin, N. Metabolite identification in complex mixtures using nuclear magnetic resonance spectroscopy, in: Modern Magnetic Resonance 1—32, Springer International Publishing, 2016, https://doi.org/10.1007/978-3-319-28275-6_6-1.

[11] R.M. Boiteau, et al., Structure elucidation of unknown metabolites in metabolomics by combined NMR and MS/MS prediction, Metabolites 8 (2018).

[12] J.L. Markley, et al., The future of NMR-based metabolomics, Curr. Opin. Biotechnol. 43 (2017) 34—40.

[13] Z. Lei, D.V. Huhman, L.W. Sumner, Mass spectrometry strategies in metabolomics, J. Biol. Chem. 286 (2011) 25435—25442.

[14] L. Brennan, NMR-based metabolomics: from sample preparation to applications in nutrition research, Prog. Nucl. Magn. Reson. Spectrosc. 83 (2014) 42—49.

[15] A. Bhatia, S.J. Sarma, Z. Lei, L.W. Sumner, UHPLC-QTOF-MS/MS-SPE-NMR: a solution to the metabolomics grand challenge of higher-throughput, confident metabolite identifications, in: Methods in Molecular Biology, vols. 113—133, Humana Press Inc., 2019, 2037.

[16] J.P.M. Van Duynhoven, D.M. Jacobs, Assessment of dietary exposure and effect in humans: the role of NMR, Prog. Nucl. Magn. Reson. Spectrosc. 96 (2016) 58—72.

[17] B.W.J. Pirok, D.R. Stoll, P.J. Schoenmakers, Recent developments in two-dimensional liquid chromatography: fundamental improvements for

practical applications, Anal. Chem. 91 (2019) 240–263.

[18] G. Wider, L. Dreier, Measuring protein concentrations by NMR spectroscopy, J. Am. Chem. Soc. 128 (2006) 2571–2576.

[19] D.I. Hoult, R.E. Richards, The signal-to-noise ratio of the nuclear magnetic resonance experiment, J. Magn. Reson. 24 (1976) 71–85.

[20] D.I. Hoult, The principle of reciprocity in signal strength calculations, Concepts Magn. Reson. 12 (2000) 173–187.

[21] J.J. Van der Klink, The NMR reciprocity theorem for arbitrary probe geometry, J. Magn. Reson. 148 (2001) 147–154.

[22] J. Willmann, K. Mahlstedt, D. Leibfritz, M. Spraul, H. Thiele, Characterization of sphingomyelins in lipid extracts using a HPLC-MS-offline-NMR method, Anal. Chem. 79 (2007) 4188–4191.

[23] S.S. Bird, et al., Structural characterization of plasma metabolites detected via LC-electrochemical coulometric array using LC-UV fractionation, MS, and NMR, Anal. Chem. 84 (2012) 9889–9898.

[24] M. Gilar, P. Olivova, A.E. Daly, J.C. Gebler, Orthogonality of separation in two-dimensional liquid chromatography, Anal. Chem. 77 (2005) 6426–6434.

[25] D.N. Bassanese, et al., Protocols for finding the most orthogonal dimensions for two-dimensional high performance liquid chromatography, Talanta 134 (2015) 402–408.

[26] K.S. Lynn, et al., Metabolite identification for mass spectrometry-based metabolomics using multiple types of correlated ion information, Anal. Chem. 87 (2015) 2143–2151.

[27] L.L. Smith, K.A. Herrmann, V.H. Wysocki, Investigation of gas phase ion structure for proline-containing b 2 ion, J. Am. Soc. Mass Spectrom. 17 (2006) 20–28.

[28] K. Wüthrich, NMR with proteins and, Nucleic Acids. *Europhys. News* 17 (1986) 11–13.

[29] G.S. Walker, T.N. O'Connell, Comparison of LC-NMR and conventional NMR for structure elucidation in drug metabolism studies, Expet Opin. Drug Metabol.

[30] G. Schlotterbeck, S.M. Ceccarelli, LC-SPE-NMR-MS: a total analysis system for bioanalysis, Bioanalysis 1 (2009) 549–559.

[31] J. Van Duynhoven, et al., Rapid and sustained systemic circulation of conjugated gut microbial catabolites after single-dose black tea extract consumption, J. Proteome Res. 13 (2014) 2668–2678.

[32] R. Espina, et al., Nuclear magnetic resonance spectroscopy as a quantitative tool to determine the concentrations of biologically produced metabolites: implications in metabolites in safety testing, Chem. Res. Toxicol. 22 (2009) 299–310.

[33] K. Shiga, et al., Metabolic profiling approach to explore compounds related to the umami intensity of soy sauce, J. Agric. Food Chem. 62 (2014) 7317–7322.

[34] S. Kaneko, K. Kumazawa, O. Nishimura, Isolation and identification of the umami enhancing compounds in Japanese soy sauce, Biosci. Biotechnol. Biochem. 75 (2011) 1275–1282.

[35] A. Apriyantono, D. Setyaningsih, P. Hariyadi, L. Nuraida, Sensory and peptides characteristics of soy sauce fractions obtained by ultrafiltration, Adv. Exp. Med. Biol. 542 (2004) 213–226.

[36] Y. Otsuka, et al., Investigation of the formation mechanism of proline-containing cyclic dipeptide from the linear peptide, Biosci. Biotechnol. Biochem. 83 (2019) 2355–2363.

[37] A.D. Borthwick, N.C. Da Costa, 2,5-diketopiperazines in food and beverages: taste and bioactivity, Crit. Rev. Food Sci. Nutr. 57 (2017) 718–742.

[38] C.J. Zhao, A. Schieber, M.G. Gänzle, Formation of taste-active amino acids, amino acid derivatives and peptides in food fermentations – a review, Food Res. Int. 89 (2016) 39–47.

[39] X. Zhu, Y. Chen, R. Subramanian, Comparison of information-dependent acquisition, SWATH, and MS All techniques in metabolite identification study employing ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry, Anal. Chem. 86 (2014) 1202–1209.

Toxicol. 4 (2008) 1295–1305.