

# Rapid visualization of chemically related compounds using Kendrick mass defect as a filter in mass spectrometry imaging

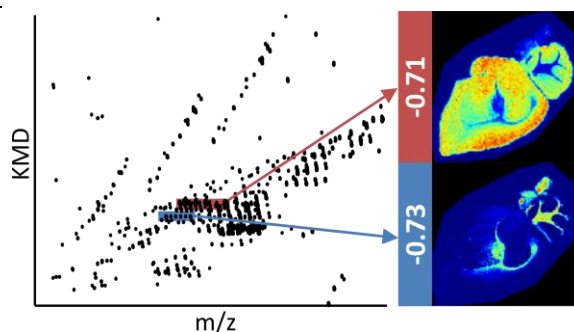
Christopher Kune<sup>a,†</sup> and Andréa McCann<sup>a,†</sup>, Raphaël La Rocca<sup>a</sup>, Anthony Arguelles Arias<sup>b</sup>, Mathieu Tiquet<sup>a</sup>, Daan Van Kruining<sup>c</sup>, Pilar Martinez Martinez<sup>c</sup>, Marc Ongena<sup>b</sup>, Gauthier Eppe<sup>a</sup>, Loïc Quinton<sup>a</sup>, Johann Far<sup>a</sup>, Edwin De Pauw<sup>a\*</sup>

<sup>a</sup>Mass Spectrometry Laboratory, MolSys Research Unit, University of Liège, Liège, Belgium.

<sup>b</sup>Microbial Processes and Interactions, Gembloux Agro-Bio Tech, TERRA Teaching and Research Centre, University of Liege, Gembloux, Belgium.

<sup>c</sup>Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University, Maastricht, the Netherlands.

**ABSTRACT:** Kendrick mass defect (KMD) analysis is widely used for helping the detection and identification of chemically related compounds based on exact mass measurements. We report here the use of KMD as a criterion for filtering complex mass spectrometry dataset. The method enables an automated, easy and efficient data processing, enabling the reconstruction of 2D distributions of family of homologous compounds from MSI images. We show that the KMD filtering, based on an in-house software, is suitable and robust for high resolution (full width at half-maximum, FWHM, at  $m/z$  410 of 20 000) and very high-resolution (FWHM, at  $m/z$  410 of 160 000) MSI data. This method has been successfully applied to two different types of samples, bacteria co-cultures and brain tissue section.



Identification of analytes in complex mixtures is still a major challenge in analytical chemistry<sup>1</sup>. For that purpose, mass spectrometers are among the most widely used instruments. They are commonly coupled to direct infusion or hyphenated with separation techniques, e.g., liquid chromatography (LC), capillary electrophoresis (CE), ion mobility spectrometry (IMS). Mass spectrometry imaging (MSI) is a particular case in which 2D distribution of analytes can be reconstructed after recording full mass spectra at the coordinate of the “pixels” sampled by a LASER beam. The resolving power and the mass accuracy of analyzers greatly improved to reach values allowing the direct determination of molecular formula of low molecular weight analytes. However, due to the massive amount of information generated now in (ultra) high resolution ((U)HR) MS data<sup>2</sup>, the extraction of chemical information from spectra is becoming more and more challenging.

Kendrick<sup>3</sup> introduced in 1963 a method based on the two-dimensional projection of the atomic composition space enabling a rapid identification of families of compounds. The Kendrick method is based on the transformation of the mass spectra (mass-to charge ratio or  $m/z$  in the X-axis and intensity in the Y-axis) to Kendrick plots where the X-axis corresponds to the Kendrick mass (KM) and the Y-axis corresponds to the Kendrick mass defect (KMD). The KM is defined by converting the exact masses of a given group of atoms (the Kendrick base) to the nearest integer value in atomic mass unit (amu). Kendrick method changes the IUPAC amu reference (i.e. 1/12 of the  $^{12}\text{C}$

mass) to another mass unit (e.g. 1/14 of the  $^{12}\text{C}^1\text{H}_2$  radical mass). KM can be calculated from  $m/z$  value by the Equation 1 where the Nominal- $M_{\text{Kendrick Reference}}$  and  $M_{\text{Kendrick Reference}}$  are respectively the nominal mass and the exact mass of the Kendrick reference used for the atomic mass unit definition.

$$KM = m/z \times \frac{\text{Nominal}-M_{\text{Kendrick Reference}}}{M_{\text{Kendrick Reference}}} \quad (1)$$

The KMD corresponds to the difference between the rounded KM values and the KM (Eq. 2).

$$KMD = \text{round}(KM) - KM \quad (2)$$

All the compounds differing only by the number of repeating unit of the Kendrick reference will have the same KMD for different KM. Consequently, chemically related compounds will be aligned in the KMD axis on the Kendrick plot, making their identification easier. Recently higher order Kendrick transformations have also been applied when oblique correlations suggest the presence of further repeating units<sup>4</sup>. Over the years, KMD analysis has been extended to the analysis of complex samples composed of chemically related structures, such as petrochemicals, but also polymers or lipids<sup>5-11</sup>.

In this paper, we show that the KMD, which is specific to chemically related homologous series of compounds, can be used to replace the  $m/z$  axis of the mean spectra classically used for the analysis of images in MSI. The KMD can be adjusted to filter data in a non-targeted, semi-targeted and targeted mode. This approach enables an automated, faster and efficient data processing, allowing the rapid identification and localization of families of compounds in the complex chemical space of the image. This method has been successfully applied to two different types of samples, bacteria co-cultures and brain tissue sections.

## MATERIALS AND METHODS

**Materials.** The MALDI matrices 2,3-diaminonaphthalene (DAN) and  $\alpha$ -cyano-4-hydroxycinnamic acid (HCCA) were purchased from Sigma-Aldrich (Belgium). Acetonitrile and Methanol were LC-MS grade from Biosolve (Belgium). Poly(ethylene oxide) monomethylether (CH<sub>3</sub>O-PEO-H) having an average nominal mass of 750 g/mol, tetraalkylammonium bromides (reagent grade >98%), trifluoroacetic acid (TFA), formic acid, bovine serum albumin (BSA), methyl-tert-butyl ether (MtBE), acetone HPLC grade and sodium chloride were purchased from Sigma-Aldrich (Belgium). Pr. Philippe Jacques (Terra Teaching and Research Center, Microbial Processes and Interactions, University of Liège, Belgium) and Lipofabrik (Lille, France) kindly provided standards of lipopeptides.

**Sample preparation for MALDI-FT-ICR MS imaging proof of concept.** DAN matrix solution was prepared at saturation (>30mg/mL) in ACN/water 70:30 vol/vol spiked with 0.2% TFA. Tetraalkylammonium bromide stock solution was dissolved in ACN to reach the final concentration of 10 $\mu$ M. Trypsin digestion of BSA using MS grade porcine trypsin (purchased from PIERCE, Thermofisher, Belgium) was prepared at a final concentration of 15  $\mu$ M in 50% methanol spiked with 0.1% formic acid following the procedure described in supporting information. Lipids were extracted from cultured eukaryotic cells based on Matysah protocol<sup>12</sup> and resuspended in MeOH. Lipopeptides were dissolved in MeOH spiked with 0.1% TFA at final concentration of 10 $\mu$ M. CH<sub>3</sub>O-PEO-H was dissolved in ACN 10 $\mu$ M NaCl to reach a concentration of 10 $\mu$ M. Each sample was mixed separately with DAN matrix at a ratio of 1:1, and were spotted randomly on a MALDI AnchorChip target plate (Bruker, Bremen, Germany) in duplicate for high and low resolution MS imaging. In addition, 1 $\mu$ L of each sample was mixed and DAN matrix was added at a 1:1 ratio and spotted onto the MALDI target plate. Finally, a blank spot of DAN matrix was added.

**Bacterial Strains, medium and Culture conditions.** The strains used for this study were *Bacillus velezensis* GAI and *Pseudomonas* sp. CMR12a. Both strains were inoculated on a semi-solid agar-based PDA medium (Potato Dextrose Agar) and incubated overnight at 30°C. At the end of the culture, petri dishes were sealed with Parafilm M® and stored at 4°C prior to analysis.

**Animals, tissue sampling and sectioning.** Mouse brain samples were provided by Prof. Martinez (School of Mental Health and Neuroscience at Maastricht University the Netherlands), and bred in house as described elsewhere<sup>13</sup>. Animals were sacrificed by CO<sub>2</sub> inhalation and brain tissues were extracted. After extraction, the brains were cut across the sagittal midline and immediately fresh-frozen in liquid nitrogen. The brain parts were subsequently stored at -80 °C. On the day of transport, samples were placed on dry ice and transferred to the University of Liege. All procedures were

approved by the Animal Welfare Committee of Maastricht University and were performed according to Dutch federal regulations for animal protection. Frozen brain tissues were sectioned into 14- $\mu$ m thick sagittal slices using a CryoStar NX70 (Thermo Fisher Scientific, Massachusetts, USA) at -20°C and thaw-mounted onto indium-tin oxide (ITO) conductive glass slides (Bruker Daltonics, Bremen, Germany).

**Sample preparation for MALDI-FT-ICR MS imaging.** Microbial colonies on agar and region of interest were cut directly from the petri dish and transferred to the target ITO plate (Bruker, Bremen, Germany), previously covered with double sided conductive carbon tape (StructureProbe INC, West Chester, PA, USA). This assembly was then placed in a vacuum desiccator until complete drying (overnight). An HCCA matrix solution was prepared at the concentration of 5mg/mL in ACN/water 70:30 vol/vol spiked with 0.2% TFA, based on the previous work of Debois et al<sup>14</sup>. Application of the matrix solution on the dried bacterial sample was performed using the SunCollect (SunChrom, Friedrichsdorf, Germany) spraying system. In total, 60 layers of HCCA matrix were sprayed on the sample. The three first layers were sprayed at 5 $\mu$ L/min, and the other at 10 $\mu$ L/min. Mouse brain tissue slides deposited on ITO slides were first put into a vacuum desiccator for approximately 20 minutes. A solution of HCCA at the concentration of 5mg/mL in MeOH/water 90:10 vol/vol spiked with 0.2% TFA was prepared, and sprayed onto the dried sample by the SunCollect. The number of spraying layers were set to 30, at a spraying speed of 10 $\mu$ L/min, excepted for the three first layers sprayed at 5 $\mu$ L/min.

**FT-ICR Mass spectrometry.** Mass spectrometry images were obtained using a FT-ICR mass spectrometer (Solarix XR 9.4T, (Bruker Daltonics, Bremen, Germany)). The mass spectrometer was systematically mass calibrated from 200  $m/z$  to 2,300  $m/z$  before each analysis with a red phosphorous solution in pure acetone spotted directly onto the ITO Glass slide or onto the MALDI plate to reach a mass accuracy better than 0.5 ppm. FlexImaging 5.0 (Bruker Daltonics, Bremen, Germany) software was used for MALDI MS imaging acquisition, with a pixel step size for the surface raster set to 150 $\mu$ m for Kendrick proof of concept imaging and to 100 $\mu$ m for bacteria and brain tissue imaging. For each mass spectrum, one scan of 20 LASER shots was performed at a repetition rate of 200Hz. The LASER power was set to 50% and the beam focus was set to “small”. The mass spectra for high and low resolution mass images were recorded by setting the number of datapoints in the transient at respectively 1 000 000 (1M) and 256 000 (256K).

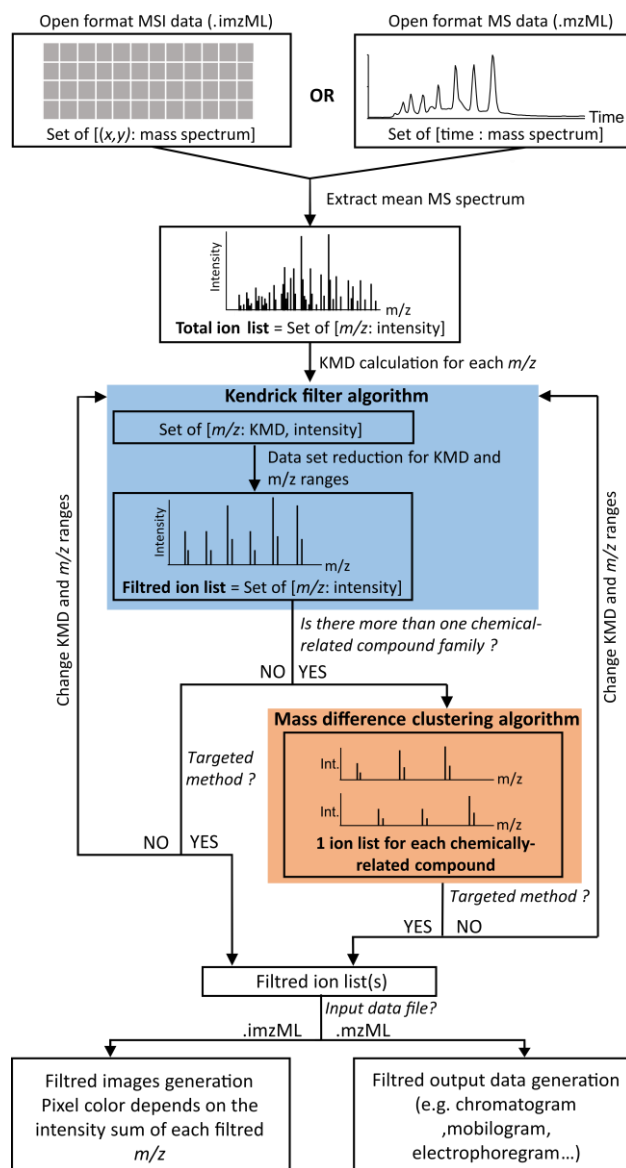
**Kendrick filtering for data reduction and clustering.** MSconvert (Proteowizard)<sup>15</sup> and Fleximaging 5.0 (Bruker) have been used to, respectively, convert MS raw files (obtained from e.g. direct infusion MS, liquid chromatograph MS, ion mobility MS...) to mzML and MSI raw files (obtained from imaging MS) to imzML files. These files are processed under Python in-house software (MSKendrickFilter, available upon request : [c.kune@uliege.be](mailto:c.kune@uliege.be)).

After a conversion to binary format files (.ICK), specifically designed for our software, data is filtrated based on the KMD. For a MS analysis, the .ICK file contains the mass-to-charge and intensity information for each detected ion of each recorded scan (or pixel in case of MSI application). A noise reduction, based on a minimum intensity threshold, can be applied to limit the .ICK file size and the CPU and RAM usage of the software.

The Kendrick filtering method workflow for MSI data or MS data (e.g., direct infusion, chromatogram, mobilogram,

electrophoregram...) is depicted in Figure 1. Once the .ICK file has been generated, an ion list containing all the  $m/z$  present in the mean spectrum (sum of each MS spectra in MS or MSI data) is created. The software calculates a Kendrick mass defect (KMD) for each  $m/z$  value according to Eq. 1 and Eq. 2 in the user-defined Kendrick reference (i.e.  $-\text{CH}_2-$  in this work). Different rounding functions are available in our software, i.e. floor, ceil, and round. Floor option has been used for the results reported here. All  $m/z$  ratio are then associated to a KMD. The second step consists in filtering the total ion list by keeping only the ions whose KMD is included in the user-defined KMD range. An  $m/z$  ratio selection range can also be applied as a criterion for data filtering. The resulting ion list contains chemically related compounds. In the case of different compound families sharing a similar KMD, an additional algorithm is added to cluster the saved ion by compound families. This clustering step (reported as mass difference clustering algorithm in Figure 1) groups ions according to their mass difference corresponding to a multiple of the chosen Kendrick reference mass. The software allows a user-defined  $m/z$  tolerance value (e.g. here we used 5 ppm and 2 ppm for  $m/z$  tolerances with resolution of, 20,000 and 160,000 full width at half-maximum, FWHM, at 410  $m/z$ , respectively). In this version of the software, isotope compounds are considered as different families due to Kendrick mass defect deviations. This algorithm generates an ion list for each putative compound family. Finally, a chromatogram (for liquid chromatography or gas chromatography MS), a mobilogram (for ion mobility MS), an electrophoregram (for capillary electrophoresis MS) or an image (for MSI) can be generated from these ion lists. The entire process can be repeated with different KMD and  $m/z$  ranges to generate additional MS data of different compound families.

In this work, we only focused the Kendrick filtering application on MSI data as a rapid visualization and localization tool of chemically related compounds. An intensity threshold value of 200,000 count per scan and a normalization to the total ion current (TIC) for each pixel was applied for all the reported images.



**Figure 1:** Kendrick filtering method workflow applied on MSI data and MS data (such as chromatography, ion mobility, capillary electrophoresis...).

## RESULTS AND DISCUSSIONS

**KMD filtering concept.** The illustration of the KMD filtering concept is shown in Figure 2 for a MALDI FT-ICR MS data. The MALDI spots were prepared with a mix of different compound families including lipids, lipopeptides, polymers, tetraalkylammonium (TAA) and peptides of BSA tryptic digest. The resulting mass spectra (Figure 2.B) is rather complex and the assignment of MS peak associated to compounds family is challenging. The Kendrick plot (Figure 2.A), generated using  $-\text{CH}_2-$  as Kendrick reference, significantly reduces the complexity by aligning the compounds bearing a repeating unit of  $-\text{CH}_2-$ . Tetraalkylammonium, lipids and lipopeptides families are then easily detected (respectively highlighted in blue, green and orange). Polymers families are revealed by changing the Kendrick reference to the repeating monomer unit  $-\text{CH}_2\text{O}-$  (See Supporting information Figure S2). Thus, the goal of the proposed Kendrick filtering method relies on its capacity to extract from a raw mass spectrum the relevant peaks associated to a given family of compounds and, by generating a new filtered spectrum, chromatogram, mobilogram or image. Depending on the mass spectrometer resolution, setting a KMD

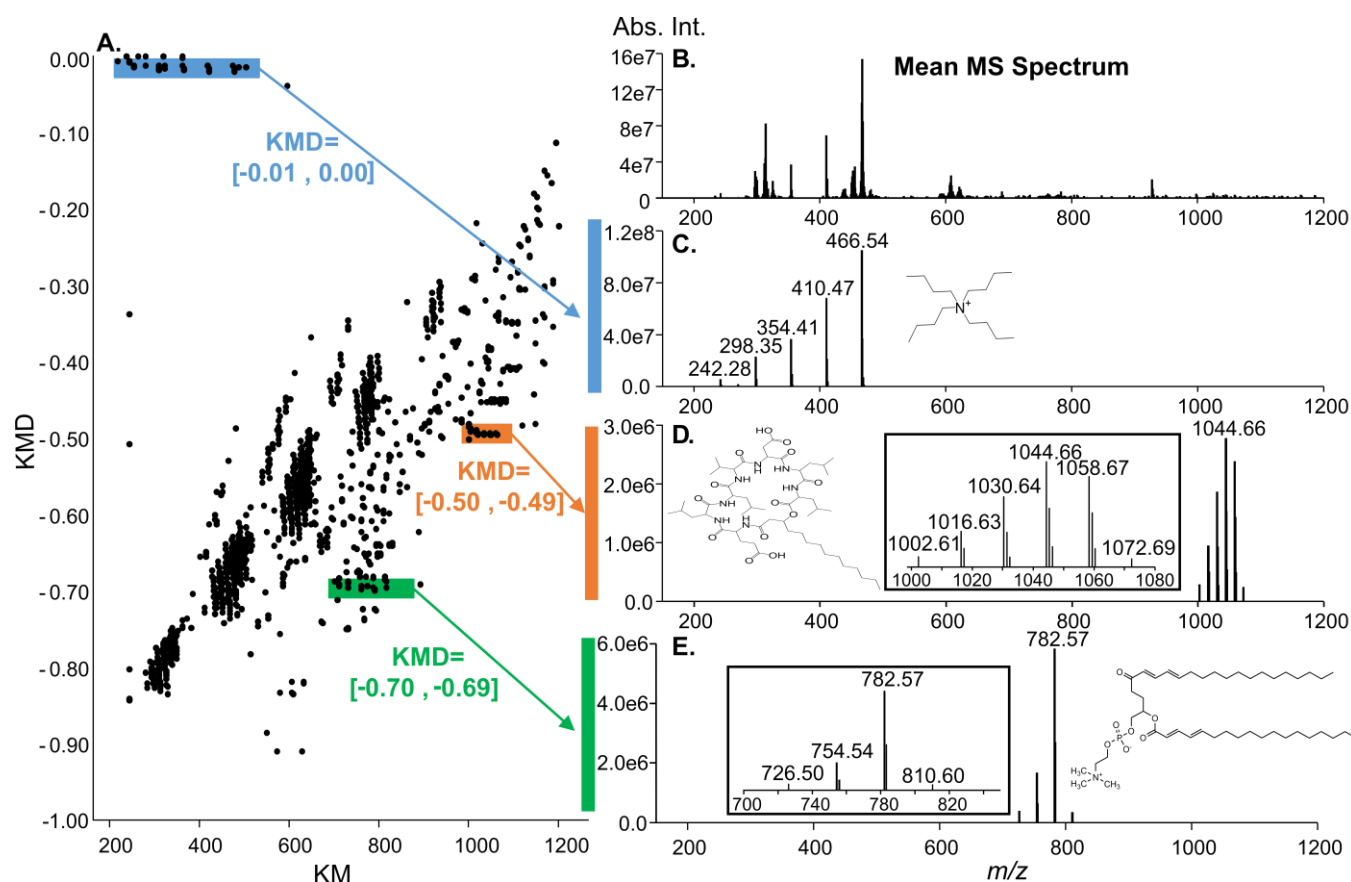
range as well as a MS range might not be sufficient to isolate a single family of chemically related compounds. Indeed, it can be explained by the presence of other compounds exhibiting similar KMD and  $m/z$ . In these cases, the mass difference clustering algorithm (see materials and methods section) can be applied to separate overlapping families of compounds. Thus, a sequential strategy consisting in first filtering by KMD and then by the mass difference clustering gives clean filtered spectra of tetraalkylammonium (Figure 2.C), surfactins (lipopeptides family, Figure 2.D) and phosphatidylcholines (lipid family, Figure 2.E) from a MALDI spot.

**KMD filtering concept.** As a proof-of-concept, but also to assess the effectiveness of KMD filtering process on MSI data, we randomly spotted a MALDI AnchorChip plate with several droplets of lipids, tetraalkylammonium salts, lipopeptides and polymers (Figure 3.A). The proposed KMD filtering algorithm

developed here was applied on the imaging dataset. The MSI data were acquired with two different MS resolution calculated at the  $m/z$  410 (i.e. 20,000 and 160,000, respectively). A mixture spot (spot 8, containing all the investigated compound families) and a blank matrix spot (Spot 3) have been added as a positive and a negative control, respectively.

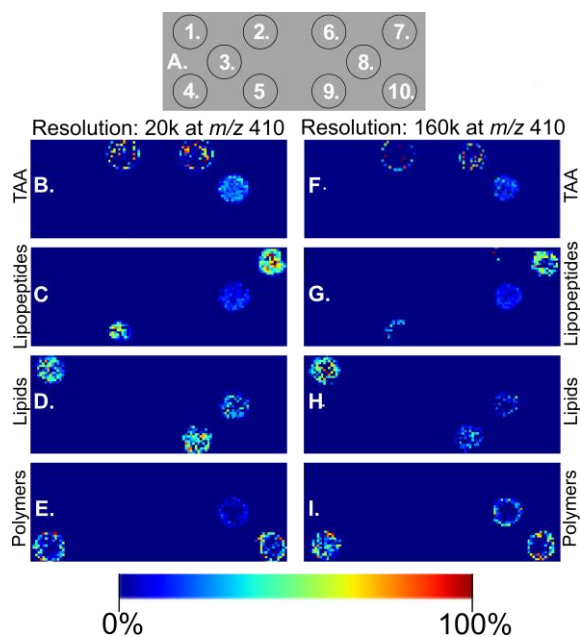
The KMD filtering process allowed to determine the precise localization of each chemically related compounds for both MS resolution acquired datasets. The similarities of the images produced at 20K and 160K resolution is in fact ensured by the mass difference clustering algorithm. In addition, one should note that none of the targeted compounds was found in the negative control spot 3 at both selected MS resolution.

These results outline the power of KMD filtering approach as a rapid visualization tool on MSI data generated from (U)HR MS analyzer with different MS resolutions.



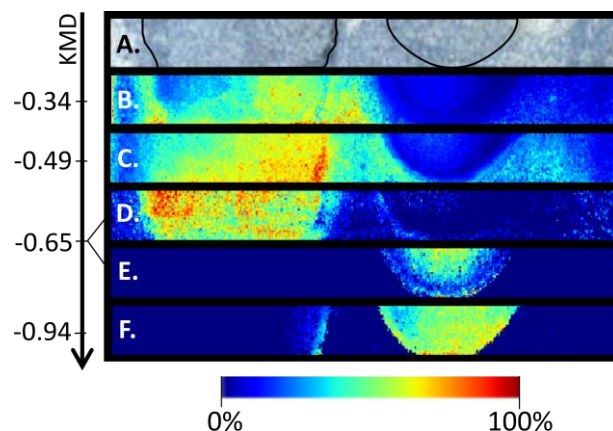
**Figure 2:** Illustration of the KMD filtering method. (A) Kendrick mass defect plot of the spectra obtained by FT-ICR MS. (B) Full MS spectra when no filtration is applied. By selecting a specific KMD range and eventually a specific mass range, it is possible to filter the MS Spectra and show only the chemically related molecules such as TAA (C), Lipopeptides : surfactins (D) or lipids : phosphatidylcholines (E).





**Figure 3:** KMD filtering applied on MALDI mass spectrometry imaging data where different chemically-related compound were randomly spotted on a MALDI AnchorChip target plate. A. Scheme of the repartition of chemically-related compounds (Spots 1. and 9.: Lipids, spots 2. and 6.: Tetraalkylammonium, spot 3.: Matrix, spot 4. and 10.: Polymers, spots 5. and 7.: Lipopeptides, spot 8.: Mix). B to E are generated image after KMD filtering for, respectively, TAA, lipopeptides, lipids and polymers from MSI data with a MS resolution of FWHM 20,000 for  $m/z$  410. F to I are generated image after KMD filtering for, respectively, TAA, lipopeptides, lipids and polymers from MSI data with a MS resolution of FWHM 160,000 for  $m/z$  410. The color scale represents a normalized intensity of the ions.

**MALDI Imaging of bacterial culture.** Mass spectrometry imaging of culture of bacteria has gained growing attention<sup>14, 16-19</sup>. By combining the chemical identification of a compound with its localization, MSI could provide more insights into the bacterial system communication. The objective would be to detect and localize compounds of interest, such as lipopeptides, produced by the bacteria in different environments or at different elapsed times of the growing process. Since most of lipopeptides contain a variable hydrophobic lipid chain-length<sup>20</sup>, applying a KMD plot with a  $-\text{CH}_2-$  repeating unit, seems to be a fit-for-purpose approach to rapidly screen the detected compounds in the region of interest. The results were analyzed according to a non-targeted approach, by screening along the KMD axis different families of compounds (Figure 4). These compounds were then identified based on their exact mass. Three different families of lipopeptides were detected, and identified as iturins (Figure 4.B), surfactins (Figure 4.C) and sessilins (Figure 4.F). As expected, the distribution of lipopeptides differs according to the different families. Surfactins and iturins are produced by *Bacillus* and excreted in the agar medium, suggesting an interplay between *Bacillus* and *Pseudomonas*<sup>21</sup>. On the opposite, the sessilins are accumulating in *Pseudomonas*, but are not detected outside of the bacterial colony.

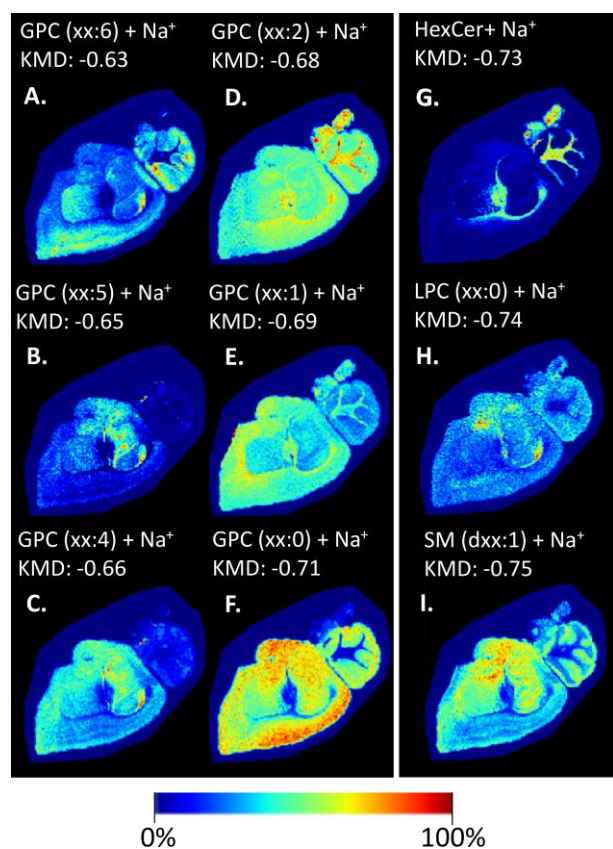


**Figure 4:** KMD filtering applied on MALDI Imaging of bacterial culture. (A) Optical image of the sample covered with matrix, *Pseudomonas* is circled at the left side, and *Bacillus* is circled at the right side. The identified families are presented according to their KMD value. (B) Iturins A (sodium adducts) at the KMD range -0.33 to -0.35 (C) Surfactins (sodium adducts) at the KMD range -0.49 to -0.51. (D) and (E) identification of two different lipid classes at the KMD range -0.65 to -0.66, (F) Sessilins (sodium adducts) at the KMD range -0.93 to -0.95. The color scale represents the normalized intensity of the ions.

In the KMD range -0.65 to -0.66, at least two different lipid families were detected (Figure 4, D and E). Thanks to the mass difference clustering algorithm, it is possible to discriminate these two groups. The first group of lipids (Figure 4.D) contained the  $m/z$  values: 710.444, 724.459 and 738.477, belonging to the sub class 1-(1Z-alkenyl)-2-acylglycerophosphoethanolamines of the glycerophosphoethanolamines lipid class. Based on their localization on the ITO slide, this lipid group was specifically related to *Bacillus velezensis* S499. The second group of lipids (Figure 4.E) composed of the following  $m/z$  values; 734.472 and 748.487 was identified as diacylglycerophosphoethanolamines or diacylglycerophosphocholines. Their spatial distribution was specific to *Pseudomonas* sp. CMR12a.

**MALDI imaging of a mouse brain section.** The last example treated in this paper belongs to MSI applied on mice tissue. MSI enables to localize a set of diverse molecules in a specific tissue, such as brain<sup>22</sup>, liver<sup>23</sup> or kidney<sup>24</sup>, providing local molecular information for a better understating of illnesses and deeper insights into physiological mechanisms. Among the detected compounds, lipids represent a high percentage of ions since they are present in every cell by constituting the membranes and energy storage vesicles<sup>25, 26</sup>. Their implication in various biological functions and their capacity to reflect the physiological and environmental conditions makes their study essential for disease understanding or biomarker discovery<sup>27-30</sup>. The results obtained by MSI of a mouse brain section (Figure 5) shows different patterns of lipid distribution belonging to the families of glycerophosphocholines (GPCs), hexosylceramides (HexCers), lysophosphocholins (LPCs) and sphingomyelins (SMs) classes, as described in the literature on brain tissue analysis<sup>31, 32</sup>. The software shows the list of peaks used to build the distribution of each family. From that list, individual distributions can be obtained. All the detected families of compounds are named according to their class (GPCs, SMs, LPCs) and according to their unsaturation (xx:n), where “xx” represents the number of carbon and “n” corresponds to the number of unsaturation. The attribution of a specific class was done based on high-resolution MS data, using LIPIDMAPS database (The LIPID MAPS Lipidomics Gateway,

<http://www.lipidmaps.org/>). The database search was performed only on even-chain lipids, with a mass tolerance of  $\pm 0.005$   $m/z$ . The list of detected compounds is available in supplementary information S3. An additional image on an adjacent section of the mouse brain tissue was recorded in the same experimental conditions to confirm the reproducibility of the spatial distribution after KMD filtering (see supporting information fig S4). Brain regions (see supporting information figure S5) were identified according to the Allen mouse brain atlas (<https://mouse.brain-map.org/>).



**Figure 5:** KMD filtering applied on MALDI imaging of a sagittal slice of mouse brain. (A) to (F) Image of sodiated glycerophosphocholines, [GPCs+Na]<sup>+</sup> with 6, 5, 4, 2, 1 and 0 insaturations respectively. (G) Image of Sodiated hexosylceramides, [HexCers+Na]<sup>+</sup>. (H) Image of sodiated lysophosphocholines, [LPCs+Na]<sup>+</sup>. (I) Image of sodiated sphingomyelins, [SMs+Na]<sup>+</sup>. The color scale represents the normalized intensity of the ions.

Interestingly, after applying our Kendrick filtering software, some of the detected GPCs from mouse brain tissue sections were easily observed to be differentially co-localized, depending on their unsaturation degree. At 6 unsaturations on the lipid chain (Figure 5.A.), GPCs are mainly detected in the isocortex, the cerebral nuclei and cerebellar cortex, while GPCs with 5 unsaturations are detected mainly in the hippocampal formation (Figure 5.B.), and the GPCs with 4 unsaturations are detected in the hippocampal formation and the olfactory areas (Figure 5.C). When the GPCs lipid chain contains one or two unsaturations, they are mainly localized in the fiber tracts (Figure 5. D and E.), but when the GPCs are fully unsaturated (Figure 5.F), they will be detected only in the cerebrum and cerebellar cortex. This variation in the localization of GPCs is in agreement with previous imaging data of lipids in brain

tissue<sup>33</sup>. Finally, additional relevant lipid distribution can also be detected, such as the specific localization of HexCers in the fiber tracts (Figure 5. G.), of LPCs mainly in the hippocampal formation (Figure 5.H.), and the SMs in the hippocampal formation, the cerebral nuclei, the olfactory area and the cerebellar cortex (Figure 5.I.).

By applying the KMD MSI data filtering approach proposed here, any type of biological modification occurring on lipids such as shorter or longer lipid chain length or lipid oxidation could be easily detected on the KMD plot. Image reconstruction based on the KMD plot provides, for a given KMD range, a specific lipid fingerprint that can further be used to compare different images.

## CONCLUSION

In this work, we have used the KMD as a filtering tool to visualize the chemically related compounds in complex mass spectra, as described in the literature. The mass resolving power and mass accuracy of modern instruments ensure an accurate KMD filtering. For overlapping families, a second filtering process using the mass distance between peaks has been introduced. Based on those selection criteria, we were able to group compounds presenting the same KMD in a single spectrum and to apply the KMD analysis to molecular images obtained by MALDI MSI where it reveals all its power. Replacing the  $m/z$  axis of the reconstructed mean mass spectrum by a KMD axis brings a real benefit for rapid and automatic image reconstruction not only for specific organic molecules but also for families, speeding up the identification process and facilitating data analysis. The developed approach has been successfully applied through two practical examples, i.e. bacterial co-cultures and histological sections of mouse brain. KMD filtering method applied to bacterial culture image data has permitted the rapid detection and localization of different groups of compounds, such as lipopeptides or lipids. The application of KMD filtering method to histological tissue sections of a mouse brain highlighted the localization of lipids according to their family and their number of unsaturation.

In light of these results, we argue that the methodology outlined in this paper can be used as a non-targeted screening tool, reconstructing the image by scanning the KMD scale according to the workflow proposed in Figure 1. Non-targeted, semi targeted or targeted mode are different options developed in the software, allowing to rapidly detect and localize metabolites containing the repetitive unit of the substrate molecule or a specific family of compounds, through their known KMD. The same approach can be extended to any type of data sets in mass spectrometry imaging with different ionization process (e.g., MALDI MSI or DESI MSI) but also when MS is hyphenated with separation methods, no matter the separation method employed (e.g., ion mobility or liquid chromatography) nor the mass analyzers used.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail : e.depauw@uliege.be.

### Author Contributions

‡These authors contributed equally.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENT

The authors want to thank Prof. Philippe Jacques, from the Microbial Processes and Interactions (MiPI) research unit of Gembloux Agro-Bio Tech (Belgium) and Lipofabrik (Lille, France) for providing lipopeptides samples.

The authors acknowledge financial support from the Excellence Of Science Program of the FNRS F.R.S (Rhizoclip EOS2018000802), from the Interreg EMR project: EURLIPIDS (R-8598) and from the European Union's Horizon 2020 research and innovation program under grant agreement No. 731077.

All experiments were done with permission from the Committee on Animal Welfare of Maastricht University, according to Dutch governmental rules.

## REFERENCES

- Schmitt-Kopplin, P.; Hertkorn, N., Ultrahigh resolution mass spectrometry. *Analytical and Bioanalytical Chemistry* **2007**, *389* (5), 1309-1310.
- Cho, Y.; Ahmed, A.; Islam, A.; Kim, S., Developments in FT-ICR MS instrumentation, ionization techniques, and data interpretation methods for petroleomics. *Mass spectrometry reviews* **2015**, *34* (2), 248-263.
- Kendrick, E., A Mass Scale Based on CH<sub>2</sub> = 14.0000 for High Resolution Mass Spectrometry of Organic Compounds. *Analytical Chemistry* **1963**, *35* (13), 2146-2154.
- Roach, P. J.; Laskin, J.; Laskin, A., Higher-Order Mass Defect Analysis for Mass Spectra of Complex Organic Mixtures. *Analytical Chemistry* **2011**, *83* (12), 4924-4929.
- Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K., Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra. *Analytical Chemistry* **2001**, *73* (19), 4676-4681.
- Marshall, A. G.; Rodgers, R. P., Petroleomics: The Next Grand Challenge for Chemical Analysis. *Accounts of Chemical Research* **2004**, *37* (1), 53-59.
- Morgan, T. E.; Ellacott, S. H.; Wootton, C. A.; Barrow, M. P.; Bristow, A. W. T.; Perrier, S.; O'Connor, P. B., Coupling Electron Capture Dissociation and the Modified Kendrick Mass Defect for Sequencing of a Poly(2-ethyl-2-oxazoline) Polymer. *Analytical Chemistry* **2018**, *90* (19), 11710-11715.
- Lerno, L. A.; German, J. B.; Lebrilla, C. B., Method for the Identification of Lipid Classes Based on Referenced Kendrick Mass Analysis. *Analytical Chemistry* **2010**, *82* (10), 4236-4245.
- Nakamura, S.; Cody, R. B.; Sato, H.; Fouquet, T., Graphical Ranking of Divisors to Get the Most out of a Resolution-Enhanced Kendrick Mass Defect Plot. *Analytical Chemistry* **2019**, *91* (3), 2004-2012.
- Zheng, Q.; Morimoto, M.; Sato, H.; Fouquet, T., Resolution-enhanced Kendrick mass defect plots for the data processing of mass spectra from wood and coal hydrothermal extracts. *Fuel* **2019**, *235*, 944-953.
- Roullier-Gall, C.; Witting, M.; Gougeon, R. D.; Schmitt-Kopplin, P., High precision mass measurements for wine metabolomics. *Frontiers in Chemistry* **2014**, *2* (102).
- Matyash, V.; Liebisch, G.; Kurzchalia, T. V.; Shevchenko, A.; Schwudke, D., Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *Journal of lipid research* **2008**, *49* (5), 1137-46.
- Youmans, K. L.; Tai, L. M.; Nwabuisi-Heath, E.; Jungbauer, L.; Kanekiyo, T.; Gan, M.; Kim, J.; Eimer, W. A.; Estus, S.; Rebeck, G. W.; Weeber, E. J.; Bu, G.; Yu, C.; Ladu, M. J., APOE4-specific changes in Abeta accumulation in a new transgenic mouse model of Alzheimer disease. *The Journal of biological chemistry* **2012**, *287* (50), 41774-86.
- Debois, D.; Ongena, M.; Cawoy, H.; De Pauw, E., MALDI-FTICR MS imaging as a powerful tool to identify *Paenibacillus* antibiotics involved in the inhibition of plant pathogens. *J Am Soc Mass Spectrom* **2013**, *24* (8), 1202-13.
- Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M. Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P., A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **2012**, *30* (10), 918-20.
- Yang, J. Y.; Phelan, V. V.; Simkovsky, R.; Watrous, J. D.; Trial, R. M.; Fleming, T. C.; Wenter, R.; Moore, B. S.; Golden, S. S.; Pogliano, K.; Dorrestein, P. C., Primer on Agar-Based Microbial Imaging Mass Spectrometry. *Journal of Bacteriology* **2012**, *194* (22), 6023-6028.
- Debois, D.; Jourdan, E.; Smargiasso, N.; Thonart, P.; De Pauw, E.; Ongena, M., Spatiotemporal Monitoring of the Antibiofilm Secreted by *Bacillus* Biofilms on Plant Roots Using MALDI Mass Spectrometry Imaging. *Analytical Chemistry* **2014**, *86* (9), 4431-4438.
- Boya, P. C. A.; Fernández-Marín, H.; Mejía, L. C.; Spadafora, C.; Dorrestein, P. C.; Gutiérrez, M., Imaging mass spectrometry and MS/MS molecular networking reveals chemical interactions among cuticular bacteria and pathogenic fungi associated with fungus-growing ants. *Scientific Reports* **2017**, *7* (1), 5604.
- Dunham, S. J. B.; Ellis, J. F.; Li, B.; Sweedler, J. V., Mass Spectrometry Imaging of Complex Microbial Communities. *Accounts of Chemical Research* **2017**, *50* (1), 96-104.
- Ongena, M.; Jacques, P., *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol. *Trends Microbiol* **2008**, *16*.
- Molina-Santiago, C.; Pearson, J. R.; Navarro, Y.; Berlanga-Clavero, M. V.; Caraballo-Rodríguez, A. M.; Petras, D.; García-Martin, M. L.; Lamon, G.; Habenstein, B.; Cazorla, F. M.; de Vicente, A.; Loquet, A.; Dorrestein, P. C.; Romero, D., The extracellular matrix protects *Bacillus subtilis* colonies from *Pseudomonas* invasion and modulates plant co-colonization. *Nature communications* **2019**, *10* (1), 1919.
- Stoeckli, M.; Staab, D.; Staufenbiel, M.; Wiederhold, K.-H.; Signor, L., Molecular imaging of amyloid  $\beta$  peptides in mouse brain sections using mass spectrometry. *Analytical Biochemistry* **2002**, *311* (1), 33-39.
- Shimma, S.; Sugiura, Y.; Hayasaka, T.; Hoshikawa, Y.; Noda, T.; Setou, M., MALDI-based imaging mass spectrometry revealed abnormal distribution of phospholipids in colon cancer liver metastasis. *Journal of Chromatography B* **2007**, *855* (1), 98-103.
- Zoriy, M.; Matusch, A.; Spruss, T.; Becker, J. S., Laser ablation inductively coupled plasma mass spectrometry for imaging of copper, zinc, and platinum in thin sections of a kidney from a mouse treated with cis-platin. *International Journal of Mass Spectrometry* **2007**, *260* (2), 102-106.
- Murphy, R. C.; Hankin, J. A.; Barkley, R. M., Imaging of lipid species by MALDI mass spectrometry. *Journal of lipid research* **2009**, *50* Suppl, S317-22.
- Burnum, K. E.; Cornett, D. S.; Puolitaival, S. M.; Milne, S. B.; Myers, D. S.; Tranguch, S.; Brown, H. A.; Dey, S. K.; Caprioli, R. M., Spatial and temporal alterations of phospholipids determined by mass spectrometry during mouse embryo implantation. *Journal of lipid research* **2009**, *50* (11), 2290-2298.
- Wenk, M. R., The emerging field of lipidomics. *Nature Reviews Drug Discovery* **2005**, *4* (7), 594-610.
- Orešič, M.; Hänninen, V. A.; Vidal-Puig, A., Lipidomics: a new window to biomedical frontiers. *Trends in Biotechnology* **2008**, *26* (12), 647-652.
- Yang, K.; Han, X., Lipidomics: Techniques, Applications, and Outcomes Related to Biomedical Sciences. *Trends in Biochemical Sciences* **2016**, *41* (11), 954-969.
- Lam, S. M.; Shui, G., Lipidomics as a Principal Tool for Advancing Biomedical Research. *Journal of Genetics and Genomics* **2013**, *40* (8), 375-390.
- Sparvero, L. J.; Amoscato, A. A.; Dixon, C. E.; Long, J. B.; Kochanek, P. M.; Pitt, B. R.; Bayır, H.; Kagan, V. E., Mapping of

phospholipids by MALDI imaging (MALDI-MSI): realities and expectations. *Chemistry and Physics of Lipids* **2012**, 165 (5), 545-562.

32. Ellis, S. R.; Paine, M. R. L.; Eijkel, G. B.; Pauling, J. K.; Husen, P.; Jervelund, M. W.; Hermansson, M.; Ejsing, C. S.; Heeren, R. M. A., Automated, parallel mass spectrometry imaging and structural identification of lipids. *Nature Methods* **2018**, 15 (7), 515-518.

33. Carter, C. L.; McLeod, C. W.; Bunch, J., Imaging of Phospholipids in Formalin Fixed Rat Brain Sections by Matrix Assisted Laser Desorption/Ionization Mass Spectrometry. *Journal of The American Society for Mass Spectrometry* **2011**, 22 (11), 1991.



# Supporting information

## Rapid visualization of chemically related compounds using Kendrick mass defect as a filter in mass spectrometry imaging

Christopher Kune<sup>a,‡</sup> and Andréa Mc Cann<sup>a,‡</sup>, Raphaël La Rocca<sup>a</sup>, Anthony Arguelles Arias<sup>b</sup>, Mathieu Tiquet<sup>a</sup>, Daan Van Kruining<sup>c</sup>, Pilar Martinez Martinez<sup>c</sup>, Marc Ongena<sup>b</sup>, Gauthier Eppe<sup>a</sup>, Loïc Quinton<sup>a</sup>, Johann Far<sup>a</sup>, Edwin De Pauw<sup>a\*</sup>.

<sup>‡</sup>These authors contributed equally

**Page 2:** Full detailed experimental section:

**Page 2: Figure S2:** KMD analysis applied on polymers:

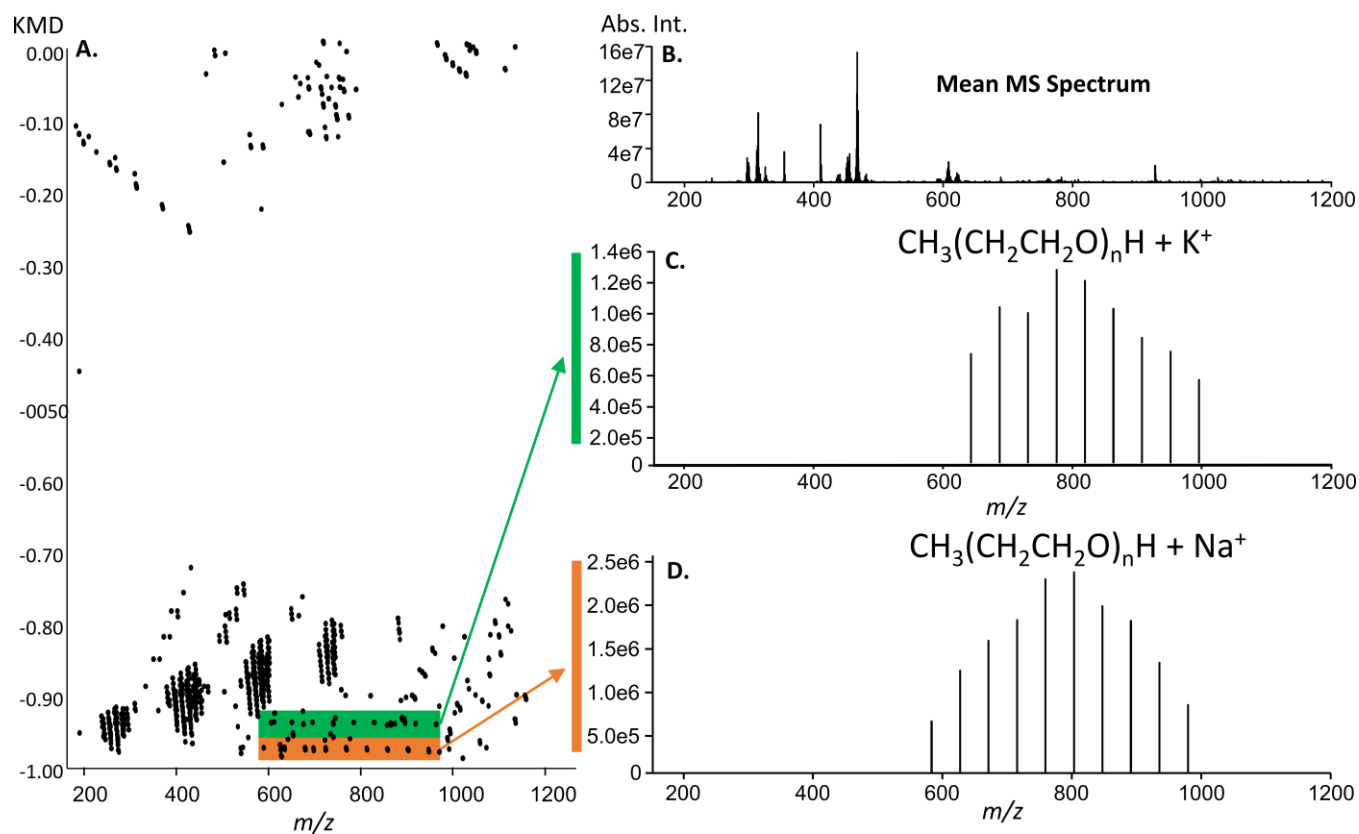
**Page 3: Figure S3:** List of identified compounds after KMD filtering applied on MALDI imaging of a sagittal mouse brain.

**Page 4: Figure S4:** KMD filtering applied on MALDI imaging of a sagittal slice of mouse brain tissue.

**Page 5: Figure S5:** Annotated sagittal mouse brain slice.

## FULL DETAILED EXPERIMENTAL SECTION:

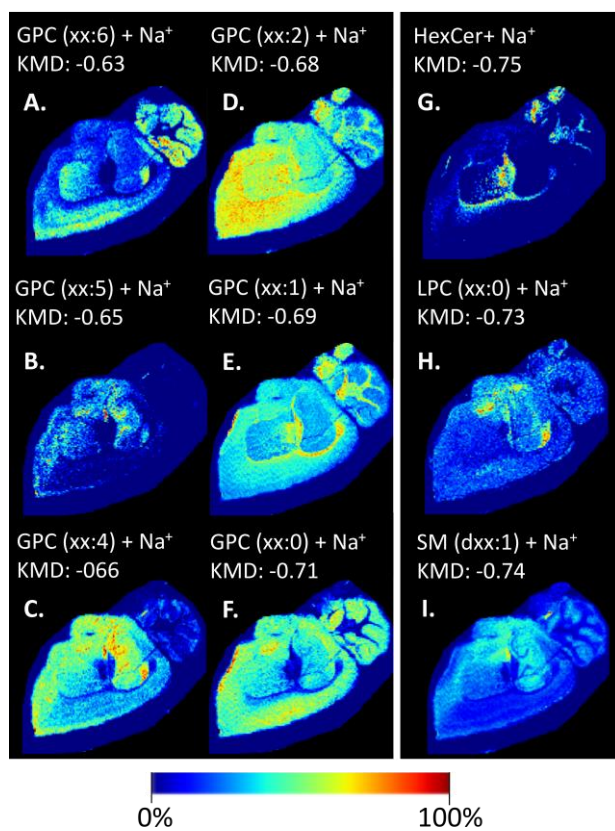
Human Glu-1-Fibrinopeptide B (Glu-Fib) and Bovine Serum Albumin (BSA) were purchased from Sigma-Aldrich (Belgium). Glu-Fib was dissolved in water with 10 % of acetonitrile and 0.1 % of formic acid to obtain a stock solution of 250  $\mu\text{M}$  and kept at  $-20\text{ }^{\circ}\text{C}$  until use. BSA was dissolved in a buffering solution of 50mM ammonium bicarbonate at pH 8. Disulfide bonds were reduced using dithiothreitol and stabilized with iodoacetamide. BSA was then digested with Trypsin (Pierce trypsin protease, MS grade from Thermo scientific) in a protein to enzyme ratio of 1/100 at  $37\text{ }^{\circ}\text{C}$  overnight. The resulting solution was dried by speed vacuum and stored at  $-20\text{ }^{\circ}\text{C}$  until use. The Glu-Fib or trypsin digested BSA was dissolved in a 50 % methanol solution with 0.1 % formic acid at a final concentration of 15  $\mu\text{M}$  for mass spectrometry infusion using 250  $\mu\text{L}$  Hamilton syringe at 4  $\mu\text{L}/\text{min}$ .



**Figure S2:** KMD analysis applied on polymers.

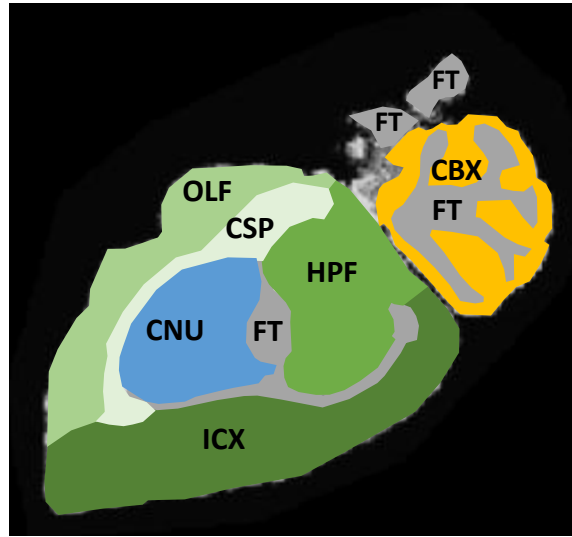
Measured $m/z$	Matched $m/z$	Delta	Name	Formula	Ion
828.5521	828.5514	0.0007	PC(38:6)	C <sub>46</sub> H <sub>80</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
856.5815	856.5827	0.0012	PC(40:6)	C <sub>48</sub> H <sub>84</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
912.6407	912.6453	0.0046	PC(44:6)	C <sub>52</sub> H <sub>92</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
746.4821	746.4731	0.009	PC(32:5)	C <sub>40</sub> H <sub>70</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
774.5091	774.5044	0.0047	PC(34:5)	C <sub>42</sub> H <sub>74</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
830.5676	830.567	0.0006	PC(38:5)	C <sub>46</sub> H <sub>82</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
748.4938	748.4888	0.005	PC(32:4)	C <sub>40</sub> H <sub>72</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
804.5523	804.5514	0.0009	PC(36:4)	C <sub>44</sub> H <sub>80</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
832.5837	832.5827	0.001	PC(38:4)	C <sub>46</sub> H <sub>84</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
860.6146	860.614	0.0006	PC(40:4)	C <sub>48</sub> H <sub>88</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
724.493	724.4888	0.0042	PC(30:2)	C <sub>38</sub> H <sub>72</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
752.5247	752.5201	0.0046	PC(32:2)	C <sub>40</sub> H <sub>76</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
780.5535	780.5514	0.0021	PC(34:2)	C <sub>42</sub> H <sub>80</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
808.5837	808.5827	0.001	PC(36:2)	C <sub>44</sub> H <sub>84</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
836.6148	836.614	0.0008	PC(38:2)	C <sub>46</sub> H <sub>88</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
864.6473	864.6453	0.002	PC(40:2)	C <sub>48</sub> H <sub>92</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
892.6779	892.6766	0.0013	PC(42:2)	C <sub>50</sub> H <sub>96</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
754.5372	754.5357	0.0015	PC(32:1)	C <sub>40</sub> H <sub>78</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
782.5684	782.567	0.0014	PC(34:1)	C <sub>42</sub> H <sub>82</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
810.5989	810.5983	0.0006	PC(36:1)	C <sub>44</sub> H <sub>86</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
838.6302	838.6296	0.0006	PC(38:1)	C <sub>46</sub> H <sub>90</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
866.6621	866.6609	0.0012	PC(40:1)	C <sub>48</sub> H <sub>94</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
728.5211	728.5201	0.001	PC(30:0)	C <sub>38</sub> H <sub>76</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
756.5518	756.5514	0.0004	PC(32:0)	C <sub>40</sub> H <sub>80</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
784.5837	784.5827	0.001	PC(34:0)	C <sub>42</sub> H <sub>84</sub> NO <sub>8</sub> PNa	[M+Na] <sup>+</sup>
794.6121	794.6116	0.0005	HexCer(t38:1)	C <sub>44</sub> H <sub>85</sub> NO <sub>9</sub> Na	[M+Na] <sup>+</sup>
822.6437	822.6429	0.0008	HexCer(t40:1)	C <sub>46</sub> H <sub>89</sub> NO <sub>9</sub> Na	[M+Na] <sup>+</sup>
850.6755	850.6742	0.0013	HexCer(t42:1)	C <sub>48</sub> H <sub>93</sub> NO <sub>9</sub> Na	[M+Na] <sup>+</sup>
753.5886	753.5881	0.0005	SM(d36:1)	C <sub>41</sub> H <sub>83</sub> N <sub>2</sub> O <sub>6</sub> PNa	[M+Na] <sup>+</sup>
781.6202	781.6194	0.0008	SM(d38:1)	C <sub>43</sub> H <sub>87</sub> N <sub>2</sub> O <sub>6</sub> PNa	[M+Na] <sup>+</sup>
809.6522	809.6507	0.0015	SM(d40:1)	C <sub>45</sub> H <sub>91</sub> N <sub>2</sub> O <sub>6</sub> PNa	[M+Na] <sup>+</sup>
837.6825	837.682	0.0005	SM(d42:1)	C <sub>47</sub> H <sub>95</sub> N <sub>2</sub> O <sub>6</sub> PNa	[M+Na] <sup>+</sup>

**Figure S3:** List of detected compounds in the MALDI imaging of a sagittal slice of mouse brain.



**Figure S4:** KMD filtering applied on MALDI imaging of a sagittal slice of mouse brain tissue.





**Figure S5:** Annotated sagittal mouse brain slice, according to the Allen mouse brain atlas (<https://mouse.brain-map.org/>). Isocortex, olfactory area, hippocampal formation and the cortical subplate are forming the cerebral cortex. Abbreviations : olfactory area (OLF), cortical subplate (CSP), cerebral nuclei (CNU), isocortex (ICX), hippocampal formation (HPF), cerebellar cortex (CBX) and fiber tracts (FT).