Application of HS-SPME-GC for Sample Classification

Vlada Olkhovych

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2023

Committee:

Robert E. Synovec
Dan Fu

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Application of HS-SPME-GC for sample classification

Vlada Olkhovych

Chair of the Supervisory Committee:
Robert E. Synovec
Department of Chemistry

The thesis presents two studies related to the analysis of volatile organic compounds (VOCs) in various samples using solid-phase microextraction gas chromatography (SPME-GC). The first study focuses on improving the reproducibility and accuracy of HS-SPME analysis by investigating standardization by use of an internal standard. The study evaluates the performance of multiple compounds as internal standards based on which effectively reduce unwanted in-class variations and have consistent peaks responses. The second study explores the feasibility of utilizing SPME-GC and chemometrics for the rapid and cost-effective classification of moisture-damaged cacao samples based on their VOC profiles. The study successfully develops a method for classifying moisture-damaged cacao samples based on their VOC profiles using SPME and GC analysis. The results demonstrate that several temperature programs were capable of distinguishing

between moisture-damaged and intact cacao samples, and several class-distinguishing peaks were found, which enabled accurate differentiation between the two sample types. The study provides valuable insights for improving the accuracy and reliability of HS-SPME analysis and contributes to the development of improved quality control protocols in the cacao industry.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor Robert E. Synovec for his invaluable guidance and support throughout this research project. His expertise, insights, and encouragement were instrumental in the successful completion of this work. I also extend my heartfelt thanks to the members of the Synovec research group for their assistance and valuable contributions to this study. Their insightful discussions and feedback were instrumental in shaping the direction and focus of this research.

I would like to extend a special thank you to my step-father, Jay, for his generous financial support that enabled me to pursue my Master's degree. Without his help, this research would not have been possible. Lastly, I would like to thank my mom for her unwavering support and encouragement throughout my academic journey. Her love and support provided me with the motivation and inspiration needed to overcome the challenges encountered along the way.

DEDICATION

For my mom. Thank you for being my best friend, my support, and my inspiration. You were never given a chance to get your Masters, but did everything so I could get mine and therefore I share this achievement between the two of us.

Chapter 1. LITERATURE REVIEW

1.1. CHROMATOGRAPHY

1.1.1. *Gas Chromatography*

Gas chromatography (GC) is a highly influential analytical technique that has revolutionized modern chemistry. Its origin dates back to the early 20th century, with the development of the theoretical basis of chromatography. However, it was not until the work of Martin and James in the early 1950s that GC became a widely accepted analytical tool.[1]

Gas chromatography, as originally developed by Martin and James in 1952, involved the use of a long narrow column packed with a solid adsorbent material such as silica gel or alumina. An aliquot of the sample mixture would be introduced into the column through which a carrier gas, such as helium or nitrogen, was flowed to carry the sample through the column. As the sample moves through the column, its components would interact with the adsorbent material in the column and be separated based on their chemical properties such as boiling point and polarity. The concept of GC has persisted as a widely used analytical technique for the separation and analysis of volatile and semi-volatile organic compounds, though details such as column construction have changed significantly. The stationary phase is now often a coating on the inside of a fused silica column as opposed to a packed material, and a variety of stationary phases are now available to allow the separation of sample components based on their physical and chemical properties such as boiling point, polarity, and size based on the compounds' affinity to the stationary phase. The separated components are then sent to a detector, which generates a signal that is proportional to the amount of each component as interpreted by a connected computer that transforms the detector signal into a chromatogram.

As mentioned, there are a wide range of choices in stationary phase from which the user selects based on the nature of the sample and the compounds of interest. Common stationary phases include polar and non-polar materials, such as polydimethylsiloxane (PDMS) and polyethylene glycol (PEG). The mobile phase, the carrier gas that is used to transport the sample through the GC column, can also be selected out of several options, with common choices including helium, nitrogen, and hydrogen. In addition to the selections of stationary phase and carrier gas, the choice

of flow rate and pressure can affect the efficiency and selectivity of the separation. The dimensions of the GC column can affect the efficiency and resolution of the separation. Furthermore, longer columns generally provide better separation, but can require extended run times and may lead to band broadening.

The temperature of the column is a critical parameter in GC, with advancements in column heating playing a significant role in method development. It remains the primary means by which the extent to which the sample components interact with and absorb onto the stationary phase, and often has significant impact on the efficiency and selectivity of the separation accordingly. The column temperature is typically programmed during the analysis to optimize the separation of the sample components. Optimization of a temperature program is an important step in GC analysis since it defines separation longevity, efficiency, and resolution.

The detector is the system component that detects the sample components as they elute from the GC column. Common detectors include flame ionization detectors (FID), thermal conductivity detectors (TCD), and mass spectrometers (MS). The choice of the detector depends on the aim of the study, investigated mixture composition, and the available budget. GC-MS systems provide the most complete qualitative information about the sample but MS detectors are expensive, slow, large and are unsuitable for on-site analysis. The choice of the detector depends on the sensitivity and selectivity required for the analysis.

An injector is part of a GC system, where a sample is injected into the instrument. An injection volume and mode are important parameters in GC. Injection modes include split, splitless, and on-column injection, each of which has different advantages and disadvantages.

GC is widely used in the analysis of volatile and semi-volatile organic compounds in a variety of industries, including pharmaceuticals, environmental analysis, food and beverage, and forensic investigations. It can be used to analyze a wide range of samples, including air, water, soil, and biological fluids.

1.1.2. *Intuvo 9000 Gas Chromatography (GC) System*

GC instruments are versatile and reliable systems, which makes them one of the most commonly used instruments in the field of analytical chemistry, though they still have some limitations. Usually trained and qualified personnel are needed to operate and maintain GC instruments due to the complexity of the instruments. Another shortcoming accompanying the application of GC in compositional analysis is the relatively long duration of the analysis, which is usually in the range of 10-60 min. This limits the use of GC in applications requiring high throughput and motivates some users to choose liquid chromatography (LC) as a more efficient (if more expensive) alternative.

The Intuvo 9000 is a GC instrument manufactured by Agilent Technologies to overcome some of the shortcomings of conventional GC instruments. It is designed to provide high-performance chromatography and accurate quantitative analysis of samples in various fields, including the food, environmental, pharmaceutical, and chemical industries. The Intuvo 9000 employs a high-speed column heating approach equipped with a novel direct heating technology, that is faster and more efficient that an air-bath oven technology used in conventional CG systems. It is capable of heating up from 30 °C to 450 °C at 250 °C/min, allowing for fast and accurate temperature programming and producing linear and repeatable temperature ramps.[2] Another remarkable feature of the Intuvo 9000 is that all the nuts and ferrules typically required in the flow path have been replaced with leak-free click-and-run direct connections. This feature reduces the time and complexity of instrument setup and maintenance, which allows a user with minimal qualifications to operate and maintain the instrument.

The Guard Chip technology implemented by the Intuvo 9000 is another significant advancement in chromatographic separation technology. It plays a crucial role in preventing the deposition of high molecular weight material on the analytical column. Before the introduction of the Intuvo, the only way to achieve this was clipping the end of the column as it became fouled. However, this practice required shortening of the column, thus often leading to appreciable retention time shifting, which resulted in additional time and effort being spent in adjusting conditions to prevent compromises in data integrity. By trapping unwanted materials before they enter the column, the Guard Chip helps to maintain constant retention times through enabling consistent column dimensions, thus avoiding these complications. Additionally, the Guard Chip

minimizes unplanned downtime resulting from faulty connections that can cause leaks or failed sample results, which could require additional troubleshooting, remediation, and repeat sample analyses. Changing the Guard Chip and so removing the trapped material from the system is a straightforward process that takes approximately one minute due to its use of click-and-run connections. Overall, the Guard Chip enhances chromatographic separation efficiency and reduces associated risks, making it a valuable addition to the analytical toolkit.

1.2. Solid-Phase Microextraction (SPME)

1.2.1. *Sample preparation techniques*

In analytical chemistry, a typical experiment involves several stages, including sample collection, sample preparation, transfer of the sample to an analytical instrument, instrumental analysis, and data analysis and interpretation. Among these stages, sample preparation can be particularly problematic, as it has perhaps the highest potential to introduce errors in the experiment outcome. Classical sample preparation techniques, such as liquid-liquid extraction (LLE), are time-consuming, multi-step processes that involve significant amounts of high-purity organic solvents. The LLE procedure is based on the partitioning of the sample between two immiscible liquids, and is labor-intensive, time-consuming, has limited selectivity, and has relatively low repeatability. Moreover, since the LLE procedure is divided into multiple individual stages, each step can introduce errors and losses, which can result in decreased analyte concentrations and experimental accuracy. The high solvent usage increases the demand for safe waste disposal, which adds another step to the experiment and creates an environmental hazard that needs to be addressed.

The use of solid-phase extraction (SPE) devices allows us to overcome some of these limitations of LLE. SPE separation is based on partitioning between a liquid sample matrix and a solid phase of sorbent material. Typical solid phase extraction procedure includes conditioning of SPE cartridge, loading of a sample into a cartridge, washing off interferences, while analytes of interest stay bonded to the inner liner of a cartridge, followed by eluting of analytes of interest.[3,4] (Figure 1.1) SPE makes sample preparation faster and less complex, requires smaller amounts of solvents as well as lower-purity solvents, and has increased reproducibility and selectivity compared to LLE techniques.[5] But despite all its advantages, SPE is still time-consuming and can

introduce contaminations into a sample due to the adsorption of compounds on the walls of the device with the accompanying risk of inadvertent reintroduction to different systems. Additionally, while solvent amounts used in SPE are smaller compared to LLE techniques, they are still significant.[6] In both LLE and SPE-based methods, volatile and semi-volatile compounds are most susceptible to losses. It makes studying volatile organic compounds (VOCs) using traditional sample preparation techniques particularly complicated.

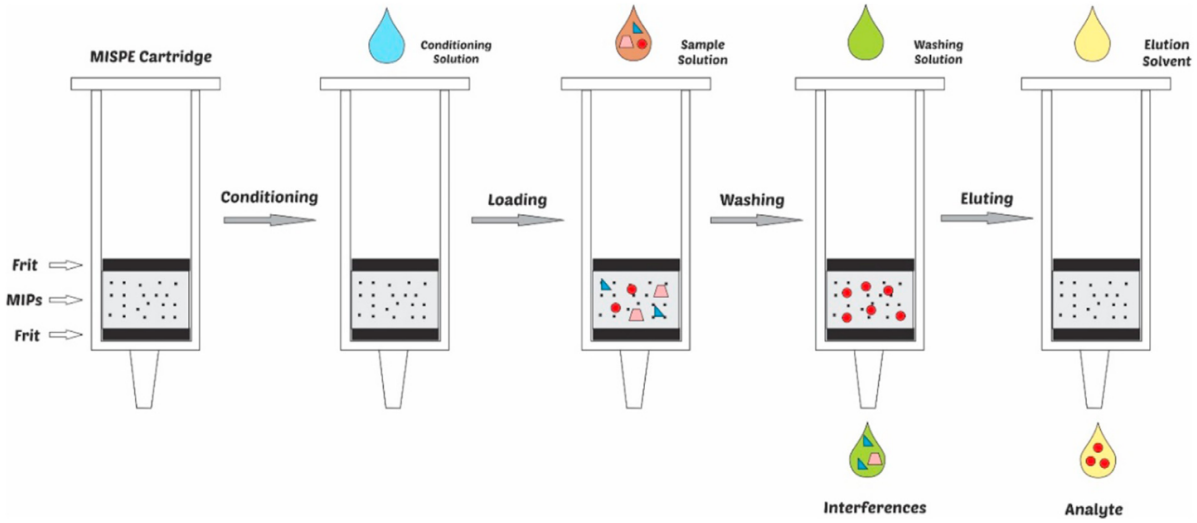

Figure 1.1. Typical solid-phase extraction procedure.[4]

1.2.2. *SPME fundamentals*

Solid-phase microextraction (SPME) was developed by Pawliszyn and Belardi in 1989 to address the shortcomings of LLE and SPE techniques.[7] SPME is a simple, non-exhaustive, and versatile sample preparation technique that proved to be well-suited for a large variety of analytes and sample classes.[8] The application of SPME allows for combining into a single-step sample collection, extraction, analyte enrichment, analyte isolation from a sample matrix, and introduction of a sample into an analytical instrument.

Key features of SPME that differentiate it from LLE and SPME methods are:

- Extractant has a substantially smaller volume (typically in the μL range) than a sample (typically in the mL range), so at equilibrium the extracted volume can be neglected, making technology non-exhaustive.[8]

- Solvent-free or low solvent consumption.
- Inexpensive and environment friendly.
- Allows for on-site and in-vivo applications.
- Applicable for all kinds of sample matrixes.
- Can be automated. [9]

The most commonly used SPME technique involves extracting a small amount of sample onto a sorbent phase of a fiber-shaped device. The extraction is complete once concentrations of compounds both in the sorbent and the sample matrix reach chemical equilibrium. After equilibrium is reached the concentrations of analytes on fiber stop increasing and further exposure of a sorbent to a sample is not needed. If we consider only two phases, the state of equilibrium between an SPME fiber material and a sample matrix can be described using equation 1.1.[10] This equation can be applied to describe direct immersion (DI) SPME since this technique involves two phases directly contacting with each other: sample matrix and a fiber.

$$C_0 V_s = C_s^{\infty} V_s + C_f^{\infty} V_f \tag{1.1.}$$

Where:

C_0 - initial concentration of an analyte in the sample

V_s - sample volume

C_s^{∞}-concentration of analyte in the sample at the equilibrium

C_f^{∞}- concentration of analyte in the fiber at the equilibrium

V_f −volume of a fiber coating

The analyte distribution coefficient (K_{fs}) between the sample matrix and a fiber coating is described as follows in equation 2.

$$K_{fs} = \frac{C_f^{\infty}}{C_s^{\infty}} \tag{1.2.}$$

The concentration of analyte in the fiber at the equilibrium (C_f^{∞}) can be found as described in equation 1.3.

$$C_f^{\infty} = C_0 \frac{K_{fs} V_s}{K_{fs} V_f + V_s} \tag{1.3.}$$

The number of moles of analyte in a sample (n) with a known volume is described in equation 4.

$$n = C_f^{\infty} V_f = C_0 \frac{K_{fs} V_s V_f}{K_{fs} V_f + V_s} \quad (1.4.)$$

When the volume of the sample is much larger than the volume of the SPME extraction phase, some parameters can be neglected and an equation 1.3 can be reduced to equation 1.5. This approach is useful when the sample volume is unknown and is applicable most of the time since normally in SPME $C_f^{\infty} \ll C_s^{\infty}$ and $V_f \ll V$.

$$n = K_{fs} V_s C_0 \quad (1.5.)$$

A typical SPME experiment consists of several stages and processes as described in the table 1.1.

Table 1.1.1. Processes involved in the SPME procedure

Process	Description
Selection of fiber	The selection of the fiber coating is a critical aspect of SPME as it determines the type of compounds that can be extracted. The most commonly used fiber coatings are polydimethylsiloxane (PDMS), polyacrylate (PA), and polydimethylsiloxane/divinylbenzene (PDMS/DVB).
Conditioning	The SPME fiber is conditioned by exposing it to high temperature and pressure to remove any contaminants or impurities.
Sample preparation	The sample can be prepared by dissolving the analyte in a suitable solvent or buffer.
Extraction	The SPME fiber is exposed to the sample for a specific time, allowing the analyte to partition between the sample and the fiber coating.

Desorption	The SPME fiber is removed from the sample and inserted into a gas chromatography (GC) injection port or a thermal desorption unit, where the analyte is desorbed by heating.
Instrumental analysis	The sample components are separated and detected in a GC instrument.

1.2.3. *Types of SPME*

There are three modes of operation in SPME: headspace SPME (HS-SPME), direct immersion SPME (DI-SPME), and a membrane-protected SPME (M-SPME) (Figure 1.2). In DI-SPME (Figure 1.2. a) a sorbent phase of an SPME instrument is inserted into a sample matrix and analytes are transmitted directly from a sample into the adsorbent. Direct immersion allows for the most complete extraction of both volatile and non-volatile compounds of a sample, but the main shortcoming of this method is the short lifespan of the absorbing materials. In HS-SPME (Figure 1.2 b) the sorbent is placed in the headspace of a sample-containing container without direct contact between an extractant and a sample. Analyte transfer in HS-SPME happens between the sample, sample headspace, and the SPME fiber, which is coated in a fused-silica sorbent material. HS-SPME (Figure 1.2 b) is the most useful for the extraction of volatile and semi-volatile analytes in complex samples but often excludes non-volatiles as well as a significant fraction of semi-volatile analytes from the analysis. Despite a less complete extraction profile, HS-SPME is usually the method of choice for the analysis of volatile and semi-volatile compounds since it leads to a purer extract, greater selectivity, and prolonged lifespan of SPME fiber. Both HS-SPME and DI-SPME have limited application when a sample contains non-volatile high-molecular-interfering compounds such as proteins, fatty materials, and humic acids. A membrane-protected SPME (Figure 1.2 c) technique allows overcoming these limitations by placing an extracting SPME fiber in a hollow cylindric membrane, which prevents interfering compounds from entering the extraction phase while still allowing a mass transfer of analytes of interest onto a fiber.[11,12] An in-tube SPME was developed for application in high-performance liquid chromatography (HPLC). In in-tube mode, an SPME device is presented as a capillary column, placed between an injection needle and an injection loop, and connected to an autosampler in an HPLC instrument.

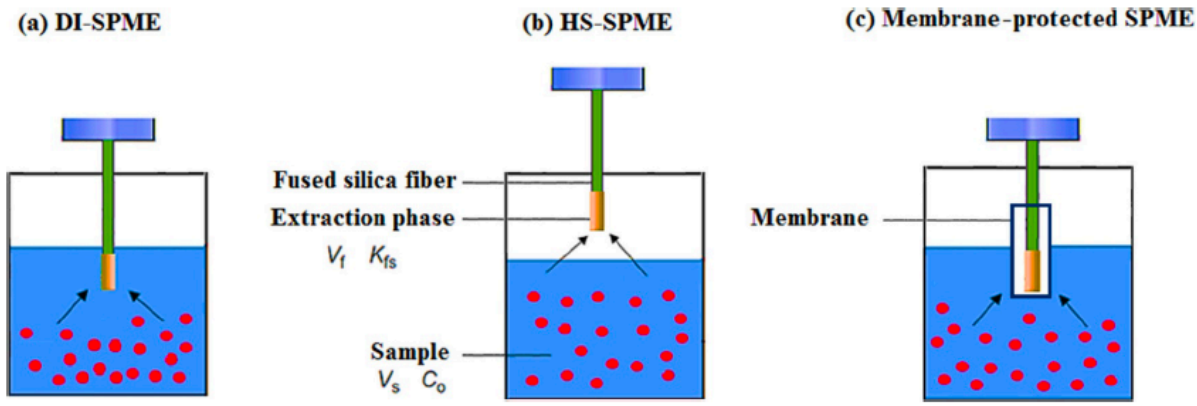

Figure 1.2. Modes of operation in SPME. (a) direct immersion (DI-SPME), (b) headspace (HS-SPME), (c) membrane-protected (M-SPME).[12]

1.2.4. *Headspace Solid-Phase Microextraction (HS-SPME)*

In HS-SPME, the fiber coated with a stationary phase is exposed to the headspace above a sample (Figure 1.3). The volatile compounds in the headspace diffuse through the fiber coating and partition between the stationary phase and the headspace until equilibrium is reached. The concentration of the volatile compounds on the fiber is proportional to their concentration in the headspace. When the fiber is inserted into a heated injection port of a gas chromatograph, the volatile compounds desorb from the fiber and are transported by the carrier gas to the column for separation and detection (Figure 1.3). The signal intensity of detected volatile compounds is proportional to their concentration on the fiber.

HS-SPME extraction involves three phases: sample matrix, sample headspace and a fiber coating. The state of equilibrium in this system can be described using equation 1.6.[10]

$$V_s C_0 = C_f^{\infty} V_f + C_h^{\infty} V_h + C_s^{\infty} V_s \qquad (1.6.)$$

Where:

C_0 - initial concentration of an analyte in the sample

V_s - sample volume

C_s^{∞}-concentration of analyte in the sample at the equilibrium

C_f^{∞}- concentration of analyte in the fiber at the equilibrium

V_f −volume of a fiber coating

C_h^{∞}- concentration of analyte in the sample headspace at the equilibrium

V_h −volume of a sample headspace

The analyte quantity (n) in HS-SPME can be calculated using Equation 1.7.

$$n = \frac{K_{fh}K_{sh}V_f C_0 V_s}{K_{fh}K_{sh}V_f + K_{hs}K_h + V_s} \quad (1.7.)$$

Where,

$$K_{fs} = K_{fh}K_{hs} = K_{fg}K_{gs} \quad (1.8.)$$

This allows us to shorten the equation above to

$$n = \frac{K_{fs}V_f C_0 V_s}{K_{fs}V_f + K_{hs}K_h + V_s} \quad (1.9.)$$

Several parameters affect the efficiency and sensitivity of the HS-SPME method. These parameters include the type of fiber coating, extraction temperature, extraction time and agitation rate. Optimal parameters for a particular application can be determined by a method validation process. The fiber coating selection is a particularly important parameter of HS-SPME as it determines the type of compounds that can be extracted. The stationary phase should have a high affinity for the analytes of interest while allowing the analytes to be released in inlet of the gas chromatograph and not interfering with the separation. The fiber should also be stable and not react with the sample matrix or the stationary phase. The selection of the fiber coating is a critical aspect of HS-SPME as it determines the type of compounds that can be extracted. The most commonly used fiber coatings are polydimethylsiloxane (PDMS), polyacrylate (PA), polydimethylsiloxane/divinylbenzene (PDMS/DVB) and three-phase divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS).

HS-SPME has numerous advantages over other extraction techniques, including simplicity, speed, low cost, and reduced sample preparation time. It is a non-destructive technique that allows for the re-analysis of the same sample if needed. Due to its non-destructive nature, high sensitivity, and ability to analyze a broad range of compounds, it is widely used for sample preparation and analysis. Some of the most common applications include environmental analysis, pharmaceutical analysis, analysis of food and beverages, forensic analysis and biomedical research. In environmental analysis, HS-SPME has been employed for detecting and quantifying

volatile organic compounds (VOCs) in air, water, and soil samples. In pharmaceutical analysis, HS-SPME is used for detecting trace impurities and degradation products in drug formulations.

HS-SPME (headspace solid-phase microextraction) coupled with GC (gas chromatography) is a particularly useful technique for the analysis of volatile organic compounds (VOCs) in food and beverages. It is widely used in food industry for quality control and safety assessment, flavor analysis, and product development. HS-SPME-GC enables the identification and quantification of volatile flavor compounds, such as alcohols, aldehydes, ketones, esters, and terpenes, which are critical for evaluating the sensory properties and quality of food and beverages. It also aids in developing new flavors and fragrances by allowing the identification of compounds responsible for desirable and undesirable flavor attributes.

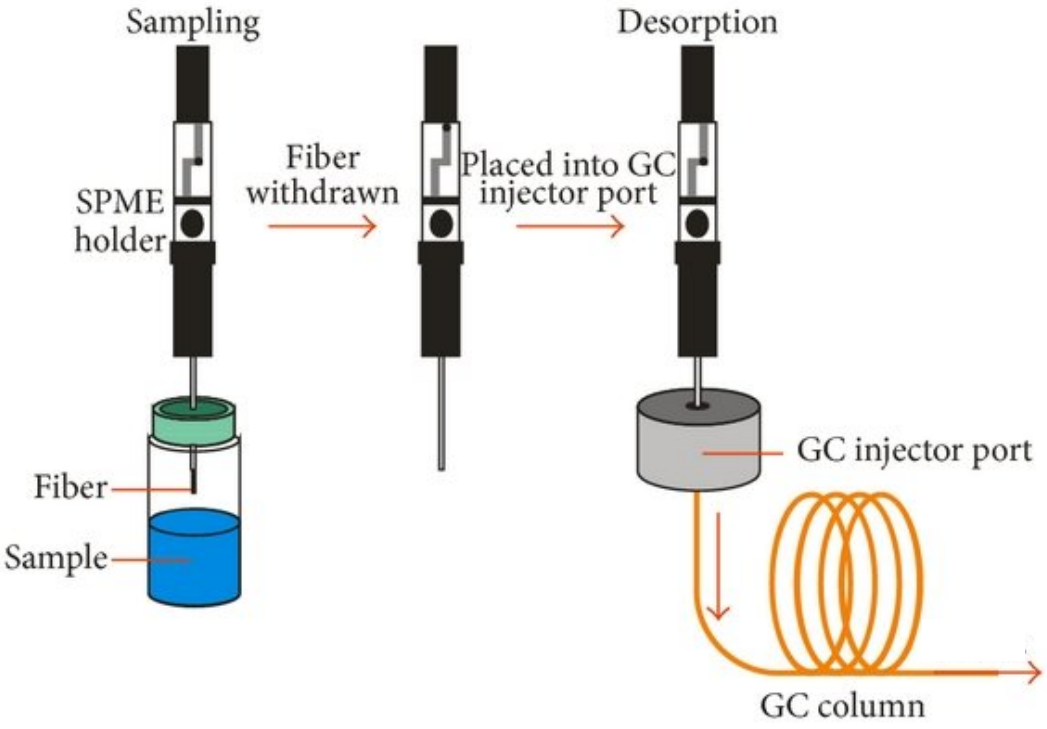

Figure 1.3. HS-SPME procedure includes sampling (extraction), injection into GC port and thermal desorption.[13]

Chapter 2. HS-SPME INTERNAL STANDARDS IN GC-FID

2.1. INTRODUCTION

Despite the advantages of HS-SPME, it has some limitations and shortcomings like any other analytical technique. The key limitations of HS-SPME are listed in the Table 2.1.

Table 2.1. Limitations of HS-SPME

Limitation	Description
Limited sample volume	The sample size for HS-SPME is limited due to the small surface area of the SPME fiber. Therefore, it may not be suitable for the analysis of large sample volumes.
Limited sensitivity	HS-SPME is generally less sensitive than other techniques of introducing the sample into gas chromatography (GC) systems. This is because the amount of analyte that can be extracted by the SPME fiber is limited.
Fiber coating interference	The fiber coating can sometimes interfere with the analysis by either reacting with the sample or masking some of the compounds.
Fiber breakage	The SPME fiber is fragile and can break easily. This can cause a loss of samples and result in inconsistent or inaccurate results.
Matrix effects	The matrix effect refers to the influence of the sample matrix on the extraction and analysis. HS-SPME can be affected by the matrix effects, which can cause issues with the accuracy and reproducibility of the results.
Limited selectivity	HS-SPME has limited selectivity, which means that it may have a limited application for the analysis of some complex samples with a wide range of analytes.

One of the most significant drawbacks of SPME compared to direct injection is the lower reproducibility of data. Some factors resulting in lower reproducibility of SPME are listed in Table 2.1. SPME has significantly higher variability compared to direct injection methods due to the additional steps during sample delivery to an analytical instrument in SPME methods. In the case of HS-SPME, such processes include the partitioning of compounds between the sample matrix and the sample headspace and the ongoing processes of absorption and desorption of analytes on the fiber. The resulting concentration and composition of the sample on the fiber depend on the state of equilibrium of the investigated system, and this state of equilibrium is highly dependent on both temperature and pressure in and outside of the system. Even minor variations in the system parameters can lead to variations in HS-SPME data.[6,7]

The carryover of analytes from one sample to the next can occur, which can lead to false positives or overestimation of analyte concentrations. Repeated use of a SPME fiber may lead to degradation of the fiber coating, affecting extraction efficiency and selectivity. Some analytes may be thermally unstable and degrade during the HS-SPME process, leading to inaccurate results.

A 2015 study performed by Aqel et al. reports how the application of different substrates affected the sample headspace composition.[14] They applied a volatile organic mix onto four different types of fabric including cotton print cloth, viscose challis, spun polyester and wool flannel. Their findings suggest that the presence of a substrate can substantially alter the sample headspace. Sample matrix composition and morphology can significantly affect relative concentrations of volatile compounds in the sample headspace.[14–17] It is important to carefully consider these potential problems when using HS-SPME and to optimize the technique for the specific analyte and sample matrix of interest.

One way to address these issues is using an internal standard. In analytical chemistry, an internal standard is a substance that is added to a sample in a known quantity before analysis, typically in a quantitative analysis, to improve accuracy and precision. The internal standard is chosen to be similar in chemical and physical properties to the analyte of interest, but with a different mass or composition, which allows it to be distinguished from the sample analytes during analysis. The internal standard is added to the sample in a known quantity, and its signal is measured along with the analyte signals during the analysis. The ratio of the internal standard

signal to the analyte signal is used to correct for any variations in the analysis that might affect the accuracy or precision of the results.

When using an internal standard in SPME, the internal standard is added to the sample before extraction, so that both the analyte of interest and the internal standard are extracted by the SPME fiber and then desorbed and analyzed using an appropriate analytical instrument. Using the extraction efficiencies of the analyte and the internal standard, it is possible to correct for any variability in the extraction efficiency of the SPME fiber. This correction can improve the accuracy and precision of the results obtained from SPME, including by helping to account for any losses of analyte that may occur during sample preparation or analysis.

In this study, we have two main goals to achieve: to compare the quality of the data collected with HS-SPME to that of direct injection and improve the reproducibility of the HS-SPME method using the internal standardization method. To the improve quality of SPME data, we investigated standardization via the internal standard method. Multiple compounds were investigated as internal standards and their performance was evaluated and compared.

A sample mixture was applied on the piece of Whatman filter paper to imitate a real-life sample that consist of three phases:

- Solid matrix (paper)
- Liquid phase (sample mix)
- Gaseous phase (sample headspace)

2.2. EXPERIMENTAL

2.2.1. *Sample preparation*

Three samples were prepared in this study (Table 2.2). All analytes were at least HPLC grade. 15 assorted compounds were diluted in methanol with the final concentration of compounds highlighted in Table 2.2. The original mixture was then split into three parts, with the first part left as a base mixture to form the first sample class investigated in the study, which is referred to as 15 Mix. The two other parts were spiked with two different and nonoverlapping sets of 5 native compounds to form two spike classes: Spike 1, as highlighted blue in Table 2.2, and Spike 2 which is highlighted green in Table 2.2. The final concentrations of the spiked analytes listed in the Spike

Concentration column in Table 2.2. The solutions were used with direct injection (DI) analysis and the Standard Addition Method to determine their retention times.

To create complex samples similar to those found in real-life HS-SPME analysis aliquots of samples were spiked onto solid matrices and sealed in headspace vials. The matrix used was a whole piece of 3 cm diameter Whatman filter paper that was cut into 8 pieces each before being added to a 10 ml SPME vial. 100 microliters of a sample solution was then added with the vial being immediately sealed to prevent volatile components from evaporating. The spiked paper samples were used for HS-SPME analysis.

Table 2.2. Sample mixtures composition. Blue rows refer to Spike 1, green rows refer to Spike 2, and unmarked rows that had consistent concentrations in all three samples including 15 Mix. sample.

Compound	Base Concentration (ppm)	Spike Concentration (ppm)
1-butanol	66.8	271
toluene	66.7	286
m-xylene	66.6	66.5
1,4-thioxane	66.5	66.4
mesitylene	66.9	285
1-decene	66.6	254
1,5-dichloropentane	95.7	374
5-decyne	66.6	66.6
1-octanol	66.7	66.6
undecane	66.6	254
1-bromooctane	66.7	66.6

2-decanone	66.8	288
methyl salicylate	66.9	363
methyl decanoate	66.5	287
bicyclohexyl	66.5	285

2.2.2. *Headspace solid-phase microextraction (HS-SPME)*

The sample was extracted and delivered to the Intuvo-GC system using HS-SPME. For the preconcentration and extraction of analytes, a 50/30 µm DVB/CAR/PMDS HS-SPME (Supelco, USA) fiber was used. Initially the extraction fiber was conditioned at 270 °C in the inlet of the GC for 30 min. During conditioning, flow in the instrument was held at a flow rate of 1 mL/min and the oven was kept at 250 °C to prevent any residual compounds coming off the fiber from contaminating the GC column. Each sample vial was placed in a bead bath and heated to 60 °C for 30 min before extraction to ensure that the system reached equilibrium. Extraction was performed for 30 min under a temperature of 60 °C. After each extraction-desorption cycle the fiber was cleaned via additional desorption in the GC inlet for 5 min at 250 °C after each GC run.

2.2.3. *Gas chromatography*

Instrumental analysis was performed using the Intuvo 9000 GC system equipped with a flame ionization detector (FID). A HP-5MS (5%-diphenyl-methylpolysiloxane) nonpolar column, with dimensions 30 m length × 250 µm inside diameter, film thickness of 0.25 µm, was used for the separation. After the SPME fiber extraction, analytes were introduced to the GC system by desorbing SPME fiber in the GC inlet at 250 °C for 1 min in splitless mode. The temperature program for the GC system was set as follows: the initial temperature was set at 40 °C and held for 1.4 min before the temperature was increased to 300 °C at the rate of 100 °C/min and then finally held at 300 °C for 1 min. The carrier gas flow rate was 3 mL/min, with hydrogen as the carrier gas. The entire GC run time was 5 min per injection. The fiber was then cleaned in the GC port for 5 min as described above. After fiber cleaning an instrument blank was collected to check for the absence of any contaminations on the column left after separation and fiber cleaning steps. For each of the 3 investigated samples, 10 chromatograms were collected resulting in 30

chromatograms in total. The FID was set at 1 kHz. The OpenLab CDS (Agilent Technologies, Palo Alto, CA, USA) software was employed for data collection, data procession, instrumental control and method adjustments.

To have a reference for proper evaluation of our data quality direct injection measurements were collected. Direct injection chromatograms were collected using the same method described above with minor modifications. 10 replicas for each sample were collected with 30 chromatograms in total. Direct injection was performed using an autoinjector. The GC temperature program used for direct injection was the same as described above.

2.2.4. *Data analysis*

Data analysis was performed using MATLAB R2022b (The Mathworks, Inc., Natick, MA, USA). A rolling ball minimum algorithm was employed to perform baseline correction on the chromatograms. Analyte identification was performed based on analyte retention times and boiling points (Table 2.3) and confirmed by the Standard Addition Method. Inter-class differences between samples were studied using principal component analysis (PCA). The success of standardization using different internal standards was evaluated using a degree of class separation (DCS) method. The average degree of class separation was calculated as a mean of DCS between all three samples.

2.3. RESULTS AND DISCUSSION

2.3.1. *Direct injection*

Figure 2.1 shows the chromatographic overlay of all 30 chromatograms collected using the direct injection method. The chromatograms colored in red on the plot are 15 Mix, green is Spike 1, and blue is Spike 2. Figure 2.1. b shows a zoom-in of two peaks with the longest retention times-methyl decanoate (14) and bicyclohexyl (15). Compound 14 is spiked in the Spike 2 sample and compound 15 is in Spike 1. Sample classes are easily identifiable visually by the difference in relative peak heights between samples. The chromatographic overlay shows some minor shifting and peak height variation, but these effects are minor enough to be neglected.

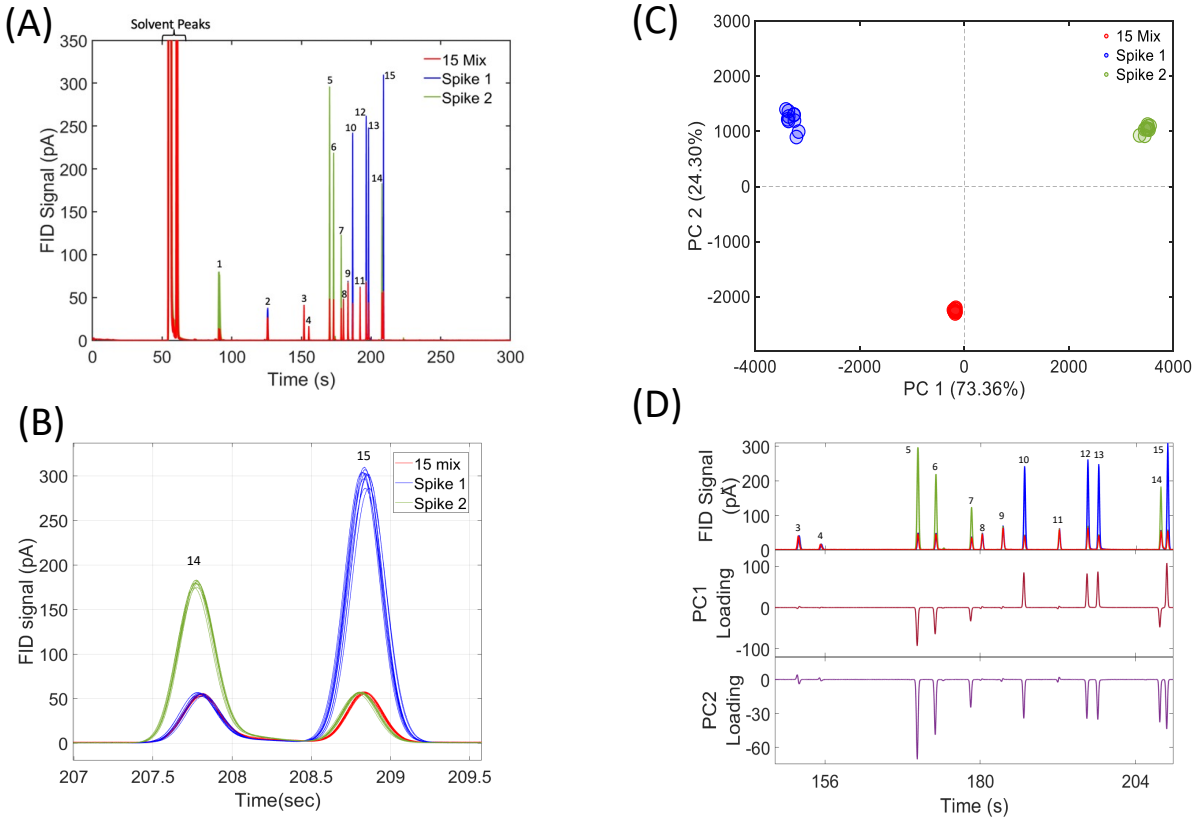

Figure 2.1. Direct injection chromatographic overlay. Red color refers to 15 Mix, blue to Spike 1, and green to Spike 2. (A) Full chromatographic overlay. (1) 1-butanol, (2) toluene, (3) m-xylene, (4) 1,4-thioxene, (5) mesitylene, (6) 1-decene, (7) 1,5-dichloromethane, (8) 5-decyne, (9) 1-octanol, (10) undecane, (11) 1-bromooctane, (12) 2-decanone, (13) methyl salicylate, (14) methyl decanoate, (15) bicyclohexyl. (B) Methyl decanoate (14) and bicyclohexyl (15) peaks zoom in. (C) PCA for DI analysis. (D) Loadings for PCA on DI analysis.

In this study, PCA was employed to analyze unsupervised classification of the three sample classes. The results, as shown in Figure 2.1 c, have a high degree of class separation (DCS) with an average separation of 67.4 (Table 2.5). The sample classes were tightly grouped, with the 15 Mix class showing the highest precision and the Spike 1 class showing the lowest precision. The first principal component (PC1) captured 73.36% of the variation in the dataset, and was able to effectively separate all sample classes.

Plotting the loadings of PCA data provides valuable insights into the distribution of analytes and their correlations in the data set that then impact sample clustering in the scores plot, which can guide the selection of suitable variables for subsequent analyses. The PC1 loadings plot showed that peaks corresponding to different sample classes were assigned different values, with Spike 2 peaks having a negative loading and Spike 1 peaks having a positive loading. The absence of unspiked peaks on the loadings plot is due to their consistent concentrations throughout the experiment and their lack of contribution to the compositional variation in the dataset (which further indicates that analyte signal responses were consistent across runs). Additionally, the second principal component (PC2) contributes 24.3% of the variation in the dataset. It successfully clusters 'spiked' and 'unspiked' samples, and the loadings correlate to spiked analyte peaks. Spread in spiked clusters could and likely are related to retention time shifting or other inconsistencies in the chromatogram though as well. The direct injection measurement results were utilized as a reference for evaluating the performance of HS-SPME.

2.3.2. *Headspace solid-phase microextraction (HS-SPME)*

A total of 30 chromatograms with 10 replicates for each of the three sample classes were collected using the HS-SPME sample preparation and delivery technique described in the sample preparation section. Since all injections were performed manually, sample-to-sample variation not caused by differences in the sample composition is more pronounced compared to automated HS-SPME systems. Table 2.3 provides a list of analytes present in the investigated samples, arranged in order of elution time and boiling points. The retention times were calculated as the mean value in the 30 direct injection chromatograms. The numbering of compounds follows the ascending order of retention time and boiling points, which is in accordance with the fundamental principles of gas chromatography. The rows in the table are color-coded to represent the classes of the compounds. Green indicates compounds spiked in the Spike 1 sample, blue indicates those spiked in Spike 2, and white indicates analytes that have consistent concentrations in all three samples, including Spike 1, Spike 2, and 15 Mix respectively. Analytes marked white were considered as potential internal standards. However, analytes 3 and 4 were not selected as potential internal standards due to their wide widths and inconsistent peak shape. It is worth noting that the first two analytes, with the lowest retention times and boiling points, co-eluted with the solvent (methanol with some minor contaminants), and thus their peaks are not visible in the HS-SPME chromatographic overlay as they are obscured by solvent peaks.

Figure 2.3 depicts a chromatographic window that features peaks 3 through 15. Out of the 15 analytes in the samples, only 11 were analyzed in detail. Compounds 8, 9, and 11 correspond to 5-decyne, 1-octanol, and 1-bromooctane, respectively, and were investigated as potential internal standards in this study. The graph appears to show compounds with higher boiling points having better peak quality (narrower peaks, smaller width/peak height ratio, higher resolution). Table 2.4 presents a comparative analysis of the peak widths obtained from the two different sample delivery methods investigated in this study. Specifically, it compares the peak widths of the direct injection (DI) method with those of the HS-SPME technique. The results indicate that on average, the DI method produces narrower peak widths than the HS-SPME method. Notably, the HS-SPME method shows wider initial peak widths, which progressively narrows as the retention times of analytes increase. Analytes desorbing from the SPME stationary phase suffer from peak broadening as analyte molecules do not desorb simultaneously but rather follow an approximately Gaussian distribution. Analytes with higher affinity to the stationary phase can in effect be trapped in the stationary phase in the start of the column, thus reducing this spreading effect and allowing for a more efficient separation.[18] This phenomenon is known as refocusing and results in narrower peaks with better resolution, making it easier to identify and quantify individual compounds.

Figure 2.4 displays the close-ups of all the investigated peaks in both the HS-SPME and DI sample delivery methods. The results reveal that the peaks obtained from the DI method are consistent and reproducible across all repetitions and the retention time scale. In contrast, the HS-SPME peaks exhibit less consistency, with broader peaks that change over time. The HS-SPME method involves several complex and interconnected steps, including extraction, sample enrichment, and delivery, which may introduce variability in the results. Moreover, the interaction of the analytes with the SPME fiber may not be uniform, resulting in varying responses and peak shapes. These factors contribute to the observed variability and wider peak widths in the HS-SPME method compared to the DI method.

The Whatman paper used as a solid matrix in this study was analyzed separately to confirm that it does not interfere with the sample peaks and does not introduce additional peaks to the chromatogram. The results can be seen in the Figure 2.5. The figure demonstrates that the paper background has some peaks at below 4 pA responses, which is significantly lower than the sample

peaks intensities. Because the paper blank contribution was so low, it was decided to not subtract the background from the sample chromatograms.

The results of PCA in Figure 2.6 are provided for the non-standardized dataset. Sample classes are clustering together but the clusters are not separated from each other. The DCS in this dataset is 0.8 (Table 2.5). Also, PC1 now has a lower contribution of 58.4% compared to 64.7% obtained with the DI measurement results. Figure 2.7 features a loadings plot that was generated to visualize the distribution of analytes in the principal components (PCs) of the HS-SPME data. As shown on the figure, all the peaks in PC1 have the same sign, indicating a positive correlation among them. Notably, all 11 analytes suitable for analysis were present in the loadings for PC1, including the internal standard peaks that ideally should not contribute in sample variation as their concentrations were kept consistent. This suggests that PC1 does not capture enough compositional variation between samples to correctly classify them.

Overall, this study aimed to determine how the presence of a substrate affects the relative concentration of a sample headspace, and whether it is effective to add internal standards to heterogeneous samples such as cacao beans investigated later in Chapter 3. The paper's weight, size, and chemical composition were kept consistent, as well as the sample volume, ensuring that the samples were quantitatively relatively close to identical. The main difference between samples was the geometry of the paper pieces, so when the sample was transferred to a sample vial, the sample's geometrical form differed each time. This might partially explain the increased variability in SPME data, as sample geometry leads to variability in a sample headspace spatial composition and sample-fiber distance, which, in turn, leads to differences in relative concentrations of sample analytes.

Table 2.3. List of compounds, their peak numbers, retention times, and boiling points.

Peak number	Compound	Retention time (s)	Boiling point (°F)
1	1-butanol	90.9	244
2	toluene	126	231

3	m-xylene	152	282
4	1,4-thioxane	155	297
5	mesitylene	170	328
6	1-decene	173	339
7	1,5-dichloropentane	179	360
8	5-decyne	180	351
9	1-octanol	183	383
10	undecane	187	385
11	1-bromooctane	192	394
12	2-decanone	197	412
13	methyl salicylate	198	428
14	methyl decanoate	208	435

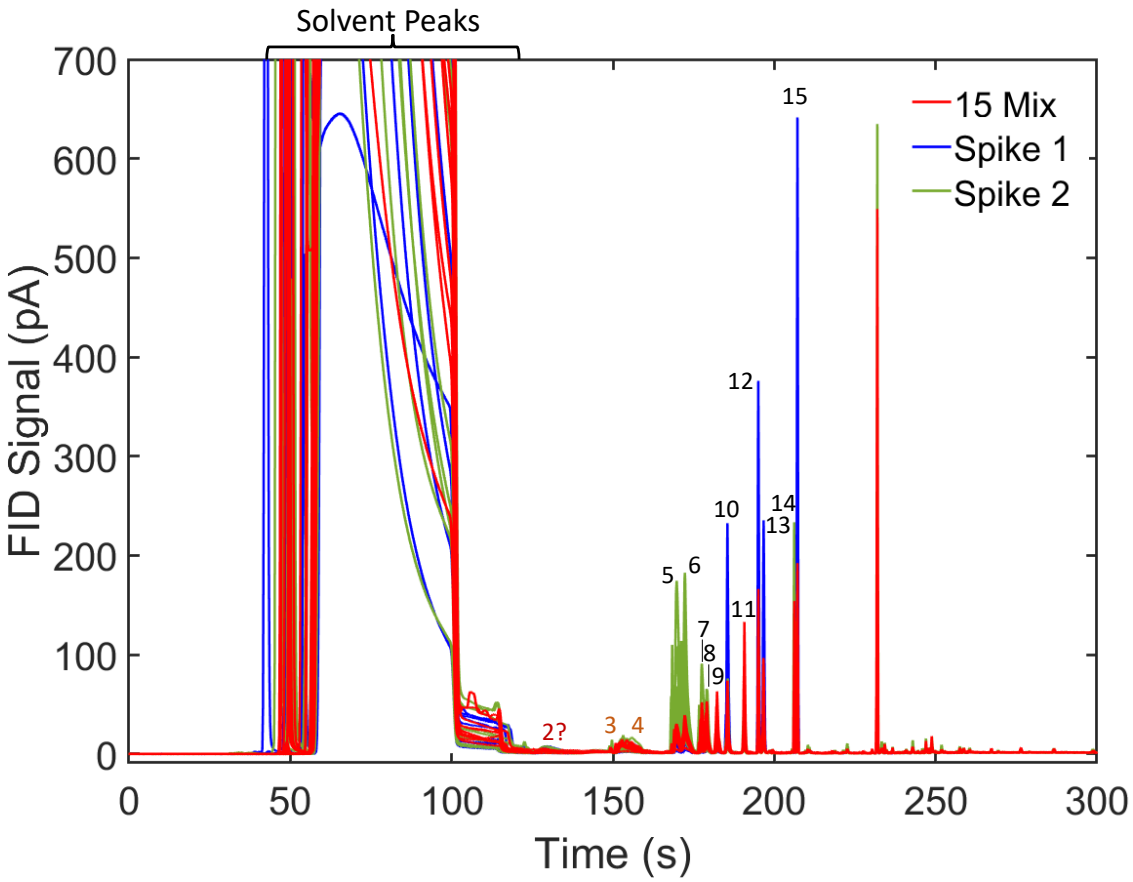

Figure 2.2. HS-SPME Full Chromatogram, (2) toluene, (3) m-xylene, (4) 1,4-thioxane, (5) mesitylene, (6) 1-decene, (7) 1,5-dichloropentane, (8) 5-decyne, (9) 1-octanol, (10) undecane, (11) 1-bromooctane, (12) 2-decanone, (13) methyl salicylate, (14) methyl decanoate, (15) bicyclohexyl.

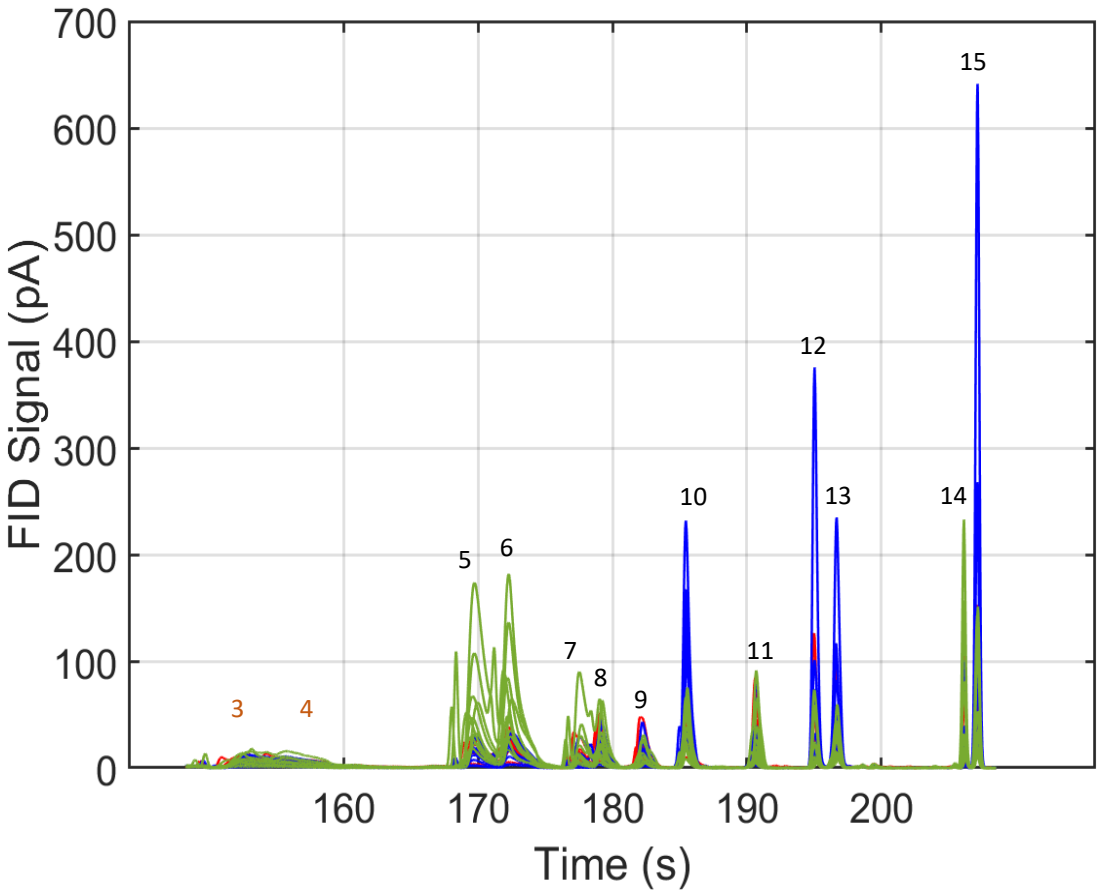

Figure 2.3. HS-SPME before normalization. Target Window. Peak numbers represent respectively (3) m-xylene, (4) 1,4-thioxane, (5) mesitylene, (6) 1-decene, (7) 1,5-dichloropentane, (8) 5-decyne, (9) 1-octanol, (10) undecane, (11) 1-bromooctane, (12) 2-decanone, (13) methyl salicylate, (14) methyl decanoate, (15) bicyclohexyl. Spike 1: (1) 1-butanol, (5) mesitylene, (6) 1-decene, (7) 1,5-dichloropentane, (14) methyl decanoate, Spike 2: (2) toluene, (10) undecane, (12) 2-decanone, (13) methyl salicylate, (15) bicyclohexyl, 15 Mix: (3) m-xylene, (4) 1,4-thioxane, (8) 5-decyne, (9) 1-octanol, (11) 1-bromooctane.

Table 2.4. Peak widths comparison between direct injection (DI) and HS-SPME

Peak number	Compound	Peak width DI (s)	Peak width SPME (s)
1	1-butanol	3.2	-
2	toluene	1.278	-
3	m-xylene	0.976	-
4	1,4-thioxane	1.071	-
5	mesitylene	1.073	3.759
6	1-decene	0.893	3.338
7	1,5-dichloropentane	0.863	2.271
8	5-decyne	0.846	1.976
9	1-octanol	1.093	2.197
10	undecane	0.906	1.876
11	1-bromooctane	1.022	1.775
12	2-decanone	1.292	1.371
13	methyl salicylate	1.255	1.534
14	methyl decanoate	1.099	0.906
15	bicyclohexyl	0.85	1.008

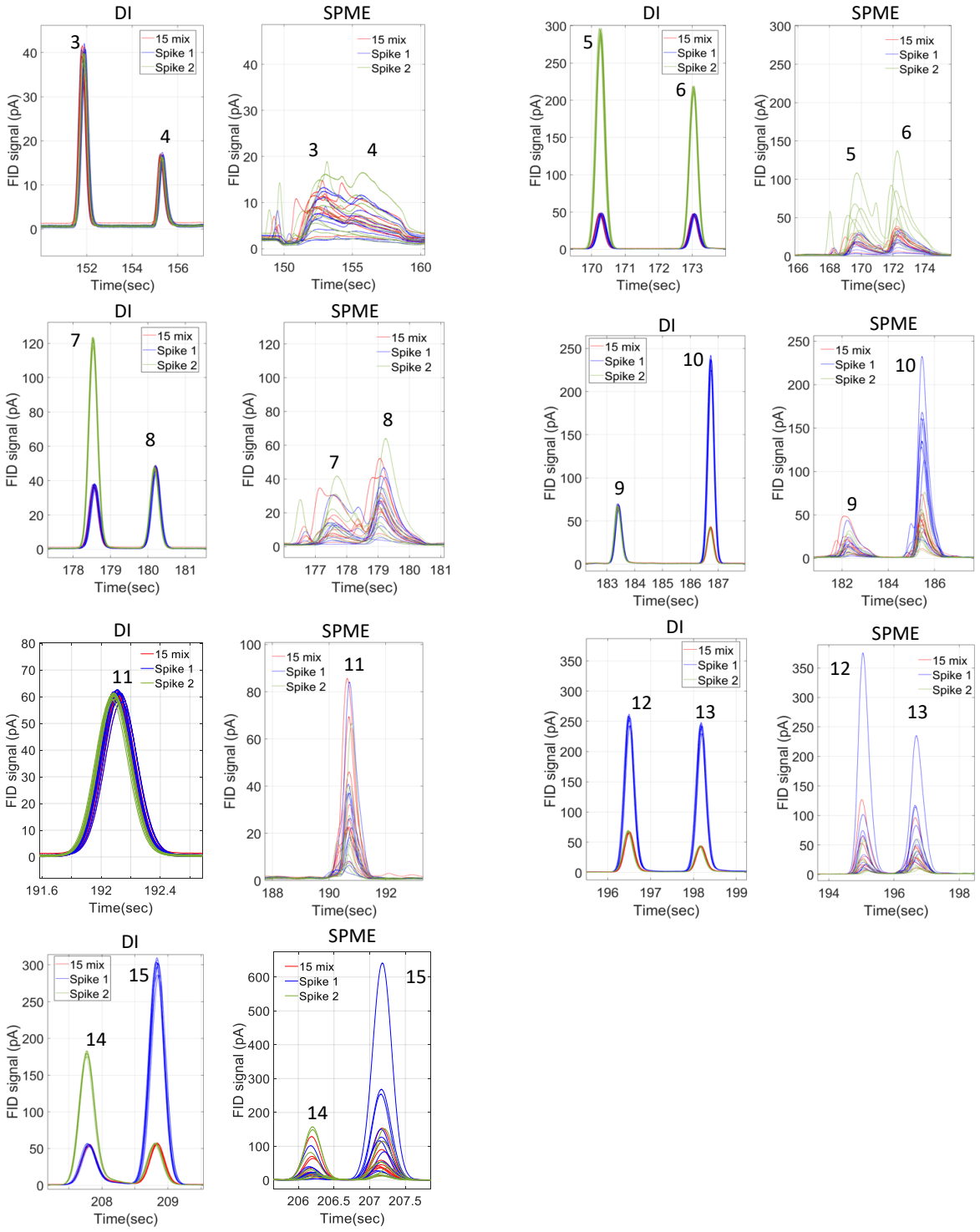

Figure 2.4. DI vs SPME peaks comparison. Peak numbers represent respectively (3) m-xylene, (4) 1,4-thioxane, (5) mesitylene, (6) 1-decene, (7) 1,5-dichloropentane, (8) 5-decyne, (9) 1-octanol, (10) undecane, (11) 1-bromooctane, (12) 2-decanone, (13) methyl salicylate, (14) methyl decanoate, (15) bicyclohexyl. Spike 1 – green, Spike 2 – blue, 15 Mix – red.

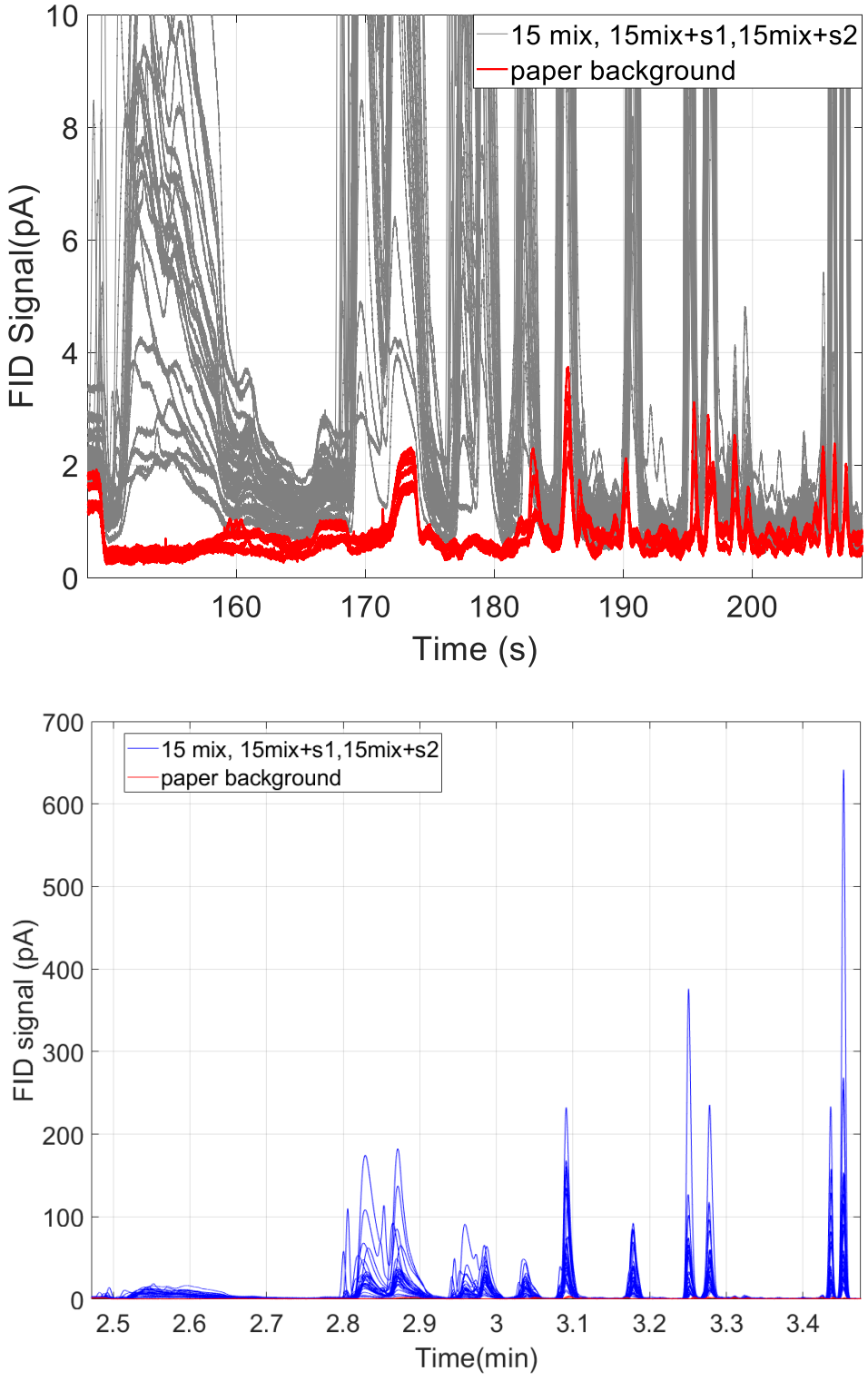

Figure 2.5. Paper background (red), full sample chromatographic overlay(gray).

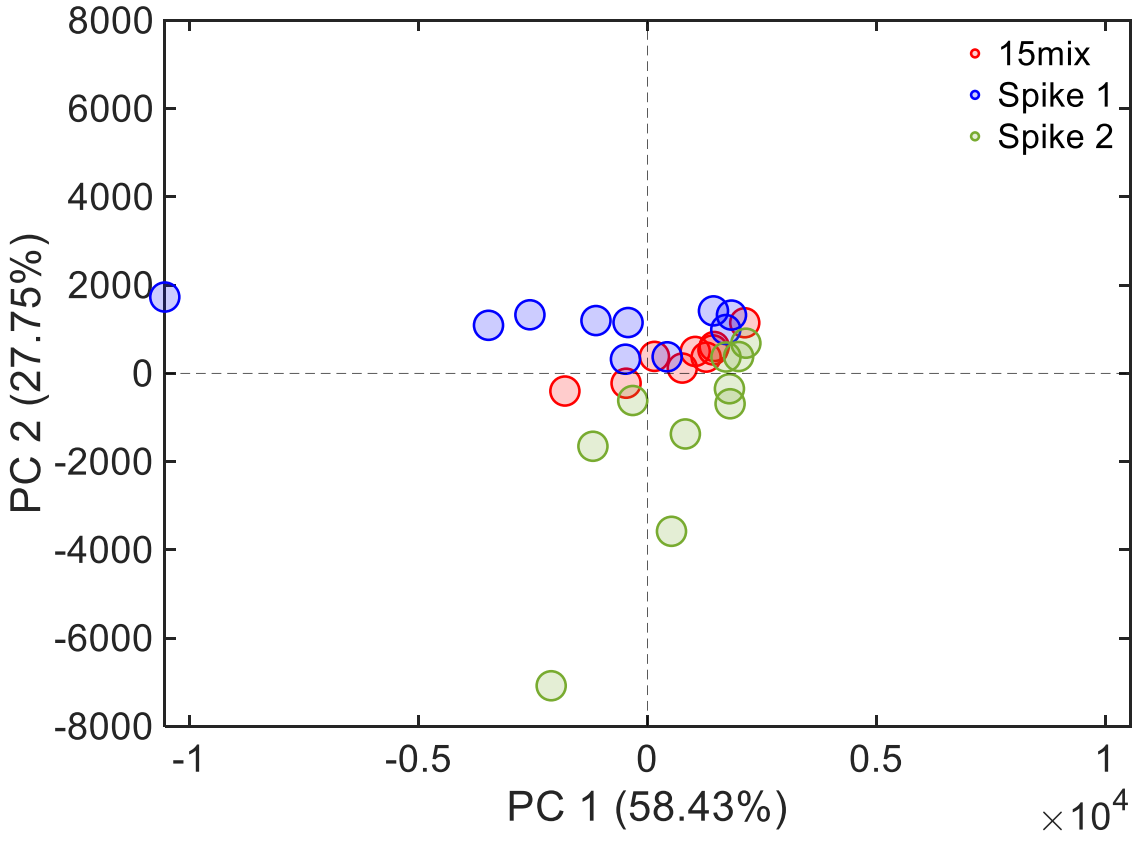

Figure 2.6. HS-SPME PC scores before normalization.

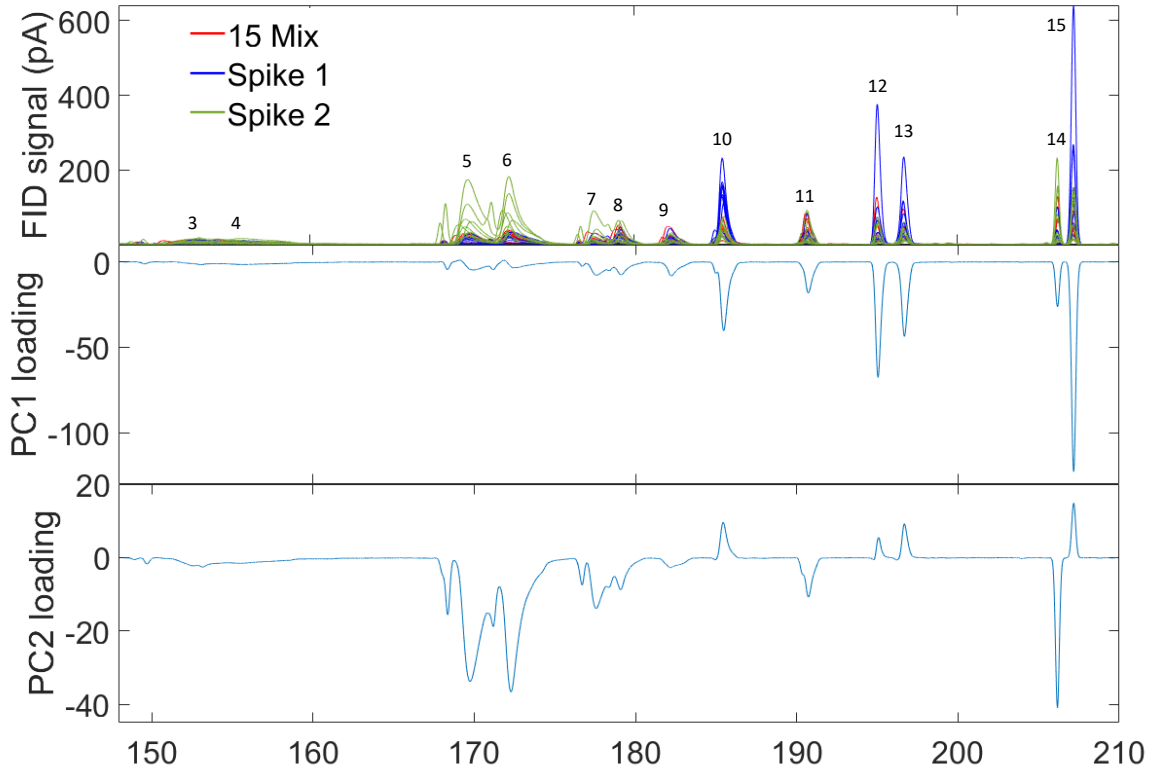

Figure 2.7. SPME PC loadings before normalization. Peak numbers represent respectively (3) m-xylene, (4) 1,4-thioxane, (5) mesitylene, (6) 1-decene, (7) 1,5-dichloropentane, (8) 5-decyne, (9) 1-octanol, (10) undecane, (11) 1-bromooctane, (12) 2-decanone, (13) methyl salicylate, (14) methyl decanoate, (15) bicyclohexyl. Spike 1: (1) 1-butanol, (5) mesitylene, (6) 1-decene, (7) 1,5-dichloropentane, (14) methyl decanoate, Spike 2: (2) toluene, (10) undecane, (12) 2-decanone, (13) methyl salicylate, (15) bicyclohexyl, 15 Mix: (3) m-xylene, (4) 1,4-thioxane, (8) 5-decyne, (9) 1-octanol, (11) 1-bromooctane.

2.3.3. *5-decyne normalization*

The objective of our study was to improve the quality and reproducibility of GC data using chemometric techniques. To achieve this, we employed a normalization method using multiple internal standard candidates, selected one at a time. The first internal standard investigated was 5-decyne. The 5-decyne peak (number 8) can be seen in Figures 2.3 and 2.4. As this peak elutes earlier in the chromatogram, it tends to be wider and displays significant variability in peak areas, shapes, and retention times between runs compared two internal standards evaluated in this study.

We normalized the peaks to an internal standard, resulting in a normalized chromatographic overlay of all 30 chromatograms (Figure 2.8). The PCA was performed on normalized dataset and the results are displayed in the Figure 2.8. b. The sample classes were more well-defined post-normalization, with 24/30 samples separated into the correct classes. However, three Spike 1 samples and three Spike 2 samples were not separated from the 15 Mix, while all 15 mixed samples were grouped into a single sample class. Furthermore, a qualitative improvement was demonstrated in the loadings as shown in Figure 2.8 c. Spike 1 analytes had the largest contribution in PC1, whereas Spike 2 analytes displayed overall lower values. It is also noteworthy that Spike 1 analytes mostly elute later so they get refocused on the column, which improves their chromatographic characteristics. Peak 14 (Spike 2) was almost absent, whereas peaks 5 and 6 appeared on the loadings plot with a negative sign but had lower values compared to Spike 1 peaks (Figure 2.8 c). Peak 8 was absent in the loadings plot as it was the peak to which the dataset was normalized. Peak 9 almost did not contribute to PC1, which was favorable since it was an unspiked peak and as such should have a consistent signal response across samples. Recall that this peak was present in the full overlay before normalization (Figure 2.7), which indicates that normalization improved the reproducibility and consistency of the data. In PC2, the most significant contributions were observed from peaks 5 and 6, which were Spike 2 peaks. The degree of peak separation (DCS) in the 5-decyne normalized dataset was found to be DCS = 2.1. These findings demonstrate that the normalization step improved the reproducibility and quality of data, leading to the separation of most replicas into corresponding sample classes.

The Figure 2.8 offers a magnified view of two specific peaks methyl decanoate and bicyclohexyl, peaks 14 and 15 respectively. Of these two, bicyclohexyl exhibits higher concentration in 7/10 Spike 1 samples, while methyl decanoate exhibits a higher concentration only in 2/10 Spike 2 samples as compared to non-spiked ones. Conversely, while methyl decanoate has a similar concentration and retention time to the bicyclohexyl, its reproducibility and consistency are comparatively lower.

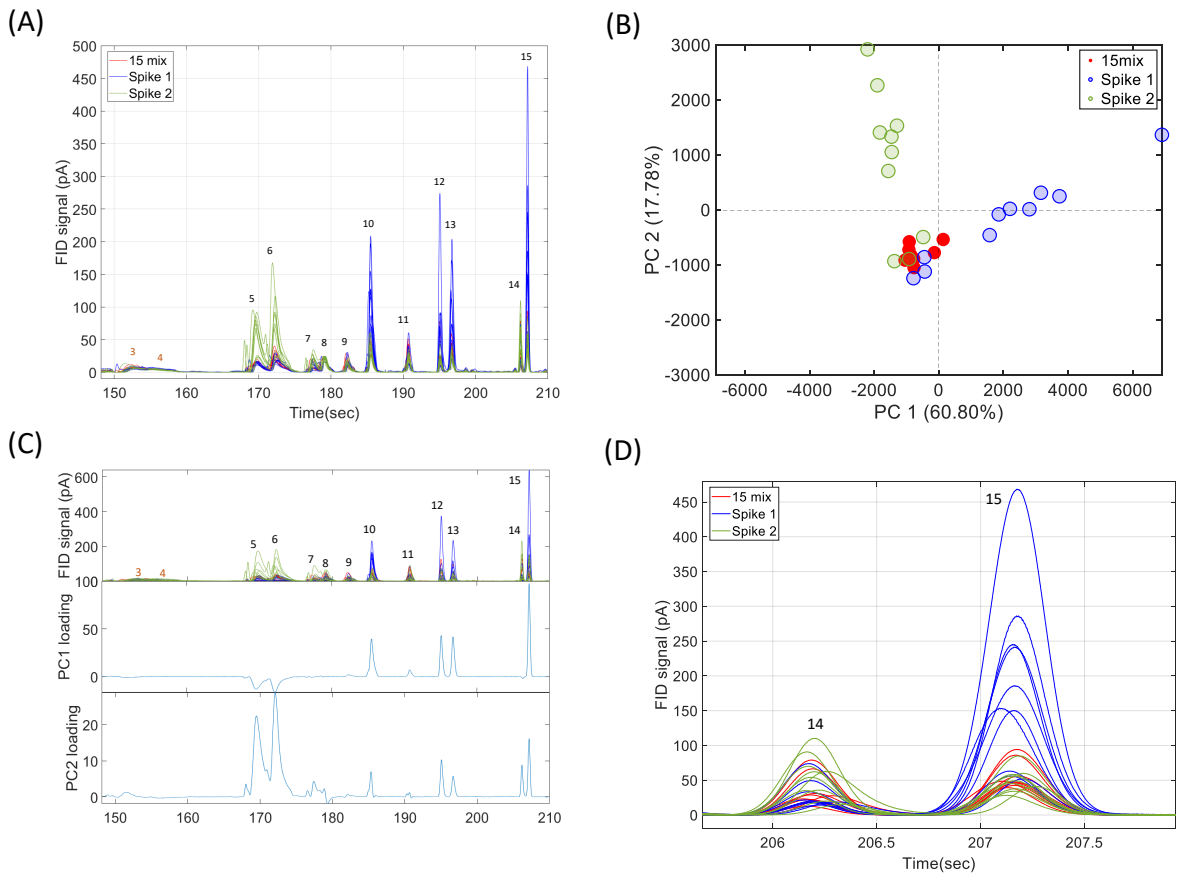

Figure 2.8. 5-Decyne normalization (A) 5-Decyne normalized chromatogram. Peak numbers represent respectively (3) m-xylene, (4) 1,4-thioxane, (5) mesitylene, (6) 1-decene, (7) 1,5-dichloropentane, (8) 5-decyne, (9) 1-octanol, (10) undecane, (11) 1-bromooctane, (12) 2-decanone, (13) methyl salicylate, (14) methyl decanoate, (15) bicyclohexyl. Spike 1 – green, Spike 2 – blue, 15 Mix – red. (B) 5-Decyne Norm. PC Scores Plot. DCS = 2.1. (C) 5-Decyne Normalized Loadings. (D) 5-Decyne Normalized methyl decanoate and bicyclohexyl peaks. Peak numbers represent respectively (14) methyl decanoate, (15) bicyclohexyl. Spike 1 – green, Spike 2 – blue, 15 Mix – red.

2.3.4. *Octanol Normalization*

The results of the 1-octanol normalization are presented in Figure 2.9. The reproducibility as seen in the reduced in-class variation, and the increased between-class variation are evident in the 15 Mix peaks (5-decyne, 1-bromoctane, and 1-octanol), which appear to be more consistent from peak to peak. Additionally, peaks 3 and 4 (m-xylene and 1,4-thioxane, respectively) are defined more distinctly and have higher relative intensity.

The results of PCA are presented through scores and loadings plots in Figure 2.9, where an evident improvement in class separation compared to non-normalized data is observed. All three sample classes are defined and visually present on the figure. 25 out of 30 samples were separated into the correct classes compared to 24/30 samples in 5-decyne normalization. Two samples in Spike 2 and three samples in Spike 1 groups cluster with 15 Mix chromatograms. The DCS was calculated as 2.2 indicating an increase of consistency in the data compared to 0.8 before normalization (Table 2.5).

Figure 2.9 C shows the loading for PCA in data normalization to 1-octanol. The loadings in this figure are similar to those in direct injection measurements. PC1 separates analytes into three well-defined groups. Spike 1 peaks have a negative sign and larger contributions to the loadings, whereas Spike 2 peaks have a positive sign and are less well-defined peaks. 15 Mix peaks have almost no contribution to PC1 loadings. PC2 loadings appear similar to the original 1-octanol normalized chromatogram, indicating that the variation captured on this PC is distributed consistently throughout the whole chromatogram, with slightly higher variation for peaks 5 and 6.

Normalization using 1-octanol as the internal standard proved to be effective in improving data quality to be more comparable to direct injection measurements in GC-MS analysis. The 1-octanol normalization resulted in improved data consistency, reduced in-class variation, and increased between-class variation. Furthermore, the PCA results indicated an evident improvement in class separation compared to non-normalized data.

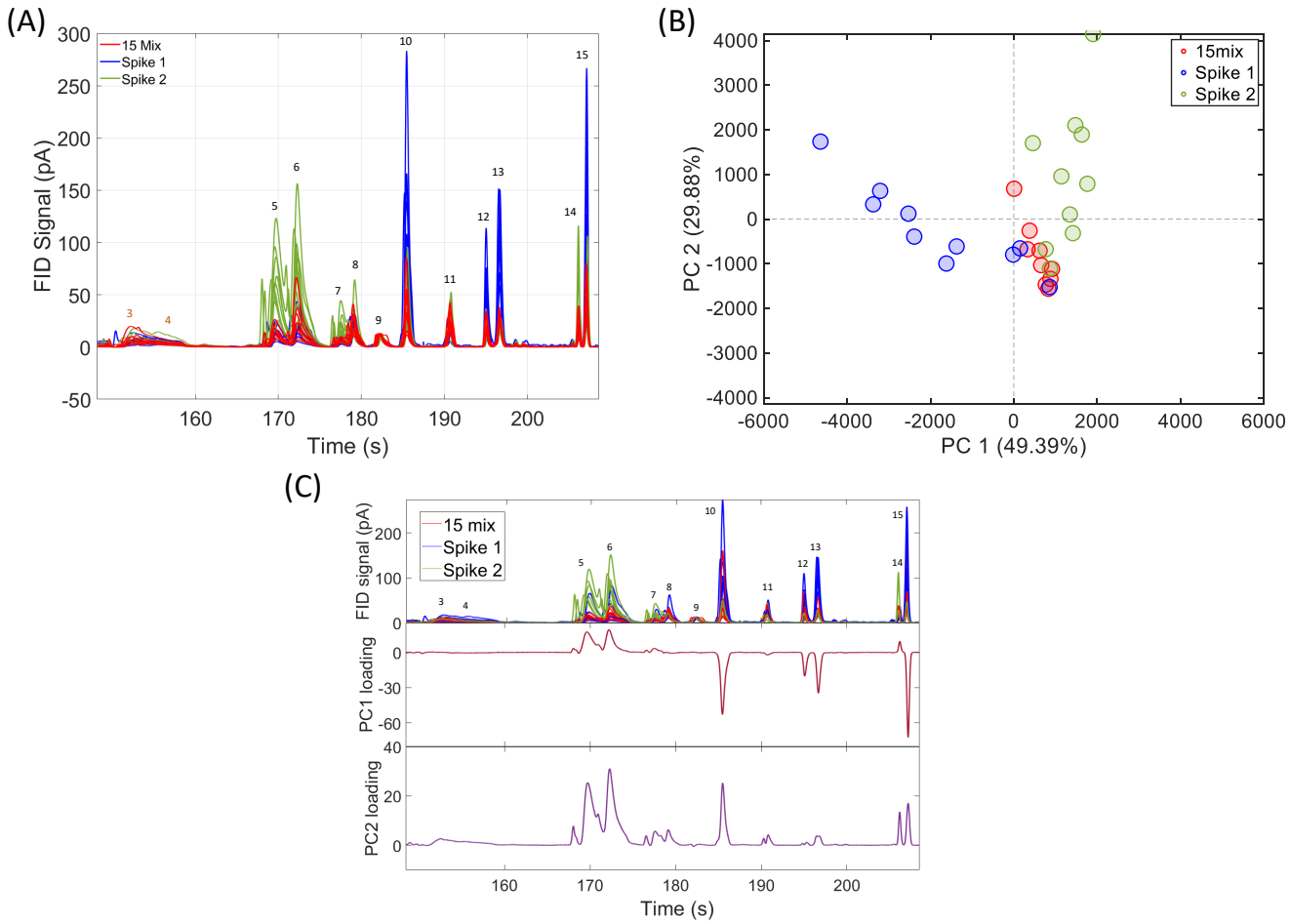

Figure 2.9. 1-Octanol Normalization. (A) 1-Octanol Normalized Chromatograms. (1) 1-butanol, (2) toluene, (3) m-xylene, (4) 1,4-thioxane, (5) mesitylene, (6) 1-decene, (7) 1,5-dichloropentane, (8) 5-decyne, (9) 1-octanol, (10) undecane, (11) 1-bromooctane, (12) 2-decanone, (13) methyl salicylate, (14) methyl decanoate, (15) bicyclohexyl. (B) 1-octanol normalized PC scores plot. (C) 1-octanol normalized loadings.

2.3.5. *1-bromooctane normalization*

Visual analysis of the chromatograms revealed that normalization to 1-bromooctane improved the consistency of internal standard peaks and reduced unwanted in-class variations, resulting in a more reasonable width-to-height ratio in the chromatogram, as seen in Figure 2.10. The chromatographic overlay of normalized data appeared neater compared to the pre-normalization data and closer to direct injection measurements than two other normalization attempts.

PCA was conducted to assess the effectiveness of 1-bromooctane normalization, as shown in Figure 2.10. The results showed that normalization to 1-bromooctane yielded the highest value of DCS = 4.7 and the separation of 27 out of 29 samples. Loadings analysis indicated that three sample classes were well-defined, with Spike 1 and Spike 2 peaks having positive and negative values, respectively, and 15 Mix peaks having almost no value.

Finally, a comparison between peaks before and after normalization to 1-bromooctane was presented (Figure 2.11). Peak 11, which is a 1-bromooctane peak used for normalization, assumed had an identical peak height in all 30 samples as a result of the normalization step. The results showed that normalization to 1-bromooctane yielded the most successful outcomes, with improved data quality, reduced in-class variation, and well-defined class separation.

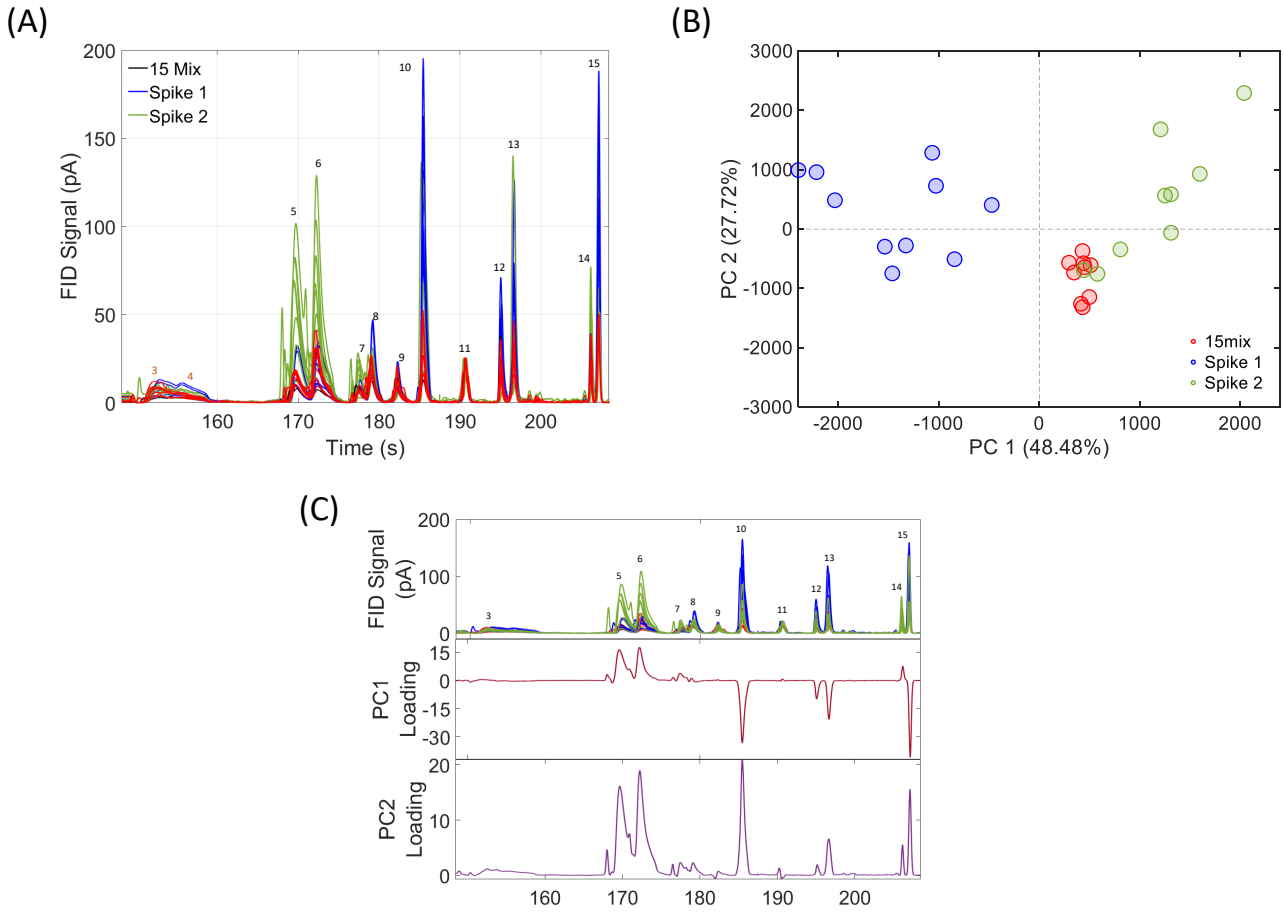

Figure 2.10. Normalization to 1-bromooctane chromatogram. (A) Chromatographic window. (1) 1-butanol, (2) toluene, (3) m-xylene, (4) 1,4-thioxane, (5) mesitylene, (6) 1-decene, (7) 1,5-dichloropentane, (8) 5-decyne, (9) 1-octanol, (10) undecane, (11) 1-bromooctane, (12) 2-decanone, (13) methyl salicylate, (14) methyl decanoate, (15) bicyclohexyl. Spike 1 – green, Spike 2 – blue, 15 Mix – red. (B) 1-Bromooctane normalized PC scores. (C) 1-Bromoctane normalized loadings.

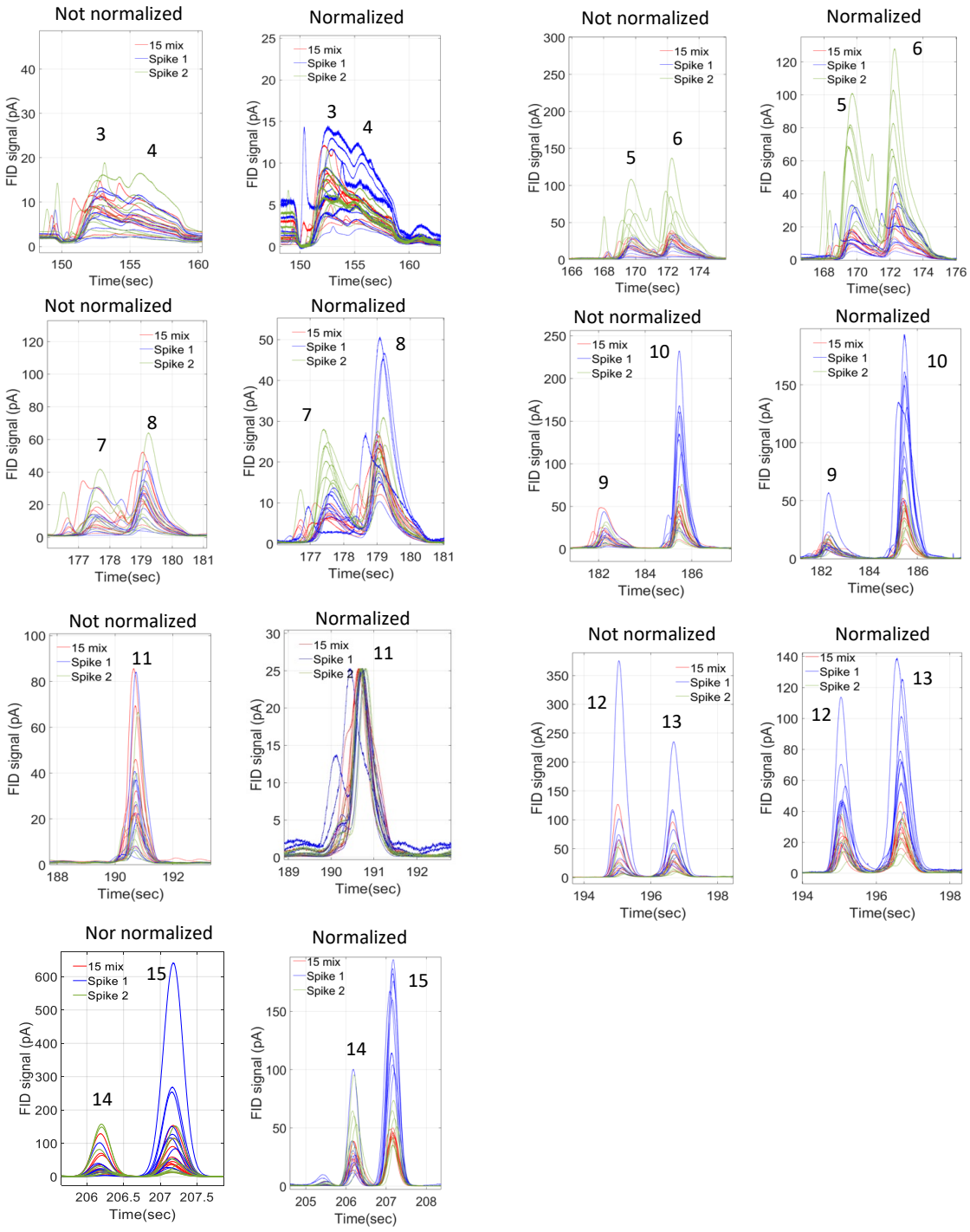

Figure 2.11. Peaks before and after normalization with 1-bromoctane. Peak numbers represent respectively (3) m-xylene, (4) 1,4-thioxane, (5) mesitylene, (6) 1-decene, (7) 1,5-dichloropentane, (8) 5-decyne, (9) 1-octanol, (10) undecane, (11) 1-bromooctane, (12) 2-decanone, (13) methyl salicylate, (14) methyl decanoate, (15) bicyclohexyl. Spike 1 – green, Spike 2 – blue, 15 Mix – red.

2.4. Conclusions And Future Prospectus

In conclusion, the study demonstrates that use of normalization to an internal standard can significantly improve the accuracy of samples classification. The choice of internal standard is important for the success of the normalization. Three internal standards were investigated and compared in this study. As a numerical measure of evaluation of their performance, the Degree of Class Separation metric was used. The calculated DCS values are presented in the Table 2.5. The study suggests that for our HS-SPME dataset later eluting compounds tend to make better internal standards due to a refocusing effect. Normalization to 1-bromooctane was found to be the most successful normalization approach, effectively reducing unwanted in-class variations and improving the consistency of internal standard peaks. However, it should be noted that some analytes (particularly early-eluting analytes) may suffer from broadening during the SPME process, and this may not be easily resolved by method adjustment. Overall, this research provides valuable insights for improving the accuracy and reliability of HS-SPME analysis, which can have important applications in various fields, including food science and environmental monitoring.

One potential future prospect for continued study could be analyzing and comparing samples of identical composition with and without the solid matrix component using HS-SPME to further investigate the influence of this matrix effect on the accuracy of HS-SPME data. This could involve investigating how different matrices affect the analysis of different compounds and how to optimize sampling methods for each matrix. Additionally, the study could explore the effect of different sample preparation techniques on the resulting data. The goal of this study could be to develop a standardized approach to sample analysis that can be applied to a wide range of matrices and compounds. This could have significant implications for fields such as environmental monitoring, food and beverage quality control, and forensic analysis.

Table 2.5. Calculated DCS values.

	HS-SPME pre-normalization	1-octanol normalization	1-bromooctane normalization	5-decyne normalization	direct injection
DCS	0.8	2.2	4.7	2.1	67.4

Chapter 3. TEMPERATURE PROGRAM OPTIMIZATION FOR CLASSIFICATION OF MOISTURE-DAMAGED AND INTACT CACAO USING INTUVO GC SYSTEM

3.1. INTRODUCTION

Cacao, a globally renowned crop, serves as a vital source of livelihood for many farmers and is greatly cherished for its sensory attributes. However, a significant concern for cacao production and supply chain management is the detrimental impact of moisture damage on the quality and safety of cacao beans. The ubiquitous nature of water in the environment, coupled with inadequate harvest and storage precautions, poses a significant challenge to the preservation of cacao quality. Moisture damage is an umbrella term encompassing a range of adverse reactions in cacao due to prolonged exposure to moisture. The consequences of moisture damage are multifaceted, including a decrease in cacao quality, alteration in flavor, and diminished food safety. The progression of moisture damage involves several intricate pathways, culminating in chemical changes in cacao. These chemical changes can result in mold growth, disruption of the fermentation process, lipid oxidation, and sugar hydrolysis, all of which affect the quality and flavor of chocolate produced from the beans (Table 3.1).

Table 3.1. Moisture damage in cacao.

Type of moisture damage	Cause	Effect	Chemical change
Mold growth	The presence of moisture can encourage the growth of mold	Musty and stale smell, compromised safety of the cacao products, unpleasant earthy flavor [19]	Mold can produce toxins, such as mycotoxins [19]
Fermentation	During the fermentation process, microorganisms break down the pulp surrounding the cacao beans. If the beans are exposed to	Can negatively affect the flavor, aroma and safety of the cacao products	Leads to the production of peptides, amino acids and alcohols [20]

	too much moisture, the fermentation process can be affected and altered.		
Lipid oxidation	Cacao beans contain lipids, such as cocoa butter, which can undergo oxidation when exposed to moisture	Rancid flavor	Hydrolysis, oxidation of triglycerides [21]
Sugar hydrolysis	Moisture can lead to the hydrolysis of sugars in cacao beans	Can negatively affect the flavor and aroma of the cacao products	Hydrolysis of sugars

The classification of moisture-damaged cacao beans is a crucial process in the cacao industry, as excess moisture can significantly impact the quality, safety, and flavor of chocolate products made from the beans. Therefore, it is important to identify and classify moisture-damaged cacao beans to ensure that only high-quality beans are used in the production of cacao products. The classification process used in industry for moisture-damaged cacao beans typically involves a visual inspection of the beans to identify any signs of moisture damage, such as discoloration, mold growth, or a musty odor. In addition to visual inspection, other methods such as moisture content analysis, microbial analysis, and chemical analysis of volatile compounds can also be used to detect moisture damage and its effects on the quality of the beans. Once the moisture-damaged beans have been identified, they are typically graded based on the severity and extent of damage. A grading system is typically used to assign a score to each bean, with higher scores indicating less moisture damage and better quality. The grading system may vary depending on the region or country of origin but typically involves assigning a numerical score or letter grade to each bean.

After grading, the moisture-damaged beans are sorted into different categories based on their score or grade. Beans with the highest scores or grades are typically sold at a premium, while beans with lower scores or grades may be used for lower-quality chocolate products or other applications. Quality control measures are put in place to ensure that the graded cacao beans meet certain standards for moisture content, flavor, and other quality factors. Samples may be taken

from each category of graded beans and tested for their quality factors. The utilization of gas chromatography for the classification of moisture-damaged cacao beans caused a significant advancement in the cacao industry.[22] GC is a widely used technique for analyzing volatile compounds in food products, and it can be utilized to identify specific volatile compounds produced as a result of moisture damage in cacao beans.

The GC methodology is based on the principles of separating, identifying, and quantifying individual components in a complex mixture. The volatile compounds in the sample headspace (HS) can be extracted using a solid phase microextraction (HS-SPME) sampling technique. The extracted compounds are then separated using a GC column with a stationary phase through which a gas mobile phase carries the compounds. The separated compounds are detected by a detector at the end of the column, which produces a signal that is proportional to the amount of each compound present in the sample. Finally, the data from the detector is analyzed to identify and quantify the specific volatile compounds present in the sample.[23–30] GC provides more accurate and quantitative information on the quality of cacao beans compared to visual inspection or other methods. It can detect and quantify specific volatile compounds produced as a result of moisture damage, such as aldehydes, ketones, and alcohols. This information allows for a more precise classification of moisture-damaged cacao beans based on the severity and extent of damage. Moreover, GC is reliable, efficient, and cost-effective, making it an indispensable tool for quality control in the cacao industry. Temperature program optimization is an important step using HS-SPME-GC analysis to ensure that the target analytes are separated from interfering compounds in the sample. Temperature program optimization involves adjusting the temperature settings of the GC oven during the SPME analysis to maximize the chromatographic resolution and peak separation of the target analytes.

The study by Elizabeth Humston et al. in 2010 investigated how moisture damage affects the chemical signature of cacao beans over time. The authors used solid-phase micro-extraction and comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry to detect and analyze changes in the headspace vapor of cacao beans. The study identified specific analytes that change in concentration due to moisture damage, which could be quantified using the F-Ratio and PARAFAC algorithms. Prediction algorithms could be used to detect moisture damage before visible signs of mold appear by analyzing subsets of the analytes.

This study has several goals related to the classification of molded cacao beans. The main objective was to develop a cost-effective and efficient method for classifying molded cacao beans based on their volatile organic compound (VOC) profiles. To this end, the HS-SPME technique was utilized to extract the VOCs from both moisture-damaged and intact cacao samples. GC analysis was then performed using varying temperature programs to optimize the separation and detection of the VOCs. The resulting data were subjected to statistical analysis to identify key features that differentiate between moisture-damaged and intact cacao samples. The impact of changes in temperature program parameters on the classification of cacao samples was also investigated. Specifically, the study sought to identify the limit of the temperature increase rate beyond which accurate classification of moisture-damaged cacao samples is not possible. This information would be valuable for developing practical and efficient classification protocols that are both sensitive and specific.

Overall, the study sought to advance the understanding of the relationship between VOC profiles and the moisture content of cacao samples. By identifying key features that differentiate between moisture-damaged and intact cacao samples with an easy-to-use instrument with fast run times, the study would contribute to the development of improved quality control protocols in the cacao industry.

3.2. EXPERIMENTAL

3.2.1. *Sample preparation*

The samples in question consisted of cacao nibs from Ivory Coast, either molded or intact. The molded cacao was prepared by introducing water to the nibs in a 1:2 ratio by mass and sealing it for ten days to allow for full mold coverage. The resulting molded and unmolded cacao was then subjected to crushing to enhance the release of volatile organic compounds (VOCs) before SPME extraction. The HS- SPME extraction procedure entailed mixing 2 g of crushed cacao nibs with 3 mL of 40% w/v NaCl. Over-concentrated NaCl solution was used to increase release of VOC from the sample.[25,31]

3.2.2. *Headspace Solid-Phase Microextraction (HS-SPME)*

During method optimization multiple HS-SPME extraction procedures from a variety of recent studies involving extraction of cacao volatiles were considered.[23,25–28,30,32–34] The selected

HS-SPME method follows the procedure described in by Barbosa-Pereira et al. with minor variations.[31] For preconcentration and extraction of analytes, a 50/30 μm DVB/CAR/PMDS (Supelco, USA) HS-SPME fiber was used. Before extraction fiber was conditioned at 270 °C in the inlet of the GC for 30 min. During conditioning, flow in the instrument was held at 1 mL/min, the oven was heated to 250 °C and the split vent was open to prevent residual compounds coming from the fiber and contaminating the GC column. Each sample vial was placed in a bead bath heated to 60 °C for 30 min to ensure the system reached equilibrium prior to extraction. Extraction was performed for an additional 30 min under a temperature of 60 °C. After each extraction-desorption cycle, the fiber was cleaned via additional desorption in the GC inlet for 5 min at 250 °C after each run.

3.2.3. *Gas-chromatography*

All data was collected using Intuvo 9000 GC system with a flame ionization detector (FID). The HP-5MS (5%-diphenyl-methylpolysiloxane) nonpolar column with dimensions 30 m length × 250 μm inside diameter × 0.25 μm film was used. Multiple temperature programs were investigated for the GC analysis, with the parameters optimized to achieve the highest degree of separation and detection of the VOCs present in the cacao samples. The chosen program was initiated with a start temperature of 40 ℃, with a hold time of 6 min. Splitless injection was used during desorption. Hydrogen was used as a carrier gas. The temperature was then ramped up to 300 ℃ at varying rates of 50, 75, 100, 200, and 250°C/min (Table 3.2), before being held at the final temperature of 300 ℃ for an additional minute. The FID detector was set at 1 kHz. The OpenLab CDS (Agilent Technologies, Palo Alto, CA, USA) software was employed for data collection, data procession, instrumental control and method adjustments.

Table 3.2. Investigated temperature ramps and their total run times.

Temperature increase rate (°C/min)	50	75	100	200	250
Time (s)	696	592	540	462	446

3.2.4. *Data analysis*

Data analysis was performed using MATLAB R2022b (The Mathworks, Inc., Natick, MA, USA). A rolling ball minimum algorithm was employed to perform baseline correction on the chromatograms. The chromatograms were aligned using the COW function to eliminate any run-to-run retention time shifting. Class distinguishing clustering was done using PCA.

3.3. RESULTS AND DISCUSSION

In this study, 5 temperature programs were investigated, two sample classes (intact/unmolded and moisture damaged) and 4 replicas of each class were collected. In total there were 40 chromatograms collected and analyzed.

3.3.1. *Temperature increase rate 50 °C/min*

The initial temperature program started at 40°C which was held for 6 min before subsequently increasing the temperature to 300°C at a rate of 50°C/min. This temperature program was the slowest of the 5 studied programs. Figure 3.1 illustrates a full chromatographic overlay of 8 chromatograms collected under the 50°C/min ramp rate. The red graphs correspond to moisture-damaged cacao chromatograms, while the blue graphs correspond to intact cacao data. The chromatogram appears to have two distinct windows, a smaller one at the beginning of the chromatogram (45-90 s) and a larger one later on the graph at around 550-650 s.

A zoomed-in view of the larger chromatographic window covering later eluting peaks is presented in Figure 3.2 (a). Visually, there were no apparent differences between the molded and unmolded cacao samples, except for certain peaks that exhibited high class-distinguishing qualities. The peaks in this region were well-defined and narrow. Figure 3.2 (b) depicts a zoomed-in view of the tallest distinguishing peak found, with a maximum signal of around 200 pA. Figure 3.2C shows a second class-distinguishing peak. Although these compounds have not yet been identified, they are likely to be byproducts of chemical changes that occur in cacao with moisture damage, such as mold growth, oxidation or microbial activity. Figure 3.2D highlights a smaller chromatographic window in the earlier eluting area that contains numerous analyte peaks in moisture-damaged samples and only a few peaks in intact samples. As these analytes elute early when the column is held at the lower temperature of 40°C, they are quite likely to have very low

boiling points and be highly volatile. Some of these analytes strongly appear to be specific to moisture damaged samples. It is difficult to determine whether the peaks present in both damaged and intact samples correspond to the same compounds or to different compounds with the same retention times. Nevertheless, this region is useful for class-distinguishing analyses due to the distinct differences in chromatographic profiles between intact and damaged samples.

The presented results illustrate the application of PCA for the classification of intact and moisture-damaged beans. Figure 3.3 (b) demonstrates results of the PCA performed on the baseline-corrected data before alignment. PC1 captured 52.7% of the variance and it separates three of the four damaged samples from the intact ones. Alignment of the dataset was determined to be necessary to address the issue of retention time shifting in the moisture-damaged sample, so COW alignment was used to resolve this. The results showed that the aligned dataset was better at separating intact and damaged samples than the unaligned dataset. It is important to note that the moisture damage in cacao was not uniformly distributed. Therefore, rather than having two distinct sample groups, the dataset was more of a spectrum ranging from no moisture damage to severe moisture damage.

The loadings plot shown on the Figure 3.3A indicates that PC1 is almost exclusively constructed of two class-distinguishing peaks in a later eluting area and two class-distinguishing peaks in the early eluting area. Class distinguishing peak 1 had the largest contribution to PC1. This means that PC1 captures the class variation in moisture-damaged and intact beans. On the other hand, PC2 has the largest contribution from peaks that are, on average, more intense in intact samples. However, due to significant variation in the areas of peaks having the largest contribution on PC2, it does not provide reliable and unambiguous results for class separation.

The early eluting area of the chromatogram (40-90 s) was identified as the section with the most useful class-distinguishing information, and PCA was performed on a dataset formed from taking a window of only this section. The results showed significantly improved class separation, with all intact samples grouped without overlap with moisture-damaged samples. The moisture-damaged samples were more spread out, which can be attributed to the uneven occurrence and varied nature of moisture damage throughout the sample volume. The variability of mold growth, which is one of the most significant types of moisture damage, results in a variety of molds at

different stages of development producing varying range of VOCs.[19] However, regardless of in-class differences between molded samples, all of them were separated from intact samples.

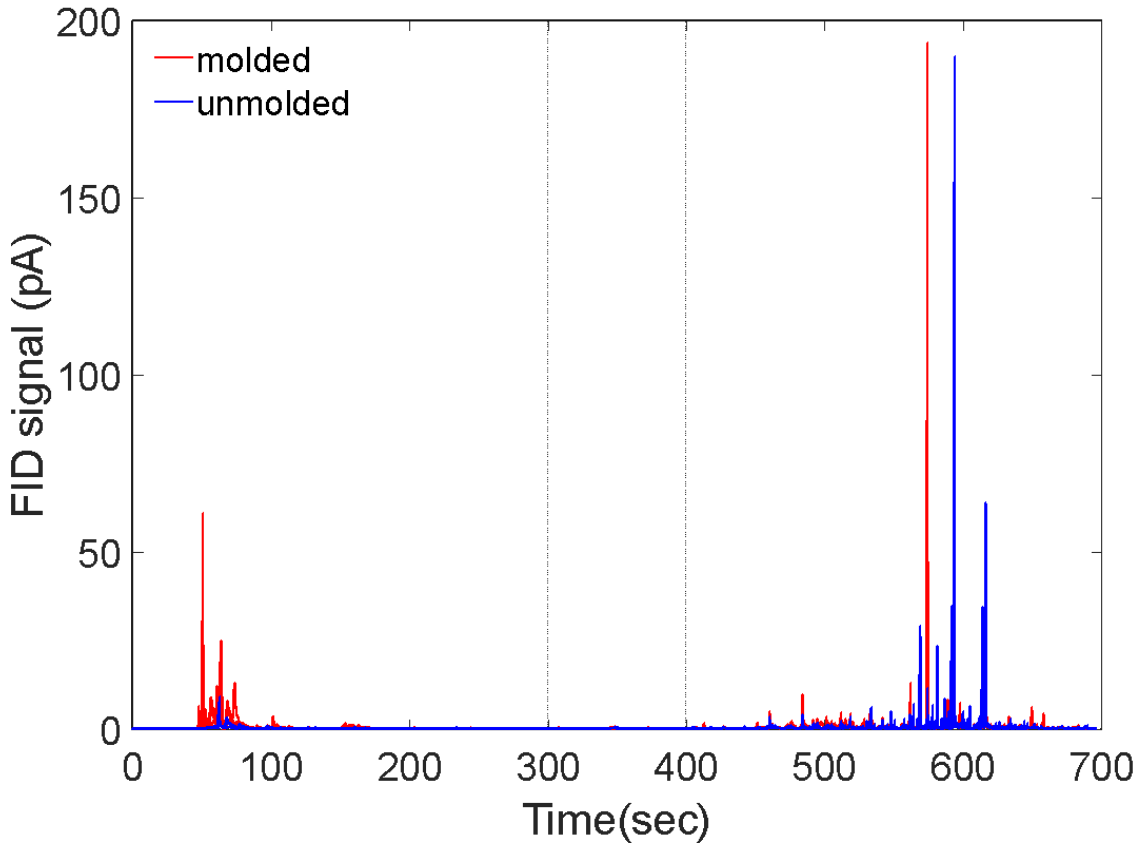

Figure 3.1. Full chromatographic overlay of the data obtained with a temperature ramp of 50 °C/min. Blue graphs correspond to the chromatograms of intact crushed cacao nibs that were kept according to the standards. The red graphs are moisture-damaged cacao.

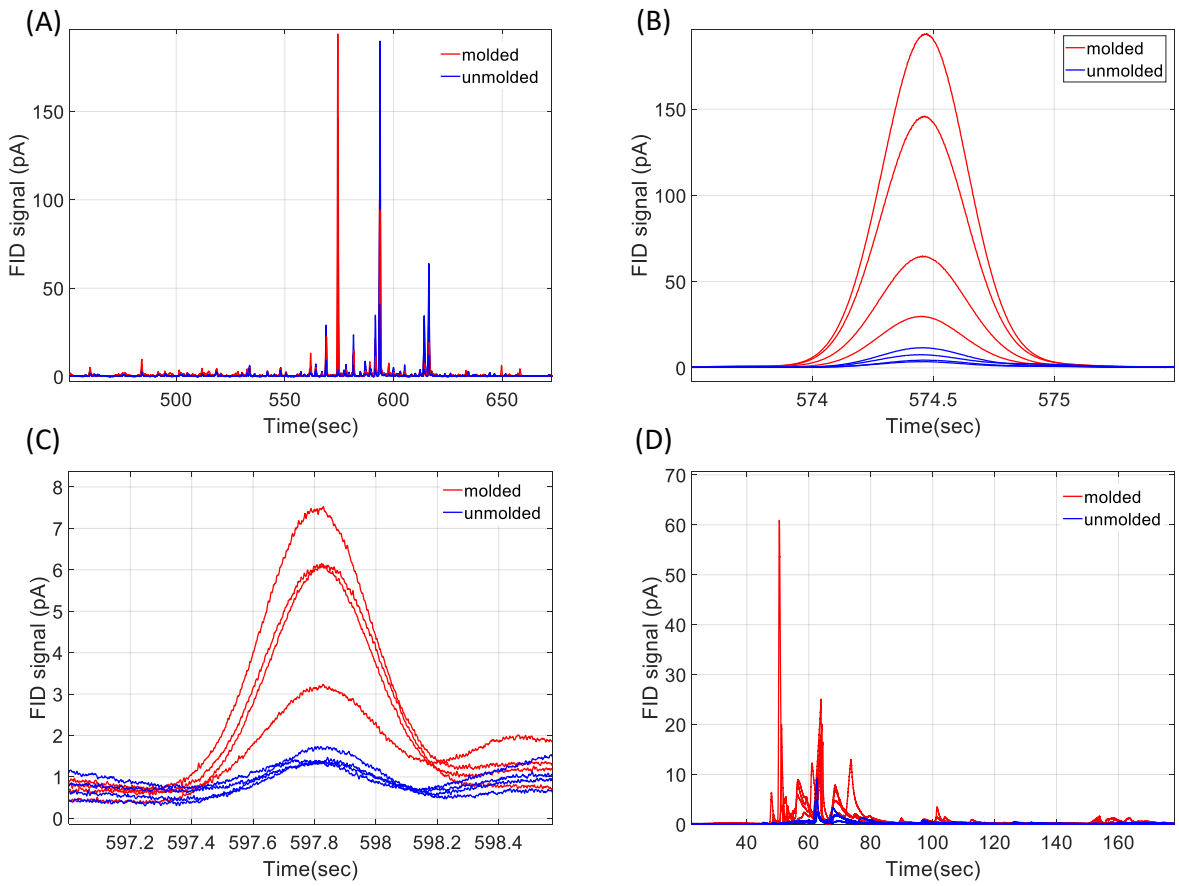

Figure 3.2. Temperature ramp of 50 °C/min. The blue chromatograms correspond to intact cacao. The red graphs correspond to moisture damaged cacao. (A) Chromatographic window: zoom-in for later eluting analytes. (B) Zoom-in for class distinguishing peak #1 (C) Zoom-in for class distinguishing peak #2. (D) Zoom in on the early eluting area. Intact cacao only has one major peak in this area, while moisture-damaged cacao has a lot of analytes eluting in the first 170 s, especially in the first 90 s.

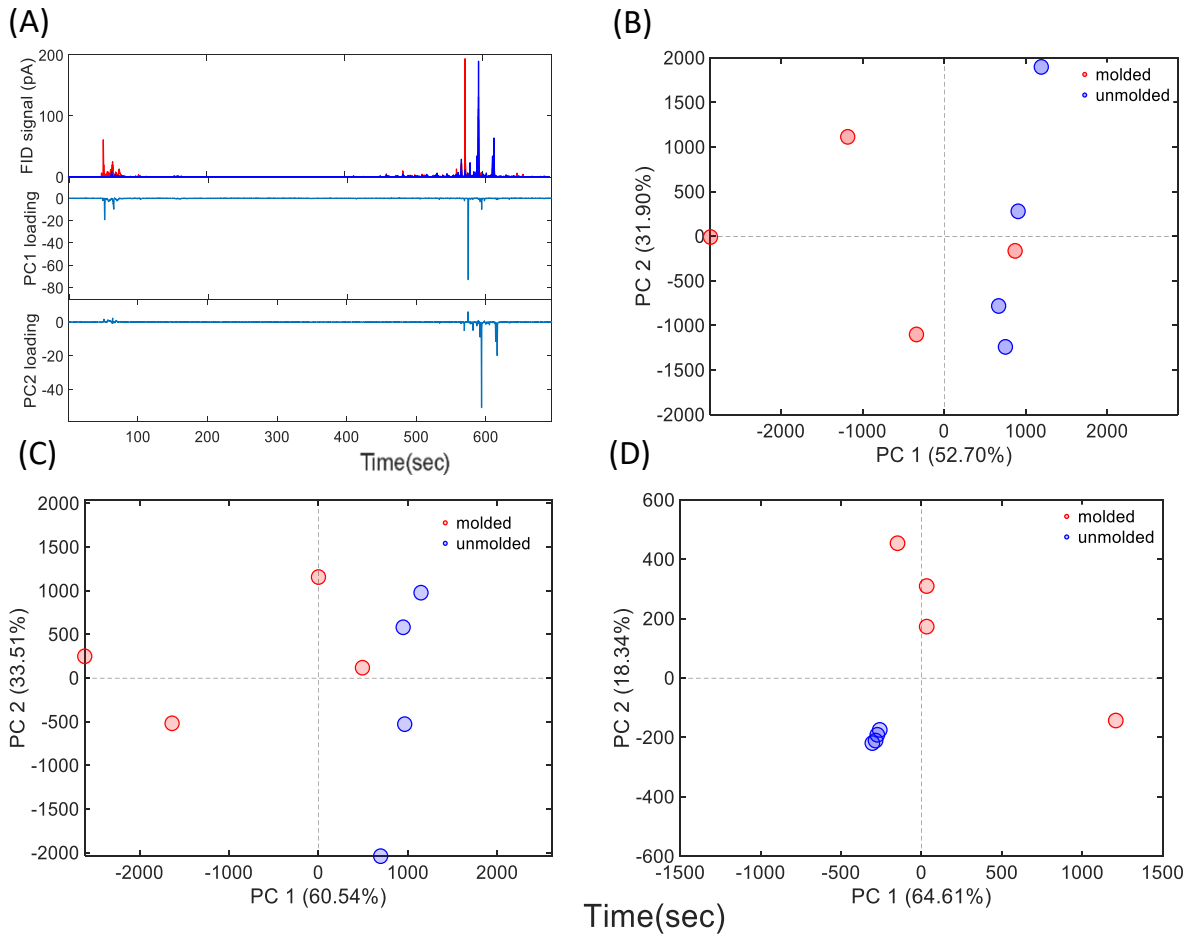

Figure 3.3. Temperature ramp of 50 °C/min. The blue markings are for chromatograms from intact cacao samples, while the red markings refer to moisture-damaged cacao. (A) PCA loadings plot. (B) PCA on full dataset before alignment. (C) PCA on full dataset after alignment (D) PCA on the first 90 s of the chromatogram.

3.3.2. *Temperature increase rate 75 °C/min*

The second temperature program investigated started at 40°C which was held for 6 min, including 5 min of injection time, with the temperature subsequently increasing to 300°C at a rate of 75°C/min. The findings were illustrated on a chromatographic overlay (Figure 3.4), which showed that the overall appearance of the chromatogram was similar to that of the 50 °C/min program, with a slightly shorter runtime after the initial 5 min injection time period when the temperature was maintained at 40°C.

The results of the PCA are displayed in the Figure 3.4B, with PC1 accounting for 64.31% of the variance, while PC2 captured 25.89%, together accounting for 91.20% of the variance in the dataset. Notably, moisture-damaged and intact samples were fully separated. Interestingly, molded samples formed two subcategories, each consisting of two samples while the intact samples exhibited higher precision for reasons previously described in the 50 °C/min dataset section.

In contrast to the 50 °C/min program dataset, where class separation primarily occurred on PC1, both PC1 and PC2 appeared to contribute similarly to class separation in the 75 °C/min dataset. This observation is better explained by examining the loadings plot (Figure 3.4C), which indicated that PC1 was primarily derived from a class distinguishing peak 1 in a later eluting area, while PC2 was primarily influenced by the early eluting area and some random variation in the later area. As both PC1 and PC2 captured different important class differentiating sections, it is logical to require both of them to fully separate sample classes. Notably, the class distinguishing peak was tall, visually appealing, and well-defined, but it did not appear in all the molded samples, which hindered its usage for classification purposes.

To examine the early eluting area of the chromatogram, PCA was conducted on that section alone, as shown in Figure 3.4D. Interestingly, even though PC1 captured 54.67% of the variance while PC2 captured only 33.19%, the class separation occurred fully on PC2. Nevertheless, the PCA plot demonstrated effective class separation and classification of moisture-damaged and intact cacao.

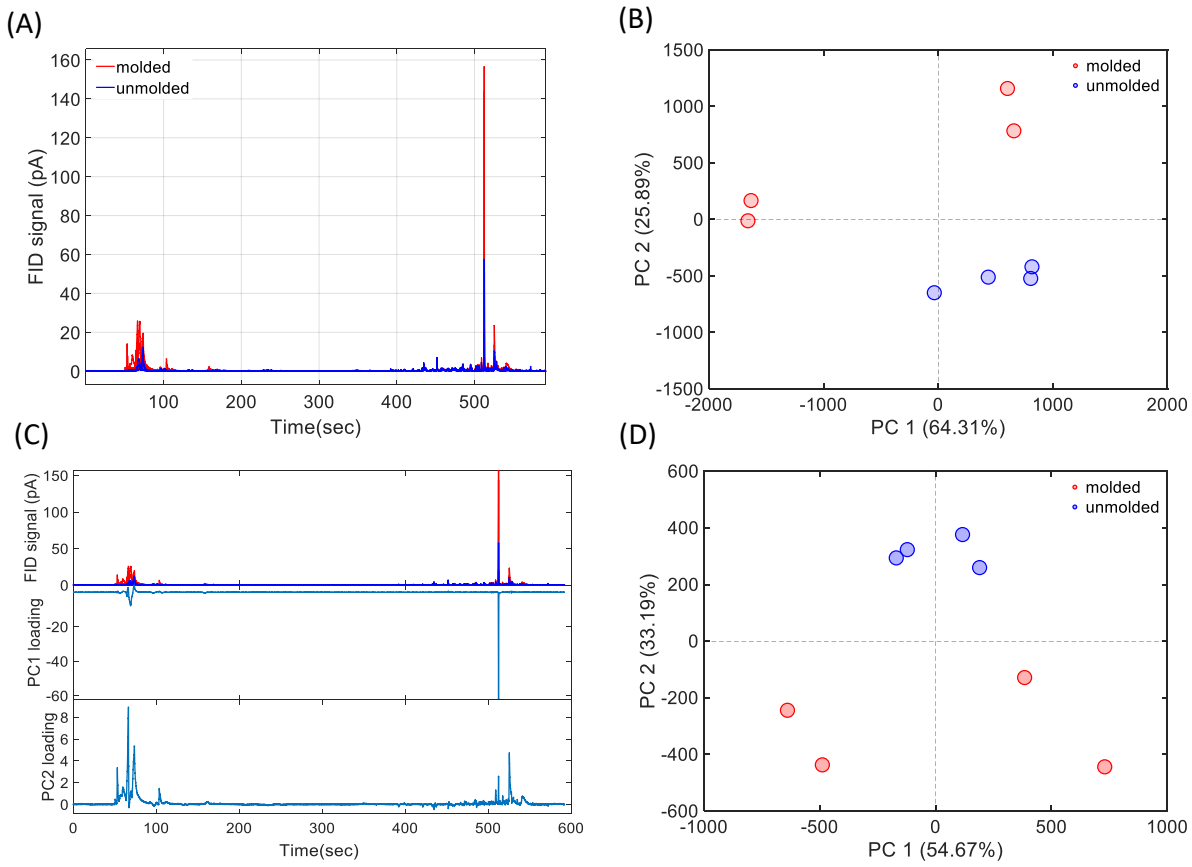

Figure 3.4. Temperature ramp of 75 °C/min. The blue marking refers to intact cacao, the red marking refers to moisture-damaged cacao. (A) Full chromatographic overlay. (B) PCA on the full dataset. (C) Loadings PCA on the full dataset. (D) PCA on the first 90 s of the dataset.

3.3.3. *Temperature increase rate 100 °C/min*

The temperature program described in this section started at 40°C which was holding for 6 min, including 5 min of injection time before subsequently increasing the temperature to 300°C at a rate of 100°C/min. For reference, this was the same as our temperature program used for the study of internal standards in the Chapter 2. It was also the first temperature ramp investigated in the study. Because it was our first investigated method, we collected 5 replicas of each sample compared to 4 replicas in other temperature programs, making a total of 10 chromatograms collected using this method. The full chromatographic overlay of 10 chromatograms is shown on a Figure 3.5A. The peaks appear reasonable following the same pattern as ones at ramps of 50 °C/min and 75 °C/min, and do not seem to lose peak resolution as it is shown on zoom in on a later eluting chromatographic window as given in Figure 3.5B. All class distinguishing peaks found in

the 50 °C/min dataset are still present and separated from other peaks. The zoom-in on an early eluting area is shown in Figure 3.5C. All intact samples have a similar response and show little to no peaks in this area. The moisture-damaged samples on the other hand all have significant peak coverage in the early eluting area.

Figure 3.6A shows the results of PCA of the full chromatograms of the described dataset. The PC1 captures 65.76% of the variance in the dataset while PC2 takes 17.3%. All samples in the two classes are fully separated on PC2 with intact samples having slightly better precision. Loadings to PCA are presented in Figure 3.6C. The PC1 loadings look almost identical to the original chromatogram. At the same time, PC2 shows the presence of class distinguishing peaks 1 and 2 and peaks in the early eluting area, these markers of moisture damage have a positive loading contribution while other peaks have a negative loading contribution. This explains why classes are separated on PC2 but not on PC1.

The results of PCA on early eluting peaks only are shown in Figure 3.6B. PC1 and PC2 have a similar contribution of 45.33% and 42.91% respectively. The class separation of only the early eluting area seems to be more effective than it was for the full chromatographic overlay. All intact samples are grouped in one tight cluster and have no overlapping with moisture-damaged sample scores. Molded samples show higher variability and are separated from the intact cluster on both PC1 and PC2. Loadings to this PCA show an interesting trend. PC1 captures mostly peaks from four moisture-damaged samples, while PC2 mimics the shape of fifth remaining moisture-damaged sample. This sample is marked with a star on the PCA plot as the only sample separated from the intact cluster primarily by PC2 and not PC1 (Figure 3.6B). This indicates that this sample may have some unique marker of a moisture damage such as a clump of mold that differentiates it from other 4 moisture damaged samples.

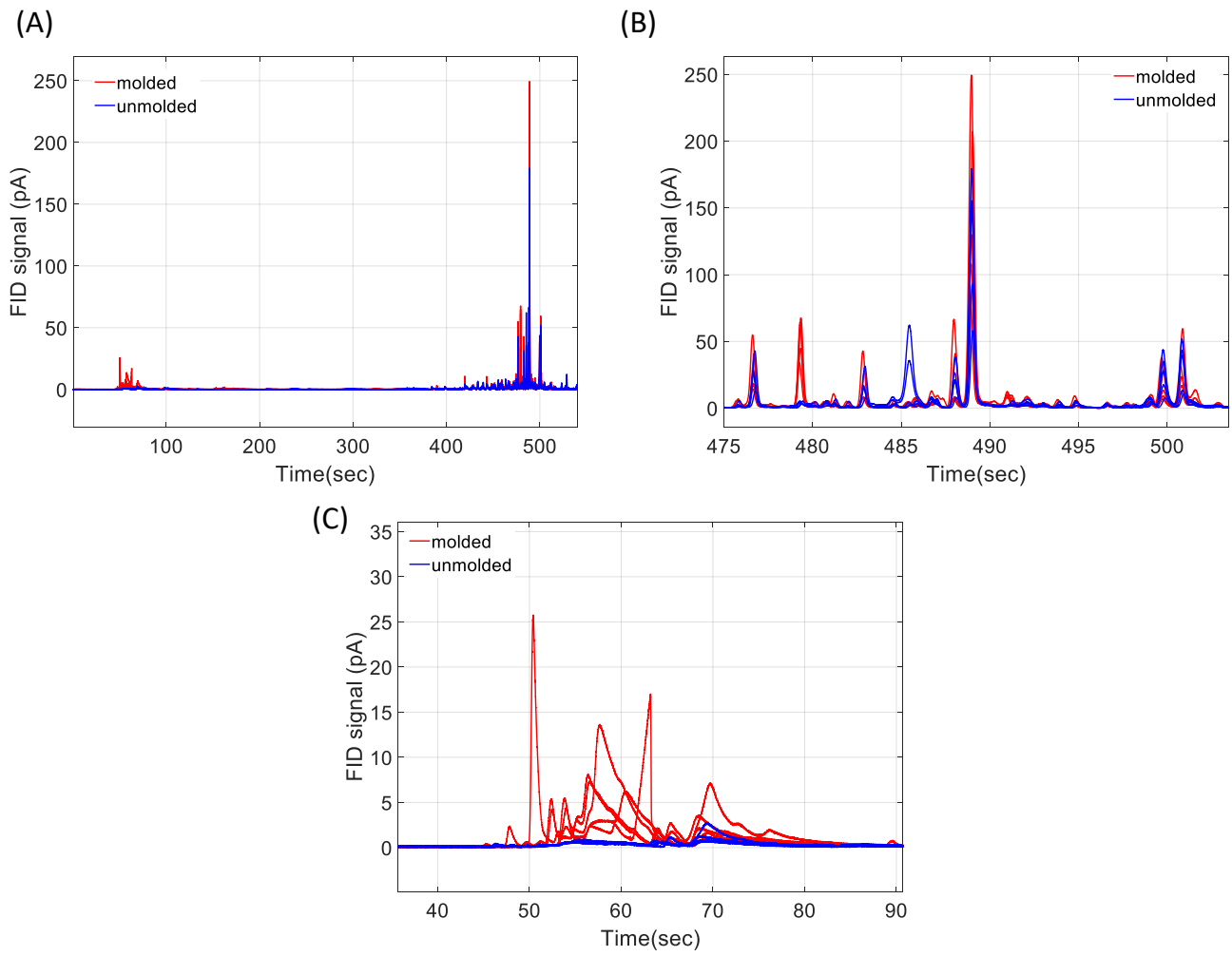

Figure 3.5. Temperature increase rate 100 °C/min. The blue chromatograms refer to intact cacao, the red chromatograms refer to moisture damaged cacao. (A) Full chromatographic overlay. (B) Zoom in on the chromatographic window. (C) Zoom in on the early eluting area.

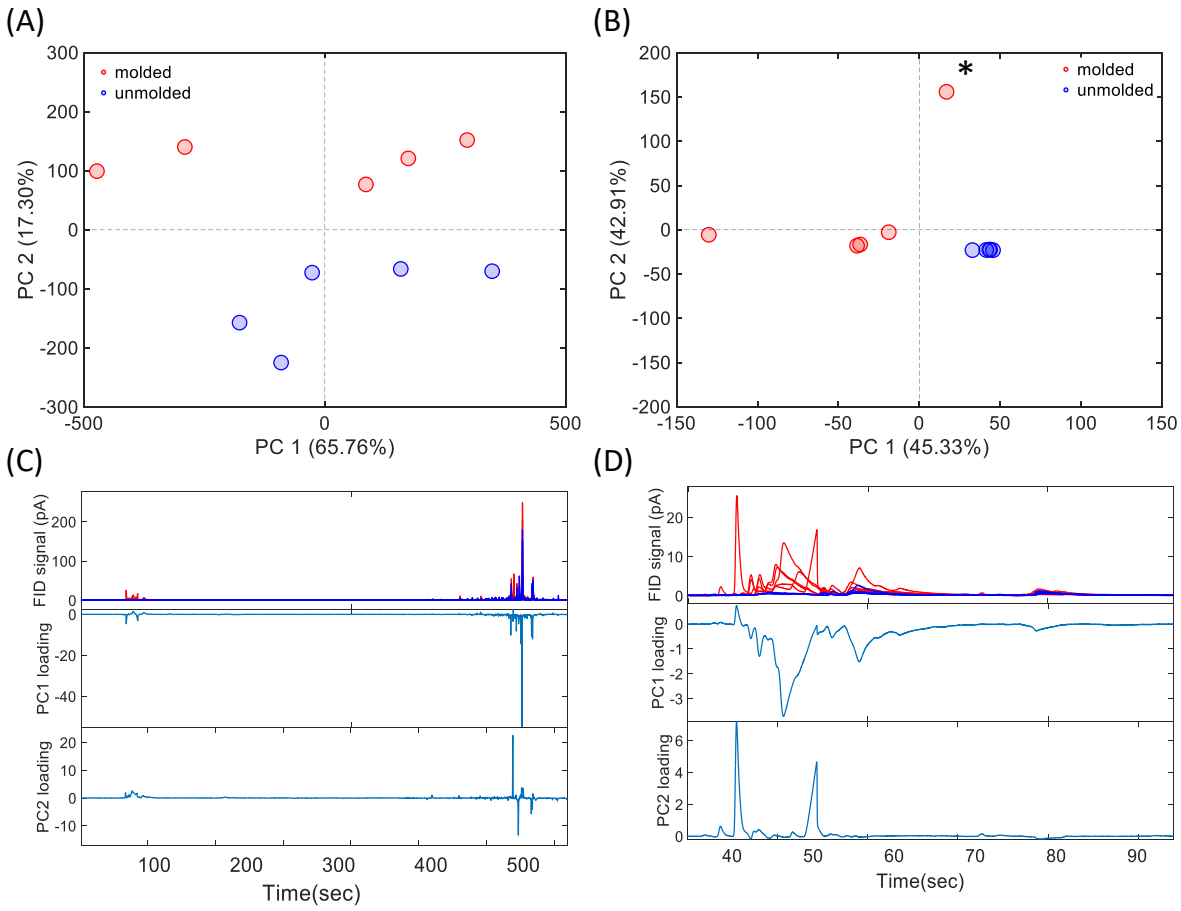

Figure 3.6. Temperature ramp of 100 °C/min. The blue marking refers to intact cacao, the red marking refers to moisture-damaged cacao. (A) PCA on full chromatograms. (B) PCA on the early eluting area (0-90 s). Starred point refers to a sample separated from an intact cluster (blue) by different features compared to other moisture-damaged samples (red). (C) Loadings to PCA on the full chromatogram. (D) Loadings to PCA on the early eluting area (0-90 s).

3.3.4. *Temperature increase rate 200 °C/min*

The temperature program described in this section started at 40°C which was held for 6 minutes, including 5 min of injection time, with the temperature subsequently rising to 300°C at a rate of 200°C/min with the final temperature being held for 1 min. The instrumental analysis took a total of 462 s to complete. The resulting full chromatographic overlay of all eight chromatograms (Figure 3.7A) revealed that while the early eluting area of the chromatogram remained relatively unchanged (as expected from the delay before the start of the temperature ramp), the later eluting area experienced a loss in resolution due to the faster temperature program. This resulted in a

smaller number of peaks being present, particularly in the later eluting area, as shown in the zoom-in on the chromatographic window (Figure 3.7C). Furthermore, while the class distinguishing peak 1 was still present, the class distinguishing peak 2 was no longer visible, as seen in the peak zoom-in (Figure 3.7B).

The PCA results using a ramp of 200 °C/min (Figure 3.7D) showed that while the sample classes were still distinguishable, they were not as well separated as in slower temperature programs due to the loss of resolution. Loadings analysis (Figure 3.7F) revealed that PC1 did not have a major contribution from the class distinguishing peaks, while PC2 had a larger contribution from class distinguishing peak 2 and early eluting peaks, explaining why the samples formed clusters on PC2. Therefore, both PC1 and PC2 did not demonstrate significant class differentiating properties.

To address the issue of lost resolution, PCA was performed on the first 40 to 110 s of the chromatographic overlay (Figure 3.7E), resulting in better PCA clustering by class than in the analysis of the full chromatogram. In this case, feature selection was necessary for effective characterization of cacao damage since the variation due to moisture difference was overpowered by random variation, which was increased by the faster temperature increase rate. Overall, this temperature program showed some limitations in its ability to effectively distinguish between sample classes due to the loss of resolution caused by the faster temperature program, although this issue was ultimately resolved by feature selection.

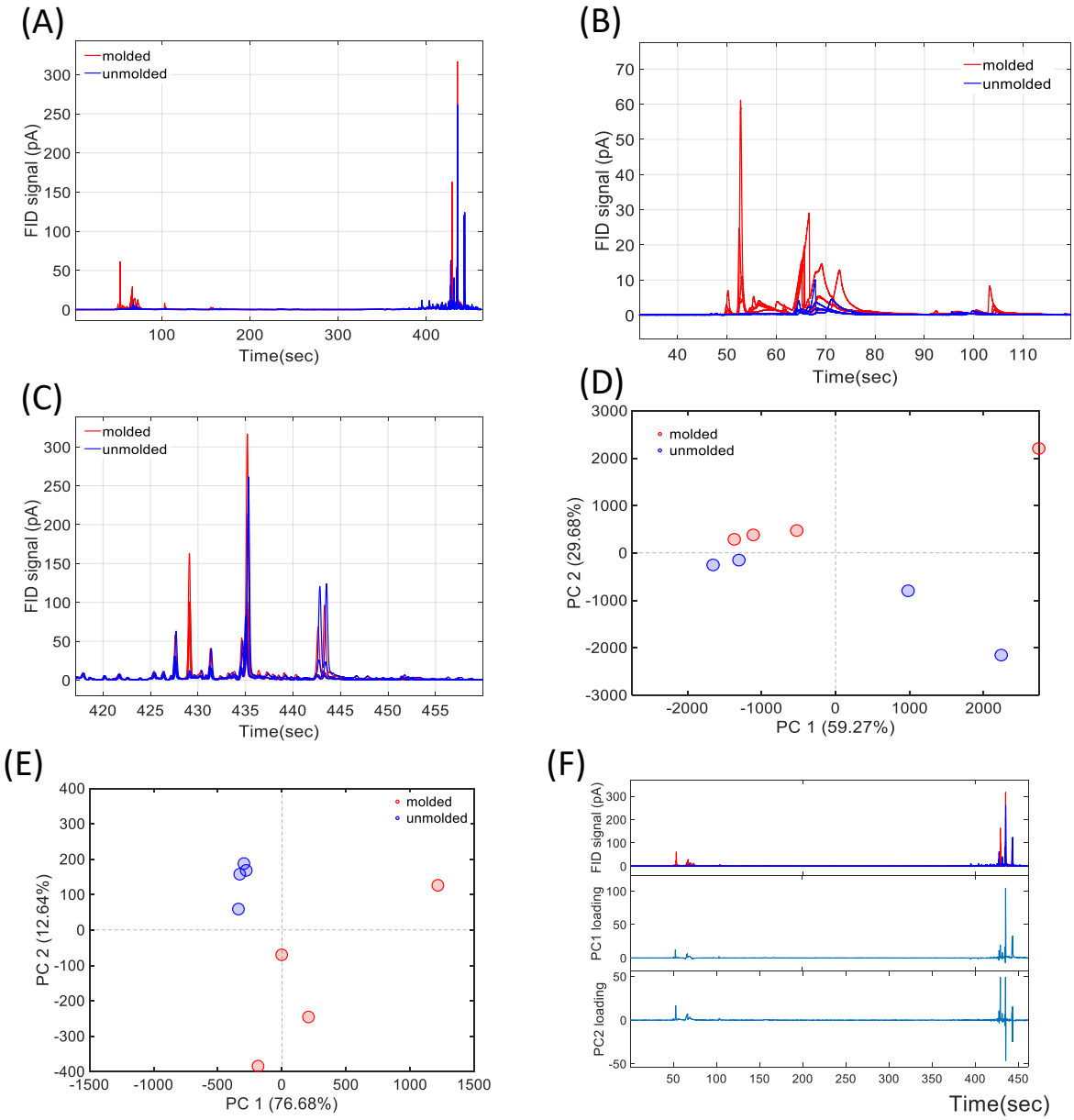

Figure 3.7. Temperature ramp of 200 °C/min chromatograms. The blue color refers to intact cacao, the red color refers to moisture damaged cacao. (A) Full chromatographic overlay. (B) Zoom in on the early eluting area. (C) Zoom in on the later eluting chromatographic window. (D) PCA on the full chromatographic overlay. (E) PCA on the early eluting area. (F) Loadings to PCA on the full chromatogram.

3.3.5. *Temperature increase rate 250 °C/min*

The INTUVO 900 GC system employed in this study is capable of achieving a maximum temperature increase rate of 250 °C/min for the HP-5MS columns used. The fastest temperature program was designed as follows: 40 °C initial temperature, followed by a 6 min hold, a temperature ramp of 250 °C/min up to 300 °C, and a 1 min hold at 300 °C, with a total run time of 446 s.

The full overlay of 8 chromatograms collected using this temperature program is presented in the Figure 3.8A. One moisture damaged sample was excluded from an analysis as an outlier as its chromatographic profile did not align with other samples collected in this study. The chromatogram showed a considerable portion of the chromatogram devoid of peaks, with a few peaks appearing at the beginning and more in a chromatographic window beginning at 400 s. Upon closer examination of the late chromatographic window (Figure 3.8B), it was evident that there was a significant loss of resolution, with only a few well-defined peaks present. One peak, marked on the chromatogram, was potentially the class-distinguishing peak 1; however, it was challenging to identify conclusively, as no statistically significant difference in intensity between the moisture-damaged samples and intact ones was observed.

PCA revealed interesting results, with PC1 capturing 90.11% of the variance in the dataset but only separating one of the three moisture-damaged samples from the intact samples (Figure 3.9B). PC2 meanwhile captured 7.61% of the variance but separated two remaining moisture damaged samples from the intact cacao cluster. The moisture-damaged samples were well separated from the intact cacao samples, with the unmolded samples forming a well-defined cluster, and the moisture-damaged samples being more spread out and forming two clusters. The PCA on the full chromatogram at a temperature increase rate of 250 °C/min appeared to better differentiate sample classes than the analysis at 200 °C/min, which was a surprising result. The explanation for this observation became apparent upon examination of the loadings plot (Figure 3.9A) and the PCA analysis conducted exclusively on the early eluting area of the chromatogram (Figure 3.9C). PC1 captured the variability from one moisture-damaged sample, exhibiting an unproportionally high area in what appears to be class-distinguishing peak 1. In contrast, PC2 loadings had the highest contribution coming from the early eluting area of the chromatogram. A

comparison of the early-eluting-area-only PCA plot with the full PCA plot revealed that the former was almost identical to the latter. This observation suggests that as the temperature increase rate was increased, the loss of resolution in the later eluting area of the chromatogram resulted in the loss of class-distinguishing variability, and the overall variability of the data decreased as well. Therefore, a PCA on the full chromatogram mimics the PCA on the exclusively early eluting area since the latter's contribution to the variability of the data increased compared to the later eluting area.

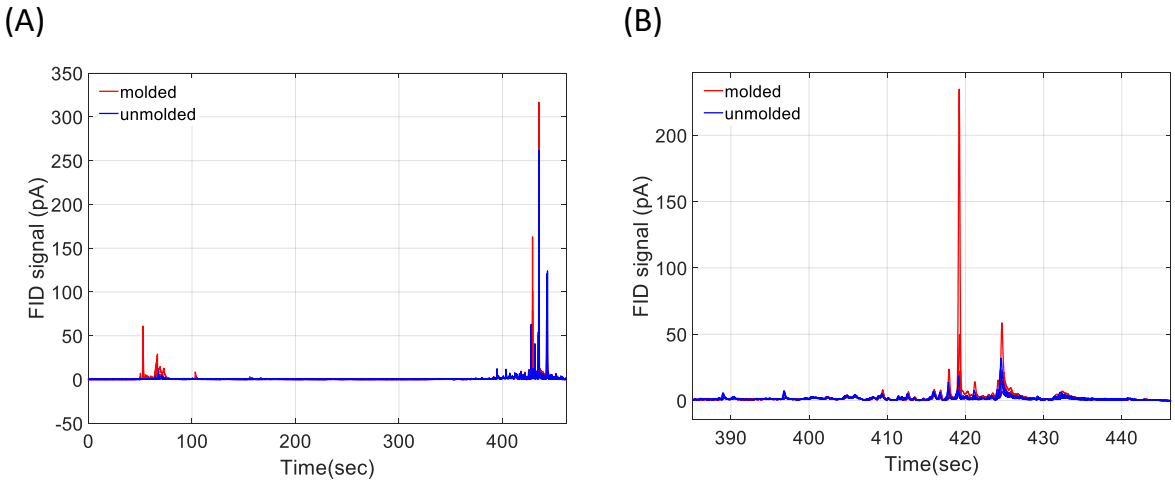

Figure 3.8. Temperature ramp of 250 °C/min. The blue chromatograms refer to intact cacao, the red chromatograms refer to moisture damaged cacao. (A) Full chromatographic overlay. (B) Zoom in on the later eluting chromatographic window.

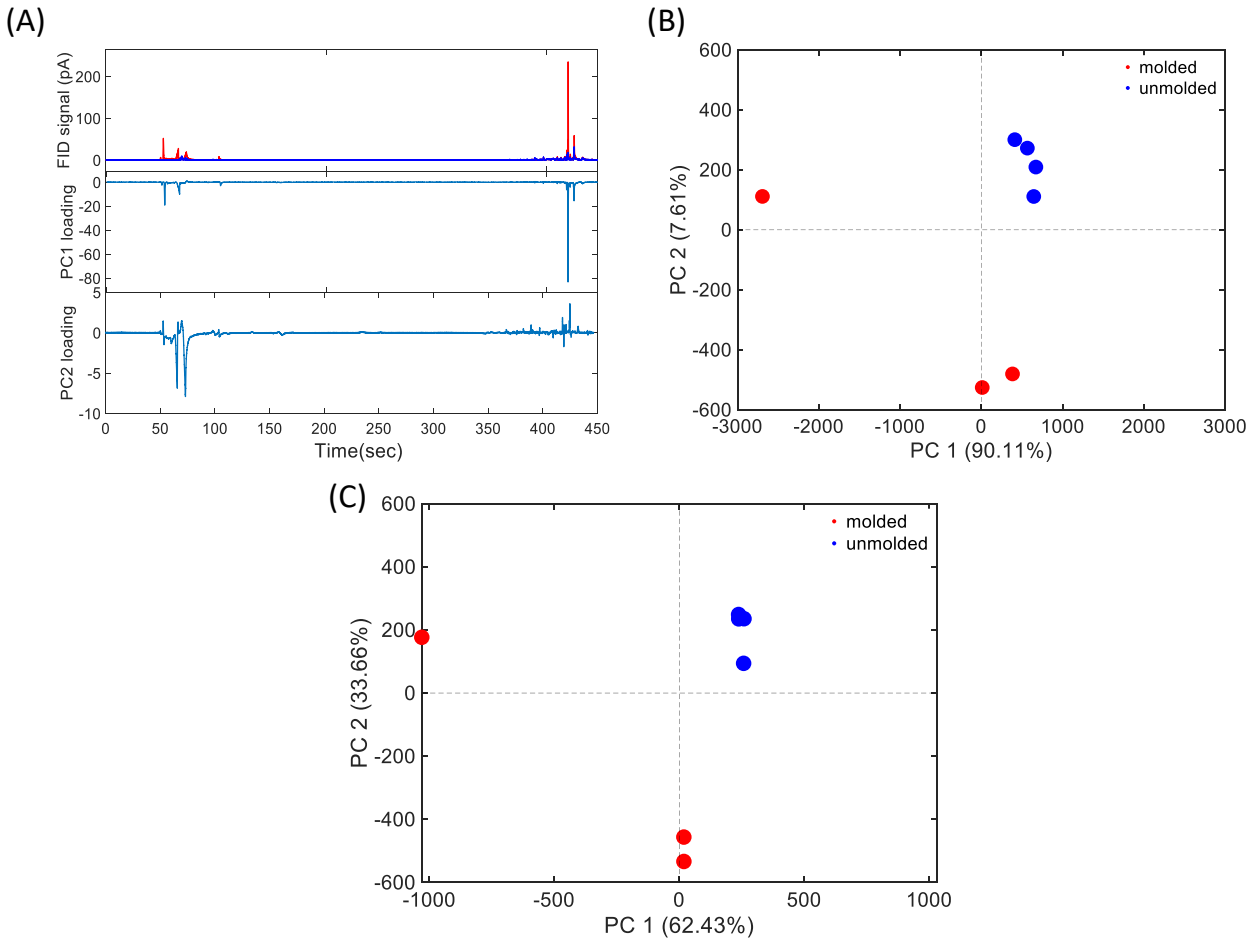

Figure 3.9. Temperature ramp of 250 °C/min. The blue circles - intact cacao, red circles-moisture damaged cacao (A) PCA on the early eluting area on the chromatogram. (B) PCA on the full chromatographic overlay. (C) PCA on the early eluting area on the chromatogram.

3.3.6. *Early eluting area*

Figure 3.10A presents a chromatographic overlay of all 40 chromatograms collected in this study within the time interval of 50-80 seconds from the overall runtime. Since the first 6 min of each run had identical temperature programs for all investigated temperature programs, it was possible to superimpose put windows from all chromatograms in a single matrix and treat it as a unified dataset. In Figure 3.10B, we present the PCA scores plot obtained from the analysis of this section of the overlay. All different temperature programs were collected on different days and were not aligned to compensate for time shifting in any way. PC1 captures 43.34% of the variance

in the dataset, while PC2 accounts for an additional 15.34%, collectively accounting for 58.68% of the variance in the dataset.

It is evident from the PCA scores plot that all 39 chromatograms are grouped along PC1 based on the presence of moisture damage. Intact cacao samples form a tightly clustered group on the plot, while the moisture-damaged samples exhibit more variability on both PC1 and PC2. The distance between two sample clusters on a PCA plot is, in a well-designed experiment, proportional to the degree of difference between the two samples data based on their quantifiable features. Given that the main source of variability in the chemical composition of our samples originate from moisture damage, it seems quite plausible that we can interpret the distance between a molded sample and the unmolded cluster as potentially proportional to the degree of moisture damage in that specific sample although, further research is needed to support this hypothesis.

The loadings plot provides a more profound understanding of the origin of the variance in the dataset (Figure 3.10C). It is noteworthy that, despite significant retention time shifting between runs, the PCA can still unambiguously identify moisture-damaged samples. This finding paves the way for the potential application of the developed method to compare and analyze data collected in different laboratories, on different instruments, and on different days without involving excessive computation as can be the case when COW alignment is needed.

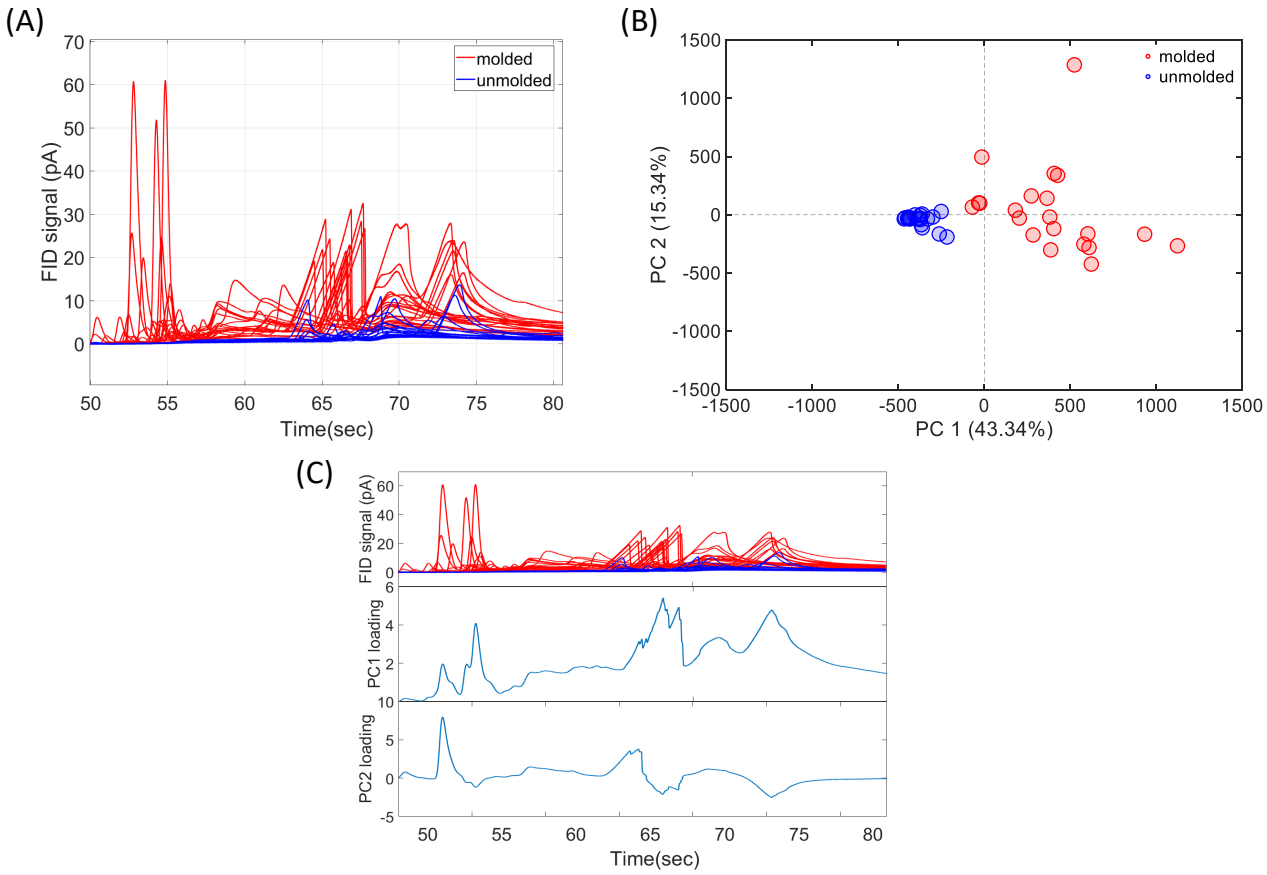

Figure 3.10. Early eluting area. Blue coloring refers to intact cacao, red color refers to moisture damaged cacao. (A) Chromatographic overlay of all 39 collected chromatograms. We were able to overlay all different temperature ramps at the first 80 s because all programs started identically for the first 6 min. (B) PCA on the early eluting area in all 39 chromatograms. (C) Loadings to PCA on full chromatographic overlay in early eluting area.

3.4. Conclusions And A Future Prospectus

The detrimental effects of moisture damage on cacao are widely recognized in the industry, as it can negatively impact both taste and safety of a final product. Various types and degrees of moisture damage exist, with resulting products being labeled and priced accordingly. However, the conventional approach to classifying moisture-damaged cacao is often subjective, labor-intensive, and time-consuming. A promising alternative involves the use of solid-phase microextraction gas chromatography (HS-SPME-GC) to detect volatile organic compounds (VOCs) associated with moisture damage. While mass spectrometry (MS) is a highly effective

tool for this purpose, its implementation may not be practical due to cost constraints and increased complexity involved. As such, in this study we explored the feasibility of utilizing HS-SPME-GC and chemometrics for the rapid and cost-effective classification of moisture-damaged cacao. Specifically, PCA was utilized to classify such products. The study successfully developed a method for classifying moisture-damaged cacao samples based on their VOC profiles using HS-SPME-GC analysis. Several temperature programs were developed, tested and compared. The results demonstrated that all of the investigated temperature programs were capable of distinguishing between moisture-damaged and intact cacao samples. Moreover, several class-distinguishing peaks were identified, which enabled accurate differentiation between the two sample types. It was observed that shorter temperature ramps resulted in a loss of resolution in later eluting peaks, although this obstacle was overcome by applying manual feature selection. A significant number of class-distinguishing analytes was found in the early eluting region of the chromatogram. More specifically, the first 80 s of the chromatogram contained enough class-distinguishing analytes to effectively separate moisture-damaged and intact cacao samples.

Moving forward, several potential areas of future research could build upon the findings of this study. Firstly, further investigation is needed to identify the specific class distinguishing analytes that are responsible for the differentiation between moisture-damaged and intact cacao samples. This would provide a more detailed understanding of the chemical changes that occur during moisture damage and allow for the development of even more sensitive and specific classification protocols. Secondly, it may be possible to reduce the analysis duration by optimizing the SPME extraction and equilibration times. This could enable faster and more efficient analysis of cacao samples, potentially allowing for higher throughput and more frequent quality control measurements. Thirdly, there is potential to develop an ultra-fast method of identifying moisture damage in cacao by focusing solely on the class distinguishing analytes that elute in the first 90 s of the GC run. This would allow for the rapid screening of cacao samples and could provide a valuable tool for the industry. Fourthly, automation of the analysis process could also be explored. Automating the SPME and GC processes would increase efficiency and reproducibility, and could enable higher throughput of cacao sample analysis. Another improvement to consider could be a development on the method that would allow for the determination of the degree of a moisture damage in a quantitative way rather than rely on classification alone. Finally, it would be valuable to explore the applicability of the findings of this study to conventional GC systems. While the

study utilized a specialized fast GC system, the development of practical and efficient quality control protocols for the cacao industry would require the application of the findings to more widely available and accessible GC systems.

Chapter 4. CONCLUSIONS

The two chapters presented in this thesis focused on improving the quality and reliability of SPME analysis for various applications. Chapter 1 investigated the use of internal standards to improve the reproducibility and accuracy of SPME analysis. The study evaluated the performance of multiple compounds as internal standards and identified 1-bromooctane as the most successful normalization approach for reducing in-class variations and improving the consistency of internal standard peaks. However, the study noted that some analytes may suffer from broadening during the SPME process, which may not be easily resolved by method adjustment.

In Chapter 2, the study developed a cost-effective and efficient method for classifying moisture-damaged and intact cacao samples based on their VOC profiles using HS-SPME-GC analysis. The study successfully developed a classification model for such products using PCA and identified several class-distinguishing peaks that enabled accurate differentiation between the two sample types. The study also investigated the impact of changes in temperature program parameters on the classification of cacao samples and identified the limit of the temperature increase rate beyond which accurate classification of moisture-damaged cacao samples is not possible. The results of this study provide valuable insights for developing practical and efficient classification protocols that are both sensitive and specific and can contribute to the development of improved quality control protocols in the cacao industry.

Overall, these two chapters demonstrate the potential of HS-SPME-GC analysis for improving the accuracy and reliability of analytical methods for various applications. The findings of this study have important implications for the development of cost-effective and efficient classification protocols for the analysis of complex mixtures.

BIBLIOGRAPHY

(1) James, A. T.; Martin, A. J. P. Gas-Liquid Partition Chromatography: The Separation and Micro-Estimation of Volatile Fatty Acids from Formic Acid to Dodecanoic Acid. Biochemical Journal 1952, 50 (5), 679–690. https://doi.org/10.1042/bj0500679.

(2) Denoyer, E. Agilent Intuvo 9000 GC System. LC-GC North America 2017, 35 (3), 5–5.

(3) Biziuk, M. Solid Phase Extraction Technique-Trends, Opportunities and Applications; 2015. https://www.researchgate.net/publication/279597851.

(4) Maranata, G. J.; Surya, N. O.; Hasanah, A. N. Optimising Factors Affecting Solid Phase Extraction Performances of Molecular Imprinted Polymer as Recent Sample Preparation Technique. Heliyon. Elsevier Ltd January 1, 2021. https://doi.org/10.1016/j.heliyon.2021.e05934.

(5) Abd-Talib, N.; Mohd-Setapar, S. H.; Khamis, A. K. The Benefits and Limitations of Methods Development in Solid Phase Extraction: Mini Review. Jurnal Teknologi (Sciences and Engineering) 2014, 69 (4), 69–72. https://doi.org/10.11113/jt.v69.3177.

(6) Vas, G.; Vékey, K. Solid-Phase Microextraction: A Powerful Sample Preparation Tool Prior to Mass Spectrometric Analysis. Journal of Mass Spectrometry. March 2004, pp 233–254. https://doi.org/10.1002/jms.606.

(7) Belardi Robert P.; Pawliszyn Janusz B. The Application of Chemically Modified Fused Silica Fibers in the Extraction of Organics from Water Matrix Samples and Their Rapid Transfer to Capillary Columns. Water Quality Research Journal 1989, 24 (1), 179–191. https://doi.org/https://doi.org/10.2166/wqrj.1989.010.

(8) Reyes-Garcés, N.; Gionfriddo, E.; Gómez-Ríos, G. A.; Alam, M. N.; Boyacl, E.; Bojko, B.; Singh, V.; Grandy, J.; Pawliszyn, J. Advances in Solid Phase Microextraction and Perspective on Future Directions. Analytical Chemistry. American Chemical Society January 2, 2018, pp 302–360. https://doi.org/10.1021/acs.analchem.7b04502.

(9) Boyaci, E.; Rodríguez-Lafuente, Á.; Gorynski, K.; Mirnaghi, F.; Souza-Silva, É. A.; Hein, D.; Pawliszyn, J. Sample Preparation with Solid Phase Microextraction and Exhaustive Extraction Approaches: Comparison for Challenging Cases. Analytica Chimica Acta. Elsevier B.V. May 11, 2015, pp 14–30. https://doi.org/10.1016/j.aca.2014.12.051.

(10) Pawliszyn, J. Theory of Solid-Phase Microextraction. In Handbook of Solid Phase Microextraction; Elsevier Inc., 2012; pp 13–59. https://doi.org/10.1016/B978-0-12-416017-0.00002-4.

(11) Basheer, C.; Lee, H. K. Hollow Fiber Membrane-Protected Solid-Phase Microextraction of Triazine Herbicides in Bovine Milk and Sewage Sludge Samples. J Chromatogr A 2004, 1047 (2), 189–194. https://doi.org/10.1016/j.chroma.2004.06.130.

(12) Jalili, V.; Barkhordari, A.; Ghiasvand, A. A Comprehensive Look at Solid-Phase Microextraction Technique: A Review of Reviews. Microchemical Journal. Elsevier Inc. January 1, 2020. https://doi.org/10.1016/j.microc.2019.104319.

(13) Schmidt, K.; Podmore, I. Current Challenges in Volatile Organic Compounds Analysis as Potential Biomarkers of Cancer. J Biomark 2015, 2015, 1–16. https://doi.org/10.1155/2015/981458.

(14) Aqel, A.; Dhabbah, A. M.; Yusuf, K.; AL-Harbi, N. M.; Al Othman, Z. A.; Yacine Badjah-Hadj-Ahmed, A. Determination of Gasoline and Diesel Residues on Wool, Silk, Polyester and Cotton Materials by SPME–GC–MS. Journal of Analytical Chemistry 2016, 71 (7), 730–736. https://doi.org/10.1134/S1061934816070029.

(15) Prada, P.; Curran, A.; Furton, K. Characteristic Human Scent Compounds Trapped on Natural and Synthetic Fabrics as Analyzed by SPME-GC/MS. J Forensic Sci Criminol 2014, 1 (1). https://doi.org/10.15744/2348-9804.1.S101.

(16) DeGreeff, L. E.; Furton, K. G. Collection and Identification of Human Remains Volatiles by Non-Contact, Dynamic Airflow Sampling and SPME-GC/MS Using Various Sorbent Materials. Anal Bioanal Chem 2011, 401 (4), 1295–1307. https://doi.org/10.1007/s00216-011-5167-0.

(17) Van Lancker, F.; Adams, A.; Delmulle, B.; De Saeger, S.; Moretti, A.; Van Peteghem, C.; De Kimpe, N. Use of Headspace SPME-GC-MS for the Analysis of the Volatiles Produced by Indoor Molds Grown on Different Substrates. Journal of Environmental Monitoring 2008, 10 (10), 1127. https://doi.org/10.1039/b808608g.

(18) Fitz, B. D.; Mannion, B. C.; To, K.; Hoac, T.; Synovec, R. E. Evaluation of Injection Methods for Fast, High Peak Capacity Separations with Low Thermal Mass Gas Chromatography. J Chromatogr A 2015, 1392, 82–90. https://doi.org/10.1016/j.chroma.2015.03.009.

(19) Porcelli, C.; Neiens, S. D.; Steinhaus, M. Molecular Background of a Moldy-Musty Off-Flavor in Cocoa. J Agric Food Chem 2021, 69 (15), 4501–4508. https://doi.org/10.1021/acs.jafc.1c00564.

(20) FORSYTH, W. G. C. A Method for Studying the Chemistry of Cacao Fermentation. Nature 1949, 164 (4157), 25–26. https://doi.org/10.1038/164025a0.

(21) Robards, K.; Kerr, A. F.; Patsalides, E. Rancidity and Its Measurement in Edible Oils and Snack Foods. A Review. Analyst 1988, 113 (2), 213. https://doi.org/10.1039/an9881300213.

(22) Tan, J.; Balasubramanian, B.; Sukha, D.; Ramkissoon, S.; Umaharan, P. Sensing Fermentation Degree of Cocoa (Theobroma Cacao L.) Beans by Machine Learning Classification Models Based Electronic Nose System. J Food Process Eng 2019, 42 (6). https://doi.org/10.1111/jfpe.13175.

(23) Humston, E. M.; Knowles, J. D.; McShea, A.; Synovec, R. E. Quantitative Assessment of Moisture Damage for Cacao Bean Quality Using Two-Dimensional Gas Chromatography Combined with Time-of-Flight Mass Spectrometry and Chemometrics. J Chromatogr A 2010, 1217 (12), 1963–1970. https://doi.org/10.1016/j.chroma.2010.01.069.

(24) Humston, E. M.; Knowles, J. D.; McShea, A.; Synovec, R. E. Quantitative Assessment of Moisture Damage for Cacao Bean Quality Using Two-Dimensional Gas Chromatography Combined with Time-of-Flight Mass Spectrometry and Chemometrics. J Chromatogr A 2010, 1217 (12), 1963–1970. https://doi.org/10.1016/j.chroma.2010.01.069.

(25) Qin, X.-W.; Lai, J.-X.; Tan, L.-H.; Hao, C.-Y.; Li, F.-P.; He, S.-Z.; Song, Y.-H. Characterization of Volatile Compounds in Criollo, Forastero, and Trinitario Cocoa Seeds (Theobroma Cacao L.) in

China. Int J Food Prop 2017, 20 (10), 2261–2275. https://doi.org/10.1080/10942912.2016.1236270.

(26) Cevallos-Cevallos, J. M.; Gysel, L.; Maridueña-Zavala, M. G.; Molina-Miranda, M. J. Time-Related Changes in Volatile Compounds during Fermentation of Bulk and Fine-Flavor Cocoa (Theobroma Cacao) Beans. J Food Qual 2018, 2018, 1–14. https://doi.org/10.1155/2018/1758381.

(27) Marseglia, A.; Musci, M.; Rinaldi, M.; Palla, G.; Caligiani, A. Volatile Fingerprint of Unroasted and Roasted Cocoa Beans (Theobroma Cacao L.) from Different Geographical Origins. Food Research International 2020, 132, 109101. https://doi.org/10.1016/j.foodres.2020.109101.

(28) Humston, E. M.; Zhang, Y.; Brabeck, G. F.; McShea, A.; Synovec, R. E. Development of a GC×GC-TOFMS Method Using SPME to Determine Volatile Compounds in Cacao Beans. J Sep Sci 2009, 32 (13), 2289–2295. https://doi.org/10.1002/jssc.200900143.

(29) Humston, E. M.; Zhang, Y.; Brabeck, G. F.; McShea, A.; Synovec, R. E. Development of a GCxGC-TOFMS Method Using SPME to Determine Volatile Compounds in Cacao Beans. J Sep Sci 2009, 32 (13), 2289–2295. https://doi.org/10.1002/jssc.200900143.

(30) Rodriguez-Campos, J.; Escalona-Buendía, H. B.; Orozco-Avila, I.; Lugo-Cervantes, E.; Jaramillo-Flores, M. E. Dynamics of Volatile and Non-Volatile Compounds in Cocoa (Theobroma Cacao L.) during Fermentation and Drying Processes Using Principal Components Analysis. Food Research International 2011, 44 (1), 250–258. https://doi.org/10.1016/j.foodres.2010.10.028.

(31) Barbosa-Pereira, L.; Rojo-Poveda, O.; Ferrocino, I.; Giordano, M.; Zeppa, G. Analytical Dataset on Volatile Compounds of Cocoa Bean Shells from Different Cultivars and Geographical Origins. Data Brief 2019, 25, 104268. https://doi.org/10.1016/j.dib.2019.104268.

(32) Torres-Moreno, M.; Tarrega, A.; Blanch, C. Effect of Cocoa Roasting Time on Volatile Composition of Dark Chocolates from Different Origins Determined by HS-SPME/GC-MS. CyTA - Journal of Food 2021, 19 (1), 81–95. https://doi.org/10.1080/19476337.2020.1860137.

(33) Utrilla-Vázquez, M.; Rodríguez-Campos, J.; Avendaño-Arazate, C. H.; Gschaedler, A.; Lugo-Cervantes, E. Analysis of Volatile Compounds of Five Varieties of Maya Cocoa during Fermentation and Drying Processes by Venn Diagram and PCA. Food Research International 2020, 129, 108834. https://doi.org/10.1016/j.foodres.2019.108834.

(34) Utrilla-Vázquez, M.; Rodríguez-Campos, J.; Avendaño-Arazate, C. H.; Gschaedler, A.; Lugo-Cervantes, E. Analysis of Volatile Compounds of Five Varieties of Maya Cocoa during Fermentation and Drying Processes by Venn Diagram and PCA. Food Research International 2020, 129, 108834. https://doi.org/10.1016/j.foodres.2019.108834.