Mémoire

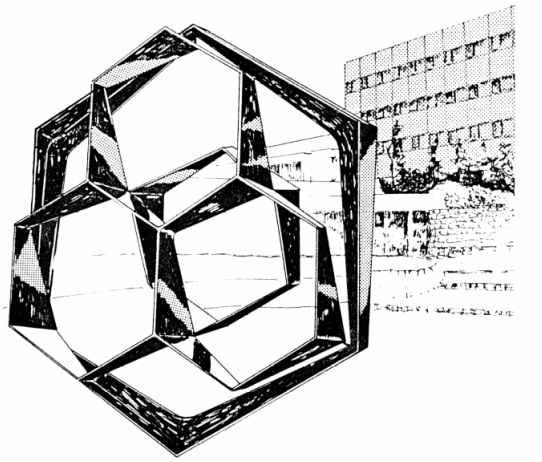
Auteur : Hansen, Maxime
Promoteur(s) : Focant, Jean-François
Faculté : Faculté des Sciences
Diplôme : Master en sciences chimiques, à finalité didactique
Année académique : 2022-2023
URI/URL : http://hdl.handle.net/2268.2/18463

FACULTY OF SCIENCES

Chemistry departement

OBiAChem - Prof. J-F Focant

Development of an analytical workflow for metabolomics profiling of in vitro gastrointestinal model



Years 2022-2023

LIÈGE université

Dissertation presented by Hansen Maxime for the fulfilment of the degree of Master in Chemical Sciences

Dedication

A special thanks to DEJONG Thibaut for his involvement in my training and guidance through this work. Without him, this work could not have been possible.

Thank you to all the members of the OBiAChem team for their good mood and guidance, especially STEFANUTO Pierre-Hugues for his time and his precious advice. I am especially grateful to FOCANT Jean-François for letting me pursue a master's thesis in his laboratory.

Thank you to BERTRAND Virginie for her good mood and precious advice during my training with cellular culture.

Thank you to the reading committee, STEFANUTO Pierre-Hugues, BERTRAND Virginie, and FAR Johann.

Thank you to my family, who supported and let me pursue my dream by giving me this opportunity.

Finally, thank you to my girlfriend, who helped me go through all this work.

Contents

1 Introduction

Inflammation serves as a defensive mechanism within the body, initiated by the immune system in response to various factors, including pathogens, physical trauma, or chemical oxidation[1], [2]. It represents the initial phase of the healing process and can be classified into two main types: acute or chronic[3]. A key distinction between these categories lies in the timing of the responses following the stimulus that triggers an immune response, with acute inflammation occurring within the first few hours or days and gradually subsiding. Conversely, chronic inflammation may persist for extended periods, spanning months or even years. In addition to the duration, chronic inflammation is characterized by other biological changes such as the presence of lymphocytes, macrophages, and plasma cells in the affected tissues[2], [4].

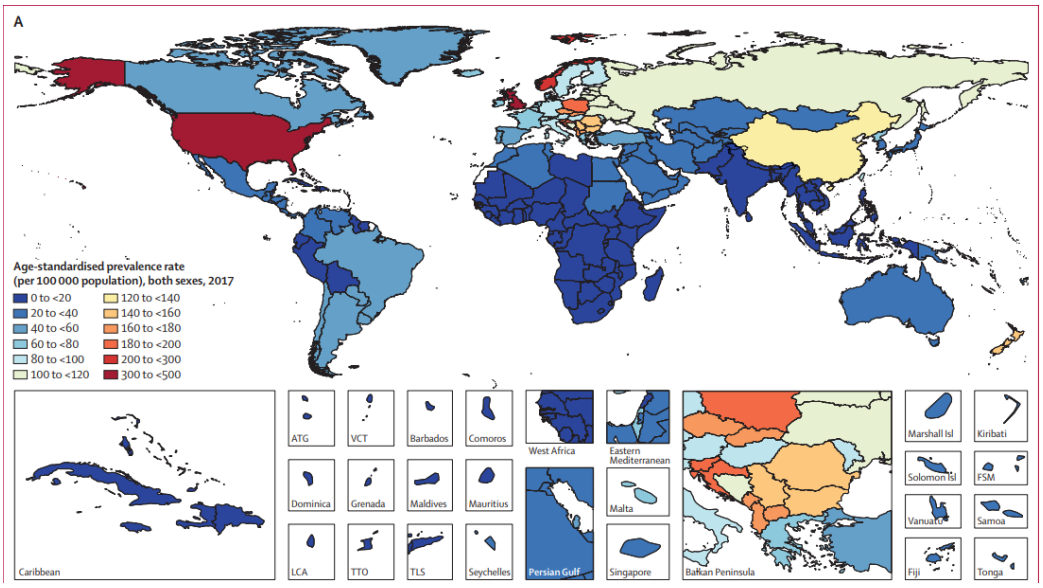


Figure 1: Age-standardized prevalence rate (per 100000 population) of IBD, both sexes, for 195 countries and territories, 2017. IBD=inflammatory bowel disease. ATG=Antigua and Barbuda. VCT=Saint Vincent and the Grenadines. LCA=Saint Lucia. TTO=Trinidad and Tobago. TLS=Timor-Leste. FSM=Federated States of Micronesia. *Adapted from [5].*

Chronic inflammation, characterized by its prolonged duration, can contribute to the development of various chronic inflammatory diseases, includ-

ing heart disorders, cancer, kidney diseases, and diabetes[2], [6]. Among these conditions, inflammation affecting the gastrointestinal cells can give rise to a range of disorders. Notably, Inflammatory Bowel Disease (IBD) encompasses two specific conditions resulting from chronic inflammations of the gastrointestinal tract: Crohn's disease and ulcerative colitis[7], [8]. As of 2017, the global prevalence of IBD worldwide was estimated to be 3.9 million cases among females and nearly 3 million among males. Given the anticipated increase in these numbers, healthcare systems and economies worldwide are expected to allocate resources more strategically to address this growing health concern[5]. Figure 1 provides a global overview of the prevalence rate of IDB. Consequently, it is crucial to comprehend the complexity of IBD and the underlying mechanisms driving inflammation.

To investigate the intricate complexity of inflammation, this study aims to develop an analytical workflow for metabolomics profiling in an *in vitro* gastrointestinal models. Subsequent sections will delve into the concepts mentioned above in greater detail.

1.1 Model for inflammation

To gain a comprehensive understanding of the intricacies involved in gastrointestinal tract inflammation, *in vivo* and clinical studies can be employed to provide valuable insights into these mechanisms. However, these approaches raise ethical concerns regarding the use of animal or human subjects for testing. Additionally, the inability to control specific parameters during experiments present a significant barrier to a specific understanding of the inflammation mechanisms[9]. To overcome these challenges, an *in vitro* model can be utilized to assess the inflammation mechanisms in gastrointestinal cells. This method offers several advantages[9], [10]:

- No ethical considerations associated with human or animal testing
- Control and reproducibility by using controlled experimental conditions
- Accessibility and cost-effectiveness by allowing researchers to develop a method for large-scale cell production and experimentation

However, it is important to acknowledge that *in vitro* models possess inherent limitations as they represent simplified versions of the gastrointestinal

tract, lacking the dynamic complexity observed within present a human living body[11]. Nevertheless, *In vitro* studies serve as fundamental models to provide valuable insights into the intricate nature of a system. On the other hand, *in vivo* studies enable the investigation of mechanisms under natural physiological conditions. Thus, both approaches complement each other and contribute to a comprehensive understanding of gastrointestinal inflammation.

An *in vitro* model can be developed using gastrointestinal-derived cellular lines to simulate the complex environment of the gastrointestinal tract. Given that the epithelial cells form the primary barrier in the intestinal tract, they play a pivotal role in initiating the immunological response and act as a source of cytokines that contribute to pro-inflammatory reactions[3]. Figure 2 illustrates the sequential events leading to an inflammatory response originating from the intestinal epithelial cells (IECs). For a more comprehensive understanding of this mechanism, the work conducted by Samoilă *et al.*[12] offers detailed insights and information.

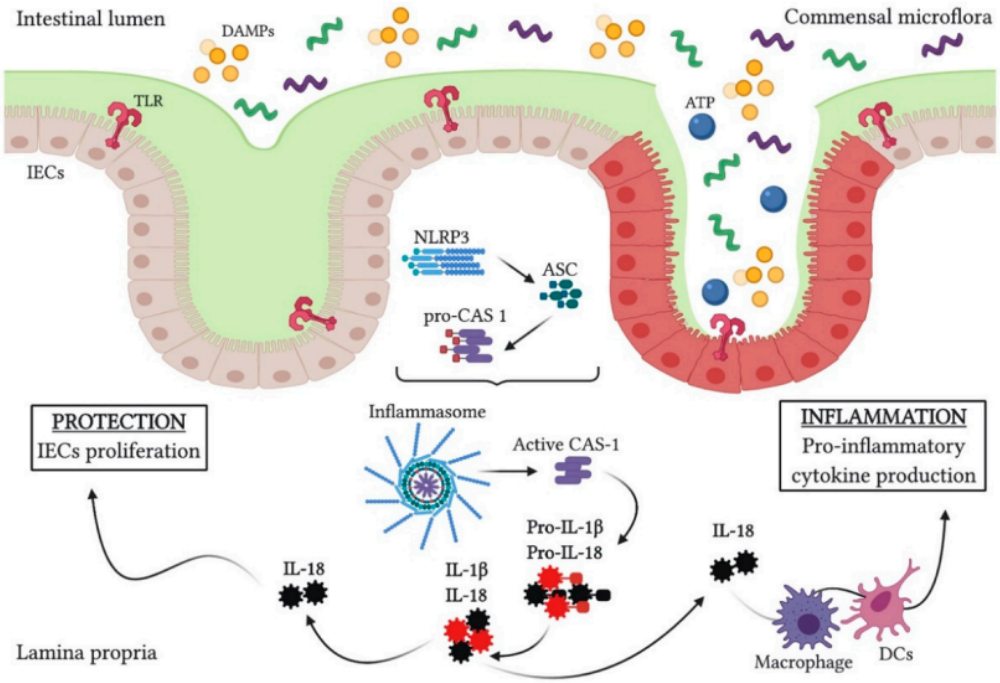


Figure 2: Chain of events leading to promotion of inflammation. *Adapted from [12]*

In summary, when the mucus layer is compromised, the contact between IECs and bacteria from the mucosa triggers an immune response. This response is mediated by innate immune sensors known as Pattern Recognition Receptors (PRRs), which recognize specific stimuli such as Damage-Associated Molecular Patterns (DAMPs). Upon recognition, NLRP3 and procaspase 1 (pro-CAS) are produced, initiating the assembly of a complex called the "Inflammasome". This complex, in turn, stimulates the production of IL-18 and IL-1β cytokines. While IL-18 promotes the proliferation of IECs, excessive production of these cytokines can activate immune cells and result in an overproduction of pro-inflammatory cytokines, ultimately triggering an inflammatory response[12].

1.2 Metabolomics to assess IBD

Inflammation involves the activation of the immune system in response to various factors, such as microbial infections or injured tissue[13]. To comprehend the intricate mechanisms underlying inflammation, it is crucial to quantify and identify the chemical mediators involved. These mediators provide valuable insights into the chain of events leading to inflammation. Several mediators have been identified[14], but due to their diverse nature, encompassing proteins, DNA, or metabolites, different techniques are employed to investigate the mechanisms of inflammation[15].

The omics approach is employed to gain a comprehensive understanding of these mechanisms. This field aims, as described by Vailati-Riboni *et al.*, to: "*identify, characterize, and quantify all biological molecules that are involved in the structure, function, and dynamics of a cell, tissue, or organism*"[16]. Figure 3 provides an overview of the different branches of omics, which encompass the study, study of various types of biological molecules, ranging from large-sized molecules to smaller ones. Within the realm of omics, specific areas of study can be dedicated to investigating the mechanisms of inflammation. Metabolomics, in particular, can be employed to study these mechanisms[17]. This area of study involves the comprehensive analysis and quantification of low molecular weight molecules lesser than 1 kDa[18], known as metabolites, present in biological systems. It aims to capture intermediates or products from enzymatic activities to food, medication or environmental factors[19]. Analytical techniques such as liquid or gas chromatography coupled with mass spectrometry or nuclear magnetic

resonance (NMR) [16], [19] are used to derive an integrated picture of the metabolome. The identified compound can be used as proximal reporters of disease and to be used as markers for clinical tests. As metabolomics can be further separated depending on different techniques, the specific field of metabolomics studying the volatile organic compounds (VOCs) released by a sample in the gaseous phase used is hence called volatolomics. Therefore, volatolomics is used for this work to study the metabolomics profiling of inflammatory gastrointestinal cells.

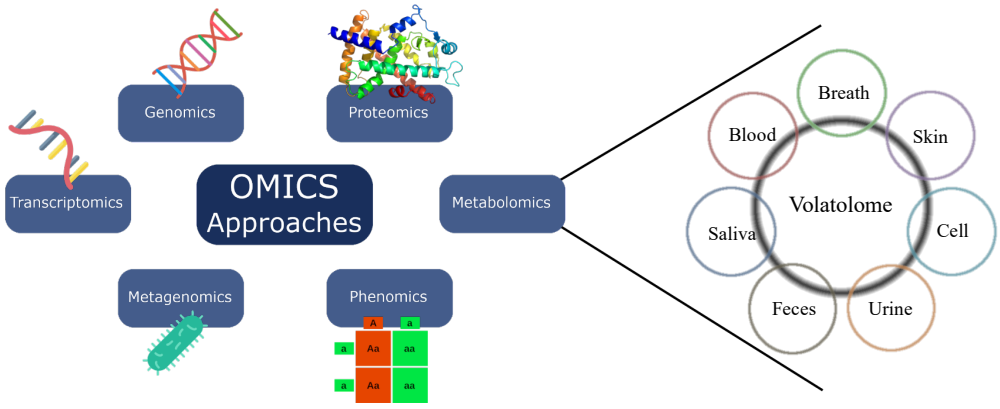


Figure 3: Overview of omics and volatolomics approaches. *Adapted from [20] and [21]*

1.3 Gas chromatography for metabolomics profiling

Metabolomic profiling is often performed using H^1-NMR or mass spectrometry coupled with chromatographic separation method[18], [22]. Particularly, comprehensive two-dimensional gas chromatography (GC×GC) exhibits promising outcomes, as seen in the realm of IDB screening for patients[23]. As Figure 3 shows, metabolomics and especially volatolomics can be performed on a variety of samples, ranging from breath to cultured cells. Different techniques are used to collect or enhance the ability to measure metabolites from the samples, such as derivatization or use of sorptive materials[18]. The present work aims to use volatolomics for the analysis of culture gastrointestinal cells using comprehensive two-dimensional gas chromatography.

1.4 Foundational research

This work builds upon the research conducted by Zanella *et al.*[20] concerning the impact of chemical and biologically induced inflammation on the volatile metabolite production of lung epithelial cells (A549) using GC×GC. The team of researchers investigated the viability and emitted VOCs by exposing *in vitro* A549 epithelial cells to a chemical oxidant, H_2O_2, at concentrations ranging from 0.1 mM to 50 mM, for varying exposure durations of 10, 30, 60 and 120 minutes. Additionally, biological oxidation was conducted *in vitro* using an inflammatory sputum mixture, exposing the same cell line for durations of 1, 3, 6 and 24 hours. However, this particular aspect of the research will not be discussed in this section. Controls were also prepared for both conditions to minimize the influence of variables stemming from the reagents employed.

Figure 4 presents the relative mitochondrial activity observed for different concentrations of H_2O_2 and various exposure times. As mitochondrial activity is linked to cell viability, A549 viability decreased with increasing concentration and duration of exposure.

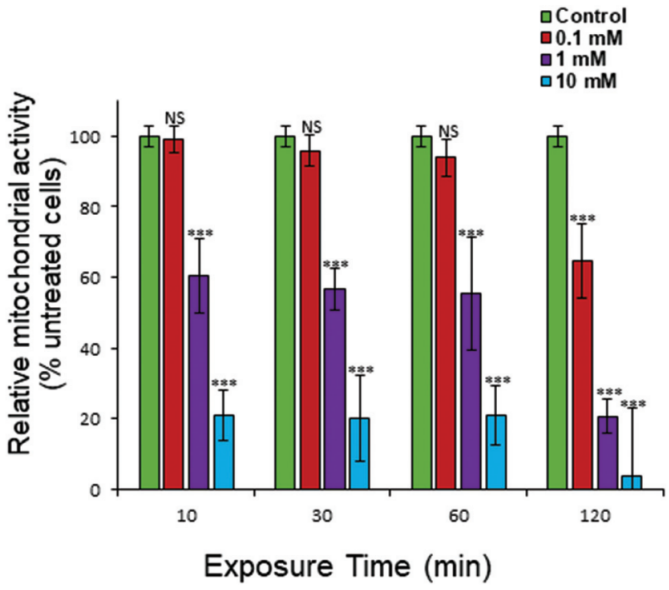


Figure 4: Mitochondrial activity of A549 cells treated with different concentration of H_2O_2. NS is non significant. *Adapted from [20]*

The findings of the study revealed that low exposure durations and concentrations of H_2O_2 did not exhibit any discernible impact on cell viability. For instance, cells exposed to 0.1 mM H_2O_2 for 1h exhibit a viability of 94%. Conversely, treatment of the cell line with a higher concentration of H_2O_2 (10 and 50 mM) resulted in significant cell death. Based on these observations, the optimal exposure time and concentration for the study were determined to be 1h and 0.1 mM H_2O_2 respectively.

Furthermore, the team aimed to demonstrate the statistical dissimilarity between the volatile fingerprints emitted by A549 cells exposed to 1 mM H_2O_2 for 1h and those from non-exposed or non-viable cells. To achieve this, principal component analysis (PCA) was conducted using VOCs emitted by non-exposed A549 cells as well as those exposed to concentrations of 0.1, 1 and 50 mM H_2O_2. The resulting PCA plot is illustrated in Figure 5.

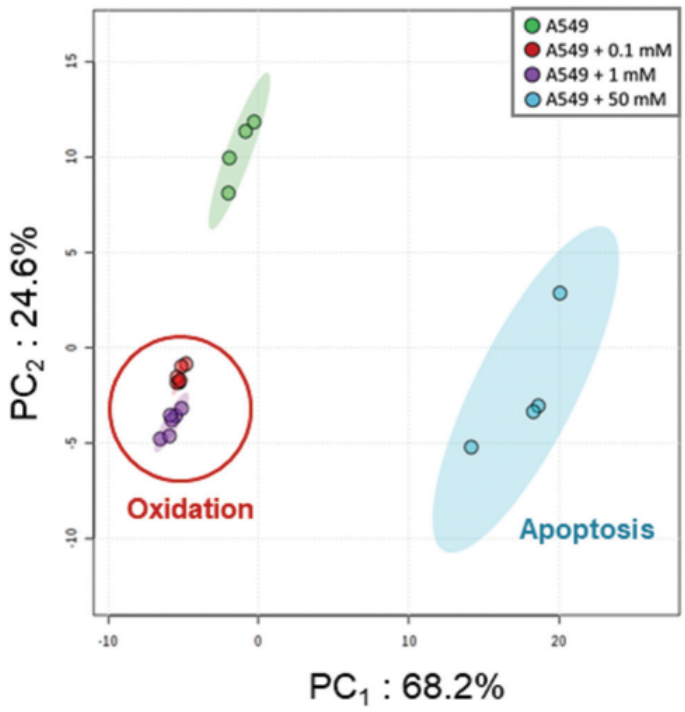


Figure 5: Principal component analysis of A549 cells exposed to different concentration of H_2O_2 using the 103 significant features highlighted by one way ANOVA. *Adapted from [20]*

PCA revealed the ability to distinguish 3 distinct groups based on the first two principal components, which collectively accounted for 92,8% of the total variance. The clustering indicated that chemical oxidation induced a metabolic alteration in the cells following exposure to H_2O_2, exhibiting a significant differentiation from cell apoptosis.

Table 1 presents an overview of the experimental conditions employed by Zanella *et al.*, with the use of medium controls to quantify the effect of this condition and the PBS controls to mimic the dilution effect caused by the addition of H_2O_2 on the cell line. Considering the results obtained, which identified the optimal concentration of H_2O_2 as 1 mM with an exposure time of 1h, leading to a discernible metabolic change without inducing cell apoptosis in comparison to non-exposed cells, this study was conducted under identical conditions.

Sample	Type	VOC source
Growth medium samples	Absolute control	Medium
	Negative control	Medium + H_2O_2
	Negative control	medium + PBS
Cell containing samples	Positive control	Cell line + medium
	Chemical oxidation	Cell line + H_2O_2
	Positive control	Cell line + PBS

Table 1: Description of the experimental conditions for cell oxidation. *Adapted from [20]*

1.5 Cell cultures

To develop a comprehensive model for inflammation, it is imperative to carefully select appropriate conditions and a suitable medium. In the present study, three epithelial cell lines, namely HCT116, Caco-2, and HT-29 cells were utilized.

HCT116 cells are human colon cancer cell line that carries a specific mutation in codon 13 of the ras proto-oncogene[24]. This genetic alteration results in a mutated variant of the Ras protein, which plays a pivotal role in the intercellular chemical exchange and cellular proliferation[25]. The Ras protein is

classified as an oncoprotein, and the corresponding gene is referred to as an oncogene. By definition, an oncogene possesses the potential to initiate cancer development. Activation of this specific gene and subsequent mutation of the expressed protein can lead to uncontrolled cell proliferation and interfere with the regulated process of apoptosis, resulting in the transformation of normal cells into cancerous cells[26].

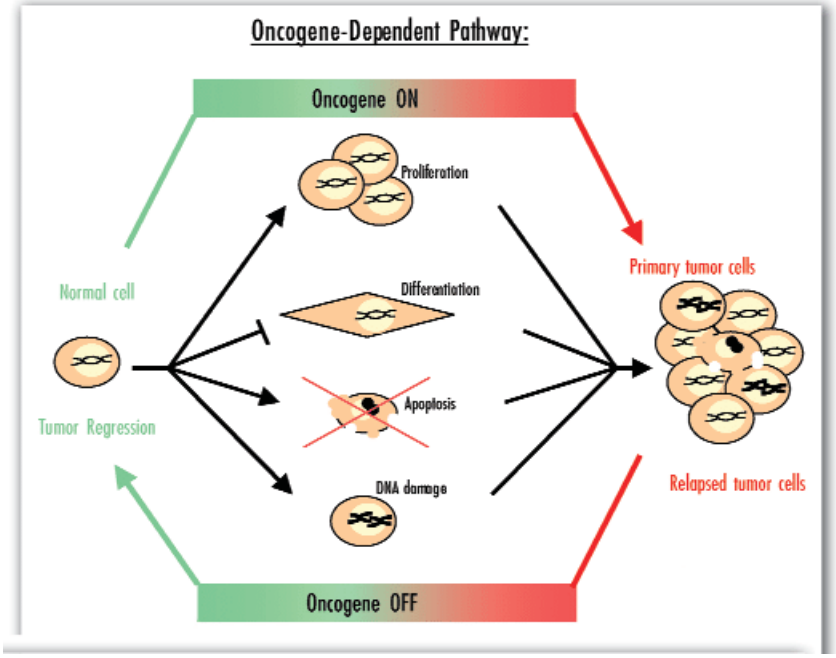


Figure 6: Mechanisms underlying cancer cells ability to evade reliance on oncogenes. Oncogene activation contributes to tumorigenesis through multiple mechanisms including: inducing autonomous proliferation, blocking differentiation, and inducing genomic destabilization. Oncogene inactivation appears to be sufficient to at least initially induce tumor regression[27]

.

Caco-2 cells are epithelial cells derived from a 72-year-old Caucasian male diagnosed with colorectal adenocarcinoma, a form of cancer that primarily affects epithelial cells of the colon. This particular cell line offers compelling features for studying CRC. Despite originating from the colon, these cells exhibit the ability to differentiate under specific conditions, resembling the phenotypic and characteristic attributes of the small intestine[28].

Finally, the HT29 cell line employed in this study was originally isolated in 1964 from a 44-year-old Caucasian female diagnosed with colorectal adenocarcinoma. Similar to the Caco-2 cell line, HT29 cells exhibit differentiating capabilities, making them valuable for establishing a model for inflammation of the gastrointestinal tract[29].

By utilizing these cell lines with distinct genetic characteristics, this research aims to establish a robust model for studying metabolomics profiling of inflammatory gastrointestinal cells.

1.6 Objectives of the study

As highlighted in section 1.4, the study successfully determined the optimal conditions for the chemical oxidation of A549 cells and identified potential biomarkers of this oxidation. However, certain questions still need to be addressed, including the applicability of the same conditions to different cell lines and whether the VOCs fingerprint is specific to a particular cell line or common across different cells. Therefore, the primary objective of this research is to provide potential insights and answers to these inquiries.

2 Theoretical concept

Effectively separating a complex gas mixture becomes a demanding undertaking, particularly when confronted with a large number of components within the mixture[30]. Numerous techniques exist for achieving this separation, with GC emerging as an efficient and widely employed method[30], [31]. Notably, bi-dimensional gas chromatography (GC×GC) has garnered significant attention due to its enhanced capabilities[31]. To comprehend the principles underlying GC×GC, it is imperative to revisit the fundamental operating principles of GC.

2.1 Basis of GC

The process of separating the components within a mixture involves introducing it into a mobile phase through a capillary column coated with a stationary phase. This separation is achieved based on the distinct affinity of individual compounds towards either the stationary or mobile phase, resulting in the effective separation of the mixture[32].

2.1.1 Theoretical model

Gas chromatography is founded upon the theoretical model proposed by A.J.P. Martin and R.L.M Synge[33], [34]. They introduced the concept of height equivalent to a theoretical plate (HETP), denoted as H, as well as the number of theoretical plates (N). The relationship between these parameters can be expressed as follows[33], [35]:

$$N = \frac{L}{H} \tag{1}$$

Here, L represents the length of the column. As chromatographic peaks typically exhibit a Gaussian distribution, the HETP can also be defined as[33]:

$$H = \frac{\sigma^2}{L} \tag{2}$$

In the above equation, σ represents the standard deviation of a Gaussian function

These equations serve as appropriate models for describing the theoretical aspects of GC. In practical experiments, the number of plates and, consequently, the column's efficiency can be determined using the equation[33], [35]:

$$N = 16(\frac{t_R}{W})^2 \tag{3}$$

In this equation, t_R, corresponds to the retention time of a compound, while W represents the width of the peak at its basis. By obtaining these 2 parameters, it becomes feasible to calculate the efficiency of a given column.

2.1.2 Practical approach

A schematic representation of the instrumental setup employed for GC-MS analysis is depicted in Figure 7. The sample is introduced into the system via an injector using a syringe, while a carrier gas serves as the mobile phase. The selection and optimization of the column, which is coated with a thin layer of the stationary phase, are crucial for achieving optimal separation conditions. During the passage of the sample through the column, the different components are separated based on their affinity for the stationary phase, as characterized by the affinity constant K[32]. The effluent from the column is subsequently collected and subject to analysis by a detector.

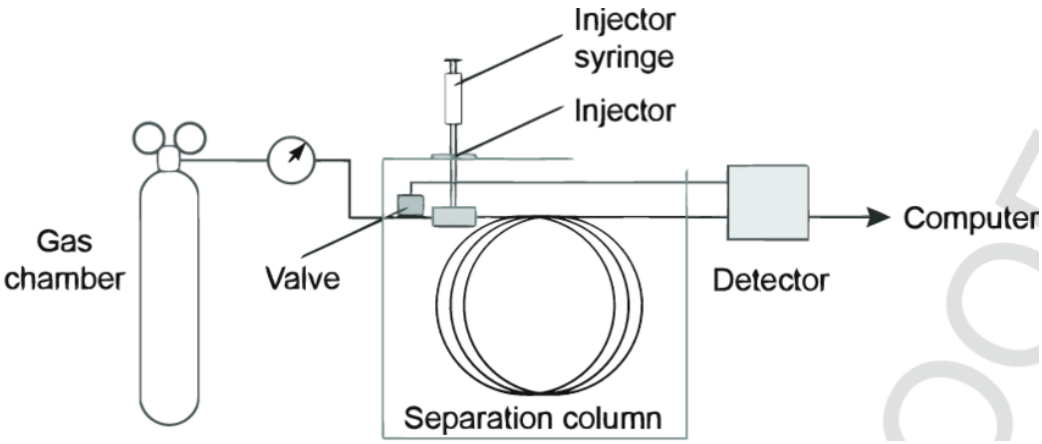


Figure 7: General scheme of GC-MS device[36]

2.1.3 Parameters that influenced retention time

There exist 7 significant parameters that have the potential to profoundly impact the quality of separation as well as the elution time required for a compound. These parameters include:

- The specific type of stationary phase[37]
- The diameter of the column[38]
- The length of the column[39]
- The thickness of the stationary phase[40]
- The temperature of the column[41]
- The type of carrier gas employed[42]
- The velocity of the carrier gas[42]

Modifying any of these parameters, either individually or in combination, can lead to enhanced resolution and faster separation. All these parameters are capable of influencing the variables present within the following equation[33]:

$$R = \frac{\sqrt{N}}{4}(\frac{\alpha - 1}{\alpha})(\frac{k_B}{1 + k_B}) \tag{4}$$

Here, R signifies the resolution of the column, N represents the number of theoretical plates, α denotes the separation factor, and k_B represents the retention factor.

2.1.4 Detector

The subsequent phase following sample separation involves the identification and detection of the array of analytes present within the mixture. Numerous ionization and detection methods are readily accessible in the commercial realm. The most prevailing method employed is time-of-flight mass spectrometry (TOF-MS). This technique facilitates the detection of compounds encompassing diverse ranges of mass-to-charge ratio (m/z). The utilization of TOF-MS offers several advantages, including the capability to detect an extensive spectrum of compounds, virtually without limitation.

Additionally, it presents a reasonable cost, moderate to high resolution, and rapid data acquisition speed, rendering it compatible with expeditious chromatographic methodologies[31], [43], [44].In this particular study, electron ionization coupled with TOF-MS is employed.

2.2 Bi-dimensional GC

Transitioning from one-dimensional GC to two-dimensional GC entails distinct advantages and disadvantages. The following section will concentrate on elucidating the fundamental principle underlying GC×GC, as well as highlighting a pivotal component that endows this technique with its robust analytical capabilities.

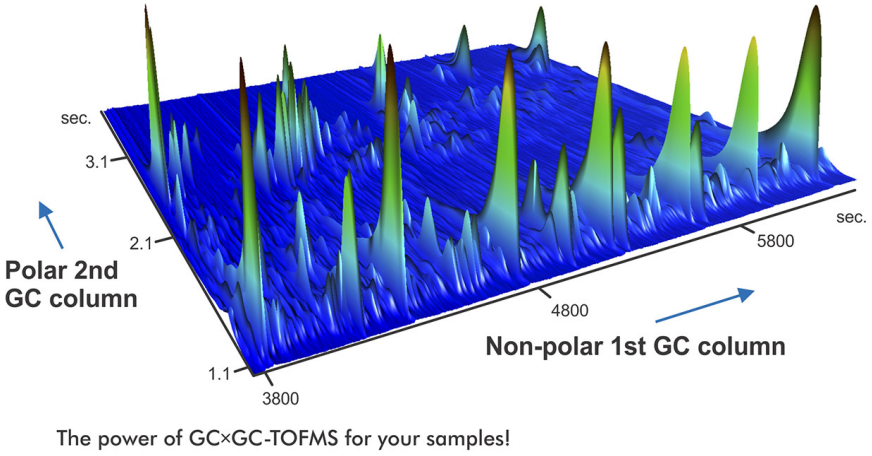


Figure 8: Example of a GC×GC chart in 3D for a petroleum analysis [45]

2.2.1 Basic principle

The analysis of complex mixtures often presents challenges when attempting to resolve coeluted compounds. While introducing a second column to enhance general resolution may seem like a viable solution, the associated technical complexities and relatively modest improvement in peak capacity render this approach less significant[31]. Instead, the optimal strategy lies in incorporating a second orthogonal dimension using a secondary column that segregates the entire sample based on a second parameter. This methodology

is known as multi-dimensional GC (GC×GC). The key distinction between one-dimensional GC and GC×GC lies in the addition of a second column and a modulator capable of separating the mixture based on two orthogonal parameters. The first column, typically a lengthy one, separates the compounds according to the first parameter. The subsequent second column, shorter in length, further resolves the refocused mixture based on a second parameter. The defining component of GC×GC analysis is the modulator, which plays a crucial role[31]. Ultimately, the resulting data matrix can be represented in 3D format, as depicted in Figure 8. Chromatograms are commonly represented in 2D, as presented in Figure 9.

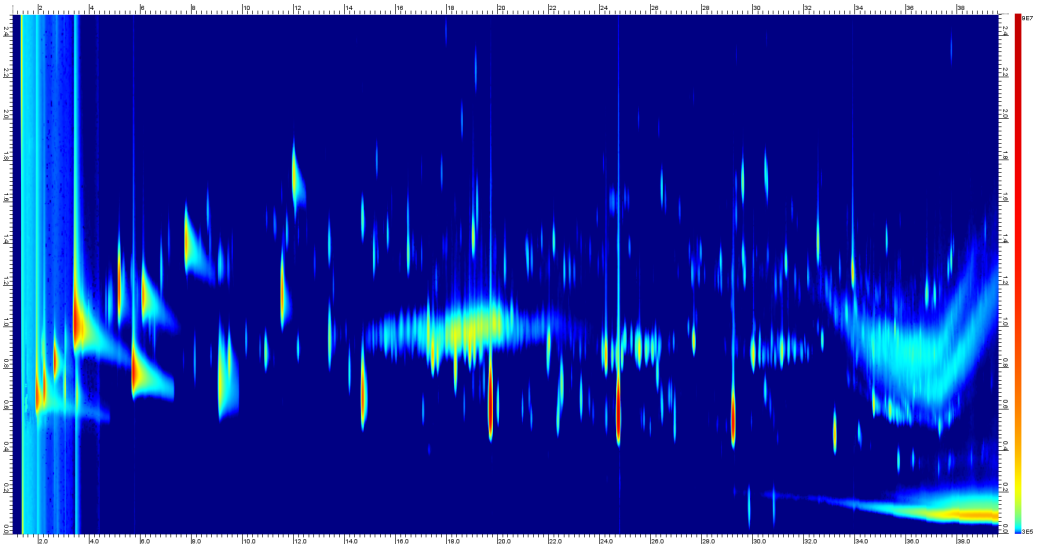


Figure 9: Illustration of a GC×GC chromatogram in 2D for metabolomic analysis. The horizontal axis represents the retention time in the first column, while the vertical axis represents the retention time in the second column. The color code, ranging from green to red, indicates the intensity of the detected compounds.

2.2.2 The modulator

An essential component of multi-dimensional chromatography, often described as the "heart" of this technique, is the modulator[46]. This piece of equipment plays a critical role in transferring the emerging compounds from the first dimension (^{1}D) to the second dimension (^{2}D) while preserving the separation achieved in the first dimension. Moreover, the primary objective of a modulator is to focus the effluents from the first dimension and subsequently introduce them into the second dimension as narrow pulses[31], [46], as depicted in Figure 10.

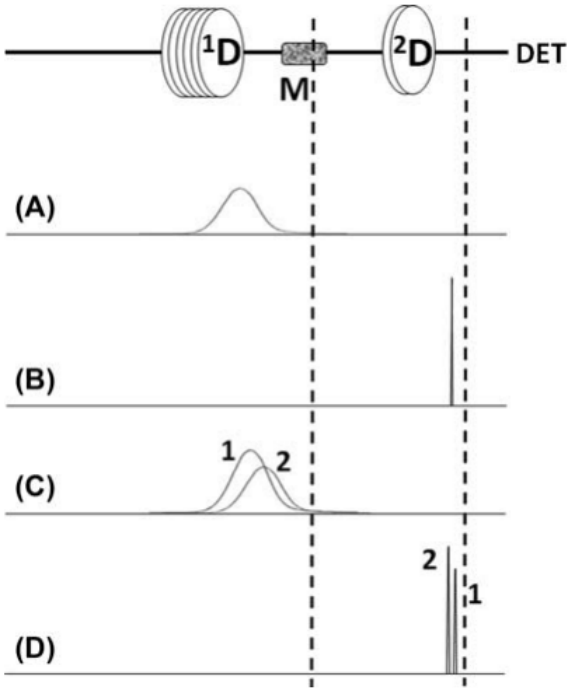


Figure 10: General scheme of a modulator. (A) A first-dimension (1D) column elutes a compound as a regular elution broadened peak that elutes from the end of the column. (B) If a modulation device completely collects the emerging compound from 1D and delivers it to a short 2D column, then the result is a much taller but narrower peak at the end of the 2D column. (C) Two overlapping compounds produce an unresolved peak at the end of 1D. (D) The overlapping compounds in (C) may then be resolved in 2D provided the inlet band is very narrow, and the stationary phase allows them to be resolved. *Figure from [31].*

There are three primary categories of modulators commonly used in GC×GC

analyses: thermal, valve, and flow-based modulators. Thermal modulation can be further classified into two subgroups: heater-based and cryogenic modulation. The latter variant was first introduced by Marriott and Kinghorn in 1997[47]. Nowadays, the dual-stage quad jet modulator stands as the most widely adopted modulator in this field.

Dual-stage quad jet modulator, introduced by Edward B. Ledford in 2000[48], consists of a two-stage modulator tube, where modulation is achieved using alternating cold and heat jets, obviating the need for any moving parts[49]. The modulation sequence of events is depicted in Figure 11.

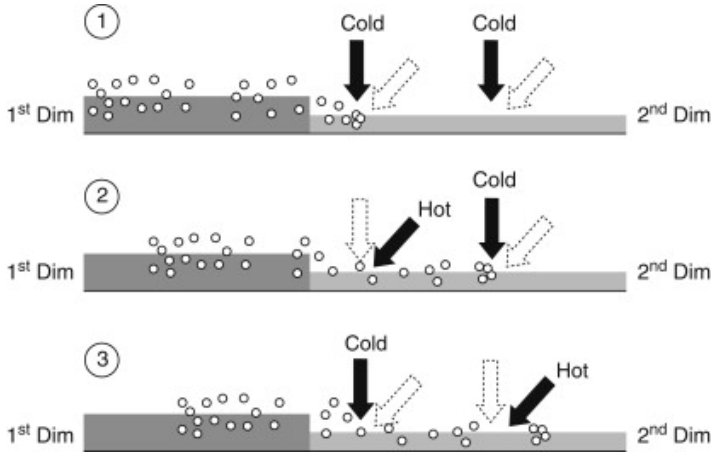


Figure 11: The sequence of events responsible for (1) trapping, (2) releasing and refocusing, and (3) re-injecting into the second column using a quad-jet cryo-modulator. *Figure from [50].*

The dual-stage quad jet modulator offers several advantages, including:

- No moving parts, resulting in reduced maintenance requirements
- Widely available commercially
- Expanded temperature range for analysis compared to thermal modulation[49]

Nonetheless, one drawback of thermal modulation is the substantial consumption of consumables such as liquid nitrogen.

2.3 Sampling and extraction

To carry out gas chromatography, the analyzed mixture must enter the vapor phase before being introduced into the column. Since samples can exist in various forms such as solid, liquid, or gas, and may be either diluted or concentrated, preliminary sampling methods must be employed. These methods include extracting the desired compounds, preparing the sample for GC analysis, and establishing the appropriate analytical concentration range of the compounds before injection into the GC instrument. Several extraction methods are commonly utilized, such as Solid Phase MicroExtraction (SPME), Stir Bar Sorptive Extraction (SBSE), and Liquid-Liquid Extraction (LLE)[31]. The choice of extraction method depends on the physical and chemical properties of the sample under investigation. For the present study, headspace extraction using SPME was employed.

2.3.1 Headspace extraction

Headspace (HS) extraction is a widely employed technique for the analysis of cells, making it particularly suitable for liquid or solid samples analysis[31], [51]. The fundamental principle underlying HS extraction involves shifting the equilibrium towards the gaseous phase, as depicted in equation 5[52]. This shift is achieved by controlling various parameters, such as temperature or the induction of salt-induced precipitation. By manipulating the parameters K or β, it becomes possible to enhance the concentration of the sample in the gaseous phase.

$$C_g = \frac{C_o}{K + \beta} \tag{5}$$

Here, C_g represents the concentration of the compound in the gaseous phase, C_0 represents the concentration of the compound in the original sample before analysis, K is the partition coefficient, and β the phase ratio (ratio of two phases volume).

Two main categories of headspace extraction techniques are commonly found[31], [52]: static headspace extraction and dynamic headspace extraction. The distinction between these categories is illustrated in Figure 12. In static headspace extraction, the VOCs present in the sample are either directly injected into the GC analyzer or collected onto a sorptive layer for subsequent

analysis. On the other hand, dynamic headspace extraction involves the continuous purging of analytes using an inert gas, with subsequent trapping on a sorbent trap. This dynamic extraction approach offers improved sensitivity compared to static headspace extraction, as it ensures a continuous displacement of analytes towards the vapor phase, facilitating their capture.

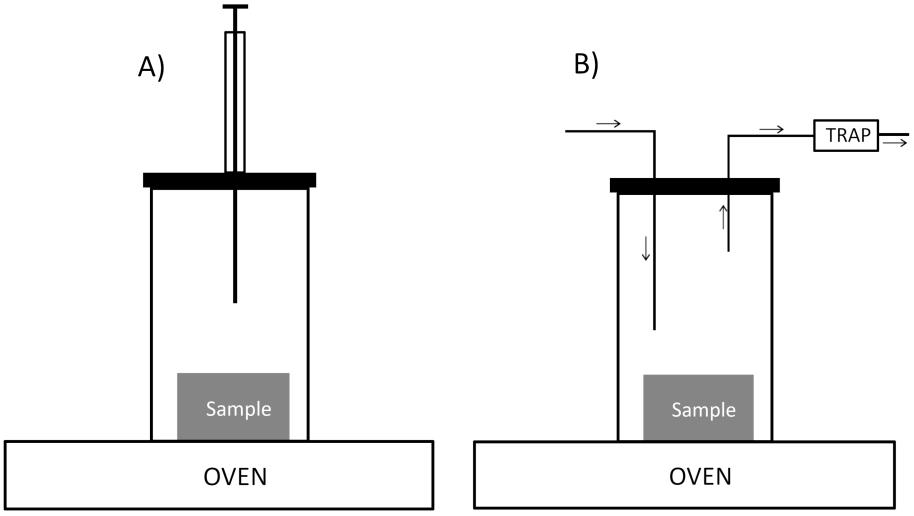


Figure 12: Difference between static headspace extraction and dynamic headspace extraction. A) Static headspace and B) Dynamic headspace. Fiber is shown in green color. The arrow represents the flow of vector gas.

2.3.2 Solid Phase Microextraction (SPME)

Solid phase microextraction is a widely employed technique utilized to enhance the sensitivity of static headspace analyses, allowing for concentration and extraction of target analytes[31]. in Figure 12, the implementation of SPME is depicted, wherein a sorptive layer, typically in the form of a syringe equipped with a coated fiber, is introduced in the headspace environment. This introduces a secondary equilibrium between the sorptive layer and the VOCs present in the gas phase. Afterward, the fiber is thermally desorbed,

releasing the capture analytes into the injection port of the GC instrument. Various fibers with distinct characteristics are available, offering flexibility in performing different (or the same) experimental setups.

2.4 Data analysis

The primary challenge associated with GC×GC-MS does not reside in the experimental procedures but rather centers around the handling and analysis of the substantial amount of data generated throughout an experiment[53]. Due to the extensive data collection involved, researchers face the challenge of developing effective strategies to interpret and present the resulting outcomes in a meaningful manner.
Metabolomics can be categorized into two distinct approaches: targeted and untargeted. The targeted approach primarily focuses on the detection and comparison of specific metabolites with well-established references, the second one aims to analyze all detectable metabolites without emphasizing any particular ones. While targeted analysis confirms a hypothesis, untargeted analyses seek to generate one. Consequently, the major limitation of the untargeted approach lies in the identification of unknown compounds[54]. For this study, the untargeted method was chosen to maximize the acquisition of information concerning the oxidation of in vitro cells.

2.4.1 Data pre-processing

The raw data generated by the GC×GC-MS instrument cannot be utilized directly and necessitate pre-processing steps before extracting valuable information for software applications such as Metaboanalyst.

GC Image^{TM} is a software tool readily available for visualization, processing, and analyzing data obtained from GC×GC. Figure 13 depicts the workflow used to process the data. Each pixel within the image corresponds to a specific data point, with the horizontal axis representing the retention time of the chemical in the first column, while the vertical axis represents the retention time of the chemical in the second column. The color code assigned to each blob (cluster of data-point pixels) indicated the quantity of the detected chemical, while the background is uniformly colorized in light blue[55]. Although performing untargeted metabolomics analysis on each image can be time-consuming, GC Image enables the concurrent analysis of multiple

samples by generating a template that encompasses the automated features detected in the samples by the software. This obtained template, which constitutes a cumulative image incorporating information from all the samples, is subsequently applied to each individual image after removing artifacts or well-known compounds unrelated to the experiment.

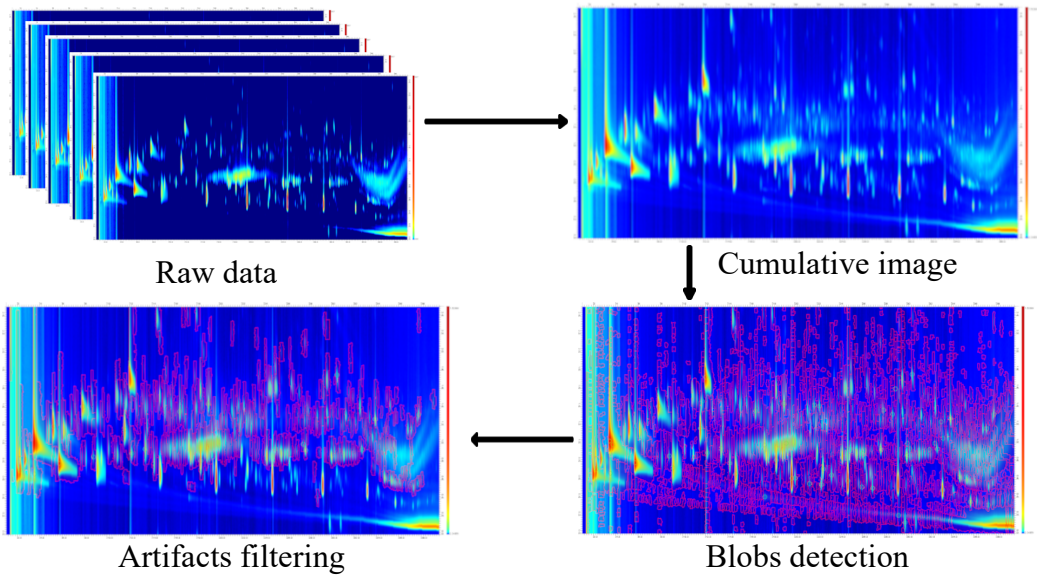


Figure 13: General workflow employed for data processing. A cumulative image is constructed using all the 1150 detected peaks, and subsequently used to create a template by automatically detecting pixel blobs (clusters of data point pixels). This template is then filtered to remove artifacts and non-relevant chemical compounds. The resulting filtered template can be applied to each of the obtained chromatograms. The number of red circles in the lower part of the figure indicates detected compounds by GC Image. This number is reduced after filtration of known compounds such as artifacts from analysis, as indicated from the right lower chromatogram to the lower left chromatogram.

2.4.2 Data normalization

Data obtained from metabolomics experiments often yield large data sets, encompassing various factors such as technical errors, samples or biological variations. These factors can pose a challenge in the identification of specific compounds that can be used as biomarkers of inflammation. To address this issue, data normalization is employed as a method to minimize unwanted variations and is an integral part of statistical analysis in metabolomics[56].

The choice of normalization method has been studied to assess its impact on data analysis[57], [58]. However, it is essential to carefully the appropriate normalization method based on the experimental procedure employed. One commonly used normalization method used is Probabilistic Quotient Normalization (PQN). This method involves determining the most probable quotient between the signals of a spectrum and signals of reference samples. Each measured signal is divided by the reference signal value (or the median value for a group of references). The median value of all obtained quotients is then computed and utilized as a normalization factor[59].

2.4.3 One-way analysis of variance

One-way Analysis of Variance (ANOVA) is a statistical technique employed to determine the statistical significance of differences among sample means, using the F distribution[60]. ANOVA operates, as the name suggests, by analysis of variances between groups by computing an F ratio as a measure, which can be defined using the following equation[60]:

$$F = \frac{Intergroup - variance}{Intragroup - variance} = \frac{\sum_{i=1}^{K} n_i(\overline{X_i} - \overline{X})^2/K - 1}{\sum_{ij=1}^{n}(X_{ij} - \overline{X_i})^2/N - K} \tag{6}$$

Here, n_i is the number of observations in the group; $\overline{X_i}$ is the mean of the group; $\overline{X}$ is the overall mean; K the number of groups; X_{ij} is the jth observational value of group i; N the number of all observational values (number of experiments).

The computed F ratio is compared to an F critical value given for an F distribution, depending on the given values of N and K described in equation 6. If the computed value exceeds the F critical value, it signifies a significance level of 0.05, leading to the rejection of the null hypothesis and indicating that the means of the different groups are indeed different. However, it is important to note that this test does not specify which means differ from one another, necessitating the use of a *post-hoc* test to provide such information. Employing ANOVA helps streamline the analysis of the collected data by excluding compounds that do not yield an F value higher than the F critical value. Consequently, ANOVA is employed in this work to reduce the complexity of the collected data.

2.4.4 Principal component analysis

Principal component analysis (PCA) is a data analysis technique employed to effectively reduce the dimensionality of a large data set while retaining the maximum amount of relevant information[61]. This is achieved through several steps, including data normalization, computation of the covariance matrix, determination of principal components, and visualizing of the transformed data)[62]. By utilizing PCA, the acquired data can be condensed and represented in a visually interpretable manner. The resulting PCA plot, as depicted in Figure 14, serves as a concise representation of the final outcomes. The horizontal axis encompassing the first principal component account for 52.5% of the total variance whereas the second principal components account for 29.9%.

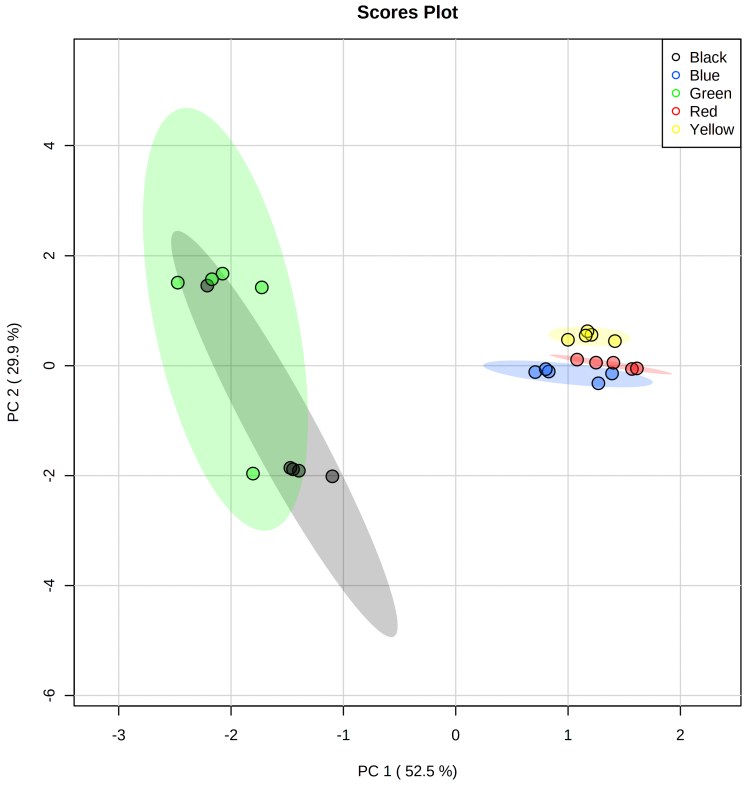


Figure 14: Example of a PCA plot obtained for the analysis of different types of beers. Two clusters can be observed, illustrating the use of PCA to show groups within an experiment.

3 Methods and experiment

To conduct metabolomic profiling of the model inflammatory cells, the three cell lines were culture and analyse using GC×GC instrument. The collected data were then subjected to statistical analysis. Figure 15 shows the analytical workflow performed for this experiment. The subsequent section provides a detailed account of the materials and methods employed for this study.

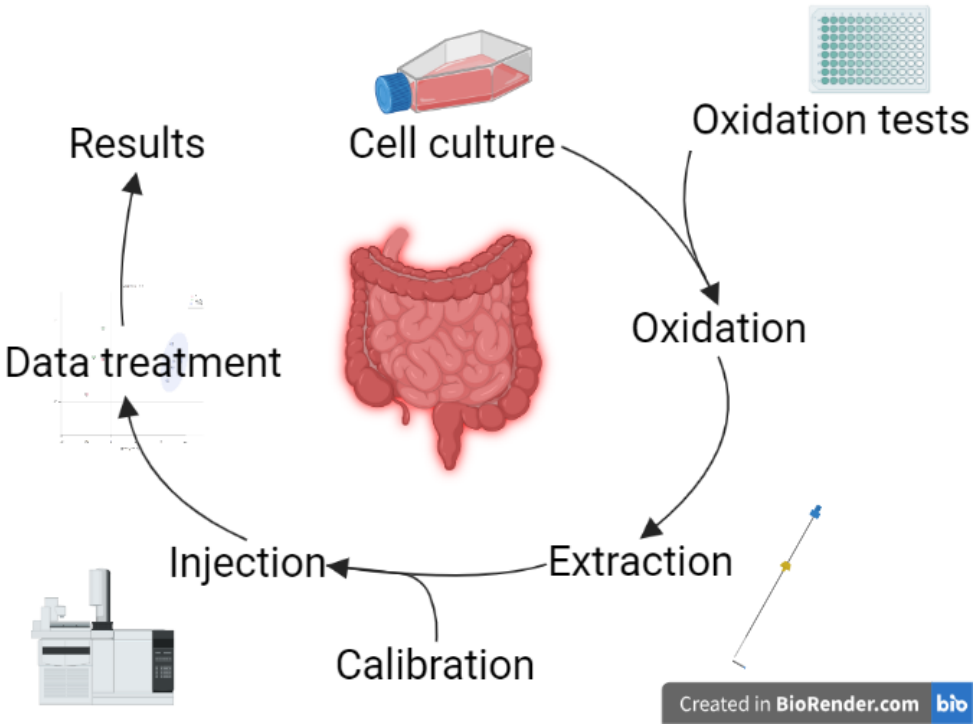


Figure 15: Analytical workflow of the experiment

3.1 Cell cultures and oxidation

3.1.1 Cell cultures

All the cell lines used in this study were initially purchased from ATCC^{TM}: HCT116 (ATCC CCL-247); Caco-2 (ATCC HTB-37); HT29 (ATCC HTB-38). However, due to unknown reasons, the purchased Caco-2 and HT29

cell lines failed to grow under designated culture conditions. Consequently, samples for these two cell lines were borrowed from the GIGA (University of Liège, Belgium).

Culture of the Caco-2, HCT116 and HT-29 cell lines was performed using a mixture composed of 500 mL of Dulbecco's Modified Eagle Medium (DMEM, Gibco), 55 mL of FBS, 5.5 mL of PenStrep (Penicillin-Streptomycin) and 5.5 mL of Ampho B (Amphotericin B). This mix is referred to as "complete media". Cells were cultivated in 25 cm^3 flasks, incubated at 37°C in 5% CO_2 until reaching 80% confluency.

The cell lines were cultured under sterile conditions, and each cell line was provided with its specific complete media prepared using the method described in this section. To initiate the culture, cells were removed from the nitrogen bath and thawed. Subsequently, they were transferred to a 50 mL Falcon® tube containing 10 mL of complete media. The Falcon® tube was then centrifuged at 200G for 5 minutes, and the supernatant was discarded. To the remaining pellet in the Falcon®, 2 mL of fresh complete media was added. The total volume was divided into four batches, with each batch transferred to a sterile T25 flask containing 7.5 mL of fresh complete media. Flasks were stored in a 37°C incubator with a 5% CO_2 atmosphere. They were kept in the incubator until 80% confluency is reached.

Once the cells reached 80% confluency, cell subculture was performed using all the newly grown cells. The supernatant in the flask was discarded and the T25 flasks were washed with 3 mL of calcium- and magnesium-free Dulbecco's phosphate-buffered saline (DPBS). DPBS was used to prevent osmotic shock to the cells and is buffered at an appropriate pH for cellular culture. Afterward, the liquid is discarded and 1 mL of trypsin was added to detach the cells from the flask walls. This process, known as trypsinization, is inhibited by calcium or magnesium. Therefore, all previous steps must avoid contamination of these two elements. Flasks were then incubated at 37°C and 5% CO_2 for 5 min. Once it was confirmed under a microscope that the cells were detached from the flask, 4 mL of fresh complete media was added to the flask to inhibit the trypsin activity. Careful rinsing of the flask walls was performed several times to collect all the cells. The liquid containing the cells was then transferred to a 50 mL Falcon® tube, centrifuged at 200G for 5 min, and the supernatant was discarded. To the cellular pellet, 1 mL

of fresh complete media was added and the total volume was divided into 4 batches to prepare additional T25 flasks.

3.1.2 Viability test

Before conducting the oxidation, a viability test was performed on HCT116 cells using chemical oxidation with H_2O_2 in a 96-well plate. This step is crucial to assess whether the experimental conditions would induce necrosis in the cells. The newly grown HCT116 cells were counted and diluted to achieve a concentration of approximately 15.10^3 cells/well. The selection of reagents and concentrations was chosen according to section 1.4 [20].

The test involved exposing the HCT116 cells to a diluted solution of H_2O_2 (0.1 mM). The hydrogen peroxide solution was prepared by filtering 5 mL of a concentrated 30% in mass solution, mixed with 495 mL of sterile water, resulting in a concentration of 97 mM. A solution of 1 mM was then prepared by diluting the 97 mM solution. Additionally, a 500 mL bottle of DMEM mixed with 5 mL of PenStrep and 50 mL of FBS was prepared, referred to as "complete media".

To carry out the experimental procedure, cultured HCT116 cells were diluted and seeded onto the 96-well plate until 60% confluency was reached. Cells were then carefully washed with 200 μL of DPBS solution containing calcium and magnesium ions. After discarding the washing solution, the testing solutions were quickly added following the pattern shown in Figure 16. Solutions were added rapidly to prevent leaving the cells without any media. Cells were then incubated for exactly one hour at 37°C with 5% CO_2. After the incubation period, all the media were discarded and the wells were washed with 200 μL of DMEM F-12, an enriched sterile culture media. The media was again discarded and 120μL of an MTS solution was added. This tetrazolium salt was used to perform a colorimetric assay to assess cell viability. The plate was inspected under a microscope and incubated for one more hour at 37°C with 5% CO_2.

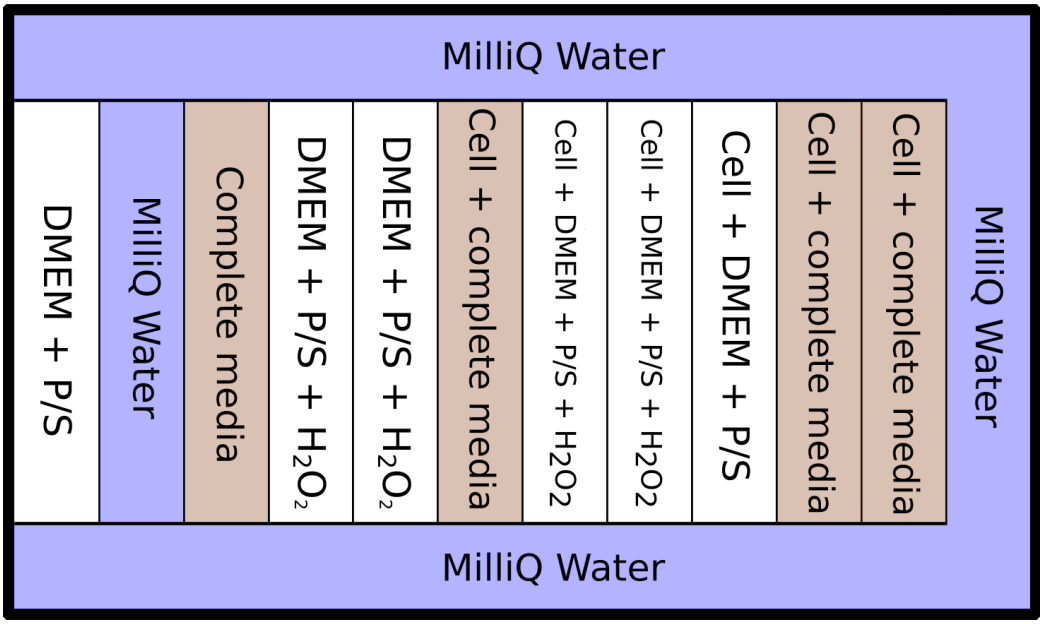


Figure 16: General representation of the 96 well-plate used for the oxidation tests performed on the HCT116 cell line. Complete media referred to a mix of DMEM, PenStrep (1%), and Fetal Bovine Serum (10%). Cell corresponds to HCT116 cell line. P/S corresponds to PenStrep. Blue cells correspond to water. Brown cells correspond to control with the presence of Fetal Bovine Serum. Cells with no coloration correspond to control without Fetal Bovine Serum.

3.1.3 Oxidation

Oxidation experiments were performed using the same protocol as described in section 1.4, utilizing the same reagent described in the previous section. Three cell conditions (oxidized cell with H_2O_2; control cell with media and PBS; control cell with complete media) in triplicate were prepared in T25 flasks. The media in all flasks was discarded, and the following reagents were added:

- In 3 flasks, 5 mL of DMEM and 1% of PenStrep were added
- In 3 flasks, 4.5 mL of DMEM and 1% of PenStrep were added, followed by 0.5 mL of PBS++
- In 3 flasks, 4.5 mL of DMEM and 1% of PenStrep were added, followed by 0.5 mL of H_2O_2 1 mM

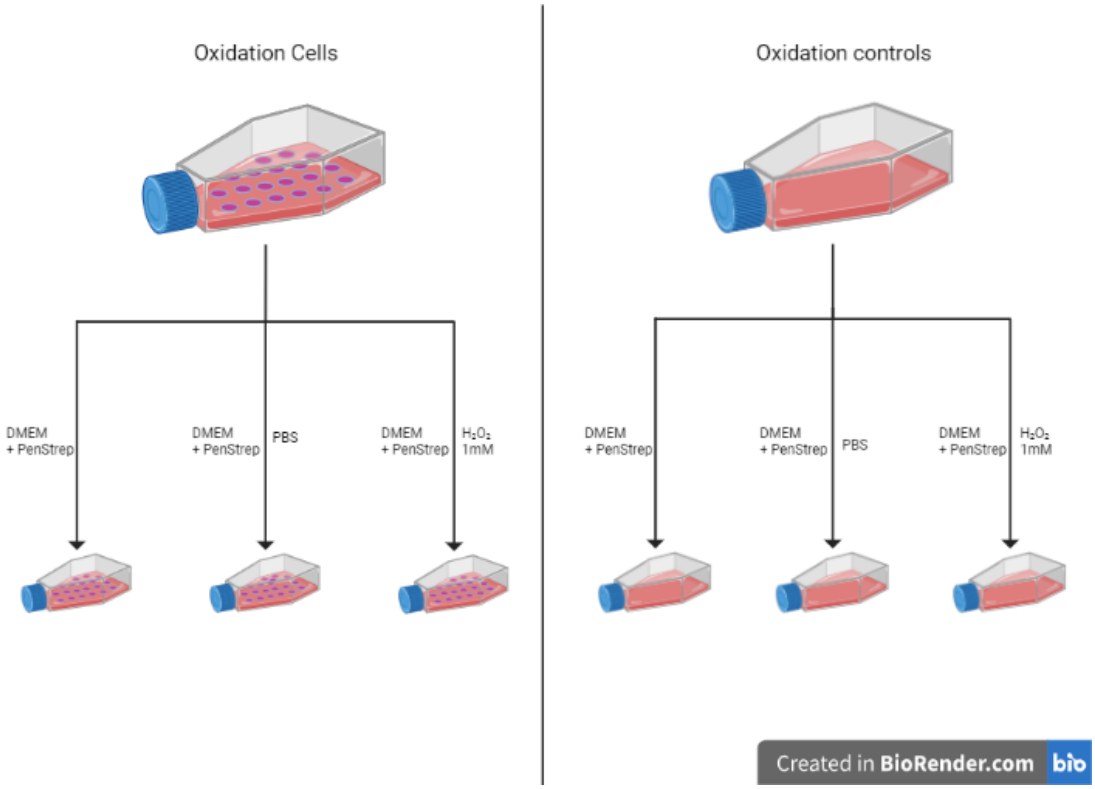


Figure 17: Experimental protocol of the oxidation. T25 flasks on the left of the figure are used for the oxidation of the cell lines. T25 flasks on the right are used for the preparation of control solutions.

The same pattern was used for the preparation of control solutions, where T25 flasks were used without any cells as absolute control conditions. After adding the reagents, all the flasks were placed in an incubator for one hour at 37°C with 5% CO_2. After the incubation period, all the solutions from the control flasks and the cells were transferred to separate vials. A quality control pool was prepared by mixing 1 mL of each collected solution. Additionally, global quality control was prepared by combining a small amount of each QC pool from each cell line. Figure 17 illustrates this experimental protocol. It is worth noting that each cell oxidation experimentation was consistently performed by the same operator, while a different operator was responsible for preparing the control samples. After the collection of all samples from cellular oxidation and controls, vials were stored at -20°C until injection.

3.2 GCxGC parameters

Table 2 lists all the parameters used for the experimentation.

Sampling parameters	Auto sampler model	LECO L-PAL3
	Fiber coating	Divinylbenzene/Carboxen /Polydimethylsiloxane
	SPME fiber color	Grey
	Agitation temperature (°C)	40
	Extraction time (min)	20
GC parameters	Instrument	Pegasus™BT 4D LECO
	Modulator type	Dual-stage quad jet cryogenic
	First column	Non polar Rxi-5Sil MS, 30m × 250 μm × 1 μm d_f
	Second column	Mid polar J& W VF-17MS, 2m × 250 μm × 0.5 μm d_f
	Modulation period (s)	2.5
	Helium rate (mL/min)	1.4
	Hot pulse time (s)	0.75
	Cool time between stage (s)	0.5
	Inlet split ratio	10
	Temperature ramp (°C/min)	5
MS method	Mass analyzer	TOF-MS
	Acquisition delay (s)	60
	Acquisition frequency (Hz)	200
	Analysis time (min)	47
	Ion source temperature (°C)	250
	Ionization energy (eV)	70
	Mass range (amu)	35 to 450

Table 2: Table of the experimental parameters for the GC×GC analysis

3.3 Calibration

To assess instrument stability, a mix of standards was injected into the instrument. The standard mixture consist of a mix of: n-hexane; 1-propanol/2-butanone; amylene hydrate; 1-ethyl-3-methyl-cyclopentane; 2-hexanone; 2,3-butanediol; p-xylene; cyclohexanone; decane; 6-methyl-5hepten-2-one; undecane; 1-octanol; acetophenone; nonanal; 2-ethylhexanoic acid; 2,6-dimethyl phenol; 2,6-dimethylaniline; methyl caprate; 1-tetradecene; methyl undecanoate; 1-pentadecene; dicyclohexylamine; 6,10-dimethyl-5,9-undecadien-2-one; methyl laurate. All vials contain 1 μL of the mix for a concentration of 1 ppm. After injection, standards retention times, forward and reverse match, probability, and area were inscribed in an Excel datasheet to build a quality control chart using 5 consecutive standard injections.
A mix of alkanes ranging from C8 to C20 was also injected to compute the Kovats retention index. The same parameters as described before for the standard injections were computed in an Excel datasheet.

3.4 Injection

All samples were randomly injected into the experimental device. To ensure quality control, after every 5 sample injections, an injection of either a blank vial or a mixture of alkanes or standards was performed. Throughout a period of 1 month, a total of 96 injections were conducted.

3.5 Data analysis parameters

All data were collected using ChromaTOF^{TM}(LECO) software and pre-processed using GC-Image^{TM} software version 2022R1. Following baseline correction, a feature template was generated by identifying features in all chromatograms using the software auto-detecting option. All artifacts arising from GC columns or well-known compounds were eliminated and the template was applied to all chromatograms to report the peak volume of each feature, along with their corresponding retention time and masses. Finally, a spreadsheet was generated, containing pertinent information such as peak volume and Fischer ratio. The statistical analyses were performed using MetaboAnalyst 5.0[63] and Microsoft Excel®. Data were normalized by PQN using Excel and transformed into log10 and auto-scaled using features provided by MetaboAnalyst.

4 Results

The ensuing chapter presents the results obtained for each step depicted in the analytical workflow (Figure 15) described in section 3. Before delving into the results of cellular oxidation, a description of a training experiment is provided in the subsequent section.

4.1 Fibroblasts analysis

Before conducting experiments on the various cell lines, extensive training and practice were undertaken to gain proficiency and become familiar with the GC×GC instrument and software. This was accomplished through the analysis of human lung fibroblasts (MRC-5) exposed to transforming growth factor and/or Pirfenidone. To assess the instrument's performance, a mixture of standards was injected to establish a quality control (QC) chart. The list of standards is described in section 3.3.

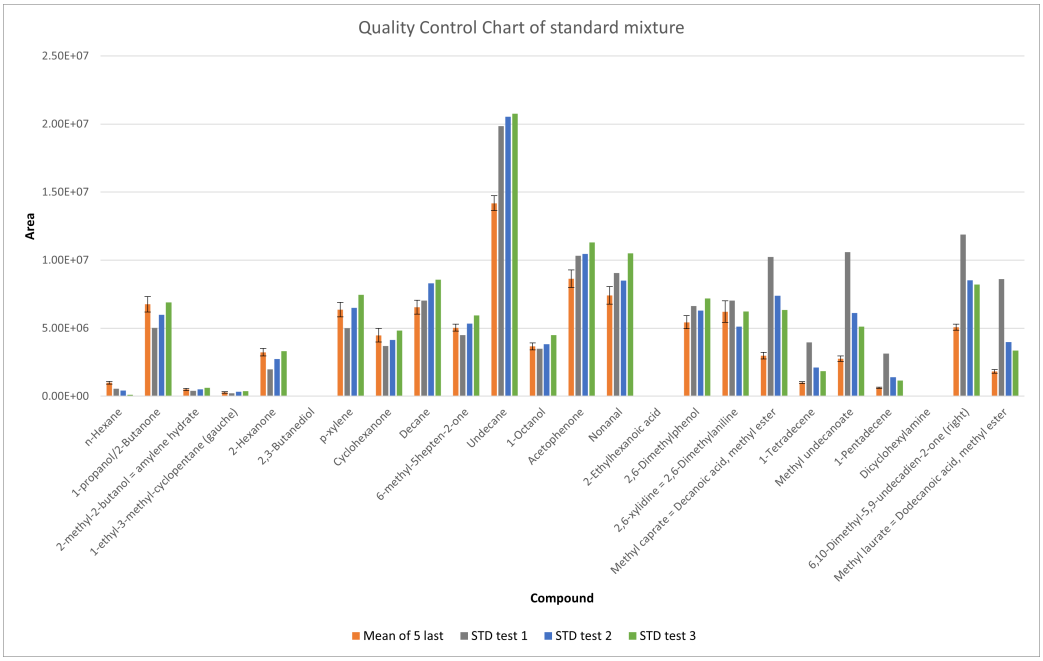


Figure 18: Measure standard areas for 5 consecutive standard injections used to built the QC chart and 3 standard injections. STD = Standard.

5 consecutive standard injections were selected to construct the QC chart.

Figure 18 compares the measured area between these 5 consecutive injections and 3 test injections. Utilizing the Relative Standard Deviation (RSD) and based on the results presented in Figure 18, various chemical compounds can be employed to create a QC chart and assess the instrument's stability.

4.2 Cell culture, viability test and oxidation

4.2.1 Culture troubleshooting

All cell lines were meticulously cultured, and the remaining culture cells were preserved by freezing and storing them in liquid nitrogen for potential future experiments. However, as mentioned in section 3.1.1, the cell lines were initially purchased from ATCC^{TM}. Only the HTC116 line demonstrates successful development under the designated culture conditions. Despite our best efforts, the Caco2 and HT29 cell lines did not flourish under the same conditions as the HCT116 cell line. Only a limited number of cells were observed under a microscope, even after multiple passages with fresh media. Consequently, to proceed with the study, the Caco2 and HT29 cell lines were borrowed from the GIGA (University of Liège, Belgium), as they had previously employed these cell lines in their research. Utilizing the identical culture conditions used for HCT116 cells, only the Caco2 cell line successfully develop. The HT29 cell line, on the other hand, failed to develop. Upon further investigations, the GIGA team discovers that improper storage conditions were utilized for this HT29 cell line, resulting in insufficient cell recovery. Recognizing this issue, the GIGA provided a new HT29 cell line for our experiment, which thrives under the specific culture conditions used for this work.

Investigation into the reasons for the inability of the purchased cell lines from ATCC^{TM} to develop was not conducted as part of this study. However, taking into account the similar issue encountered by the GIGA team, it is hypothesized that inappropriate storage conditions during transportation for the Caco2 and HT29 cell lines may have led to cellular death, resulting in insufficient cell recovery. Consequently, it is of utmost importance to exercise caution and ensure proper storage conditions when handling cellular lines in experimental studies.

4.2.2 Viability test

To evaluate cellular viability under the applied oxidation conditions, a colorimetric test was performed on the oxidized HCT116 cells. This particular test relies on the reduction of an MTS tetrazolium salt by the seeded cells, leading to the formation of a color compound that absorbs light at 490 nm. By employing Beer-Lambert's law, the absorbance was measured, providing a proportional indication of cell viability[64]. The oxidation conditions, as described in section 3.1.3, involve the use of a mixture comprising DMEM, PenStrep and H_2O_2 in conjunction with the cells. After subtracting the absorbance of the control samples, the measured viability was determined to be 90%. This observation serves to confirm that the applied conditions utilized in this work do not induce cellular necrosis or apoptosis.

4.2.3 Oxidation

Oxidation of the 3 cellular lines was going well and no specific troubles were encountered. All oxidation conditions, especially the time of exposition to the oxidation agent, were strictly respected.

4.3 Extraction, calibration and injections

4.3.1 Calibration

Figure 26 within the Appendices illustrates the measured area pertaining to 4 consecutive standard injections. Examination of this figure discerns that several compounds within the mixture can be used to build a QC chart. Notably, among these compounds, cyclohexanone, 2-butanone and 6,10-dimethyl-5,9-undecadien-2-one exhibit low RSD values, measuring 3, 3.1 and 3.4 respectively. Conversely, 8 compounds inherent to the mixture exhibit RSD values exceeding 10, encompassing 1-ethyl-3-methyl-cyclopentane, p-xylene, decane, 6-methyl-5hepten-2-one, 1-octanol, 2,6-xylidine, 1- pentadecene and dodecanoic acid, methyl ester. These compounds are not used for assessing the instrument's stability.

Figure 19 portrays a QC chart built upon the measured area for cyclohexanone, incorporating data from both 4 consecutive standards and standards injected during the analysis. Additional QC charts constructed through other chemical compounds are presented in the Appendices. Given the pronounced

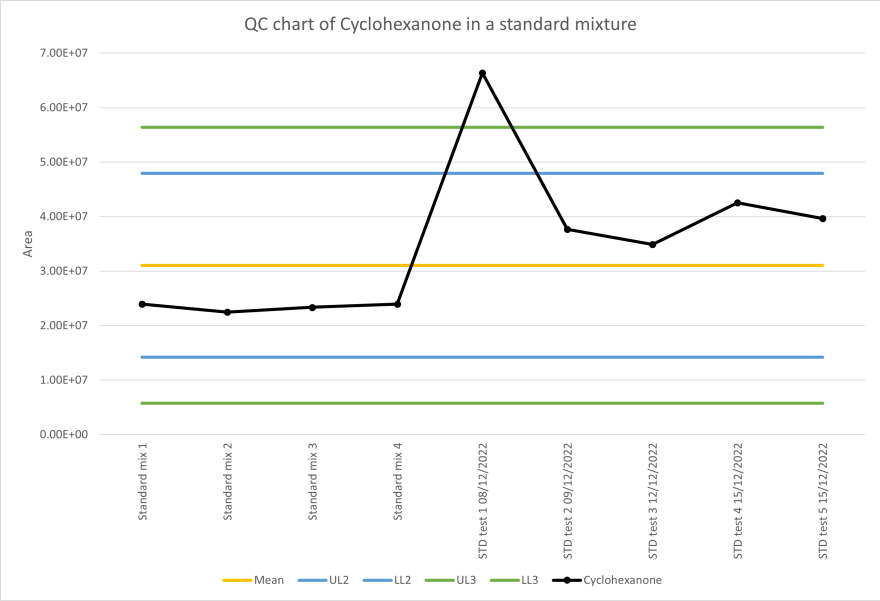


Figure 19: QC chart of cyclohexanone in a standard mixture. LL = lower limit. UL = Upper limit.

distinction between the measured area of the consecutive injections and the standard injected across the analysis, data normalization was applied via PQN using the values from the 4 consecutive standards to compute the normalization factor. Upon normalization, the measured area from the consecutive standard and standard injected during analysis was used to compute the mean, lower and upper limit for the QC chart. However, the "STD test 1 08/12/2022" was excluded from this computation, owing to its marked disparity from all the other standard injections.

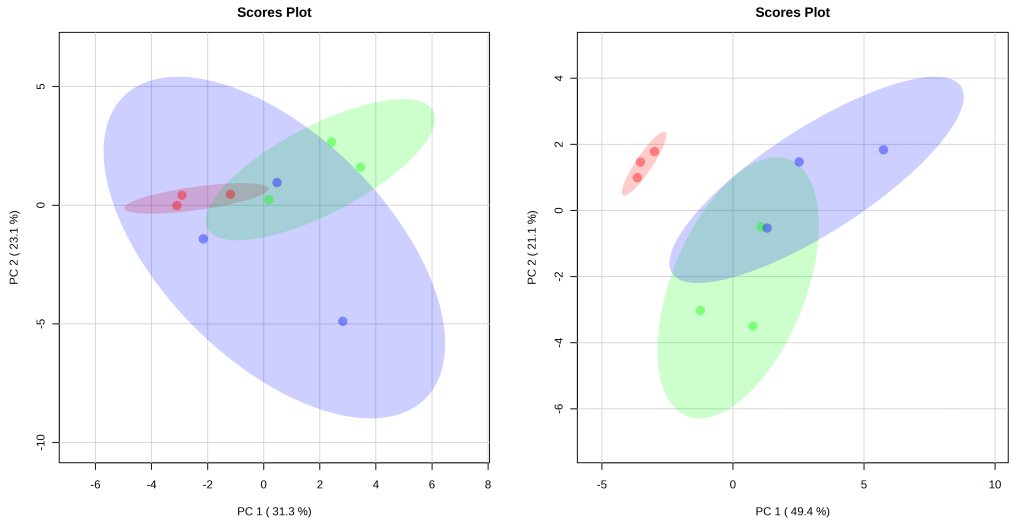
4.3.2 Extraction and injection

All samples were initially scheduled for injection on the 23rd of September 2022. However, during the night of the planned injection, an ice plug formed, rendering the instrument unable to proceed with the analysis. Since the samples were left on the machine sampling rack, temperature control was not possible, and consequently, the samples were discarded. To mitigate the risk of ice plug formation, the injections were spread over several days, starting on the 15th of November 2022, with a maximum of 19 sample injections in a single day.

4.4 Data treatment

4.4.1 Effect of normalization

The impact of data normalization on the quality of the separation of different experimental conditions is emphasized in Figure 20. It can be observed that the application of Probabilistic Quotient Normalization enhances the separation between the 3 experiment conditions, whereas, without this normalization method, the 3 groups remain confounded. This effect is consistent with the findings reported by Dieterle *et al.*[59]. As a result, PQN normalization is applied to all statistical analyses.



(a) Top 20 without PQN normalization

(b) Top 20 with PQN normalization

Figure 20: Comparison of PCA plots obtained for different normalization methods. Red dots correspond to Caco2 + H_2O_2, green dots correspond to Caco2 + media, and blue dots correspond to Caco2 + PBS. (a) PCA plot for the top 20 compounds without PQN normalization (sum=54.4%). (b) PCA plot for the top 20 compounds with PQN normalization (sum=70.5%).

4.4.2 Global view

Figure 21 presents the PCA plot encompassing all cell lines without any statistical data exclusion. Due to the inherent complexity of PCA and the challenges associated with identifying meaningful clusters, the subsequent

section focuses on the individual analysis of each cell line to facilitate a more comprehensive examination. However, the distinction between the oxidized cells, controls, and pool QC is highlighted by the picture, as this PCA plot can be separated into 3 clusters encompassing the mentioned groups.

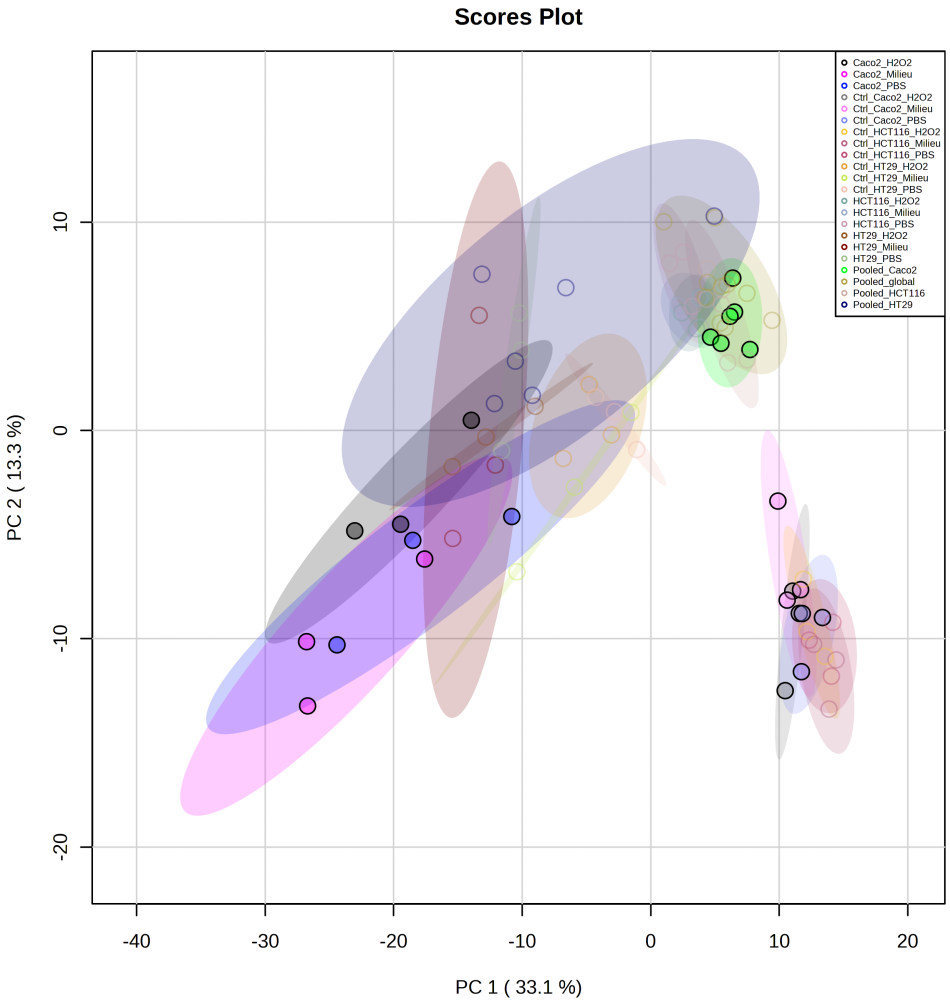
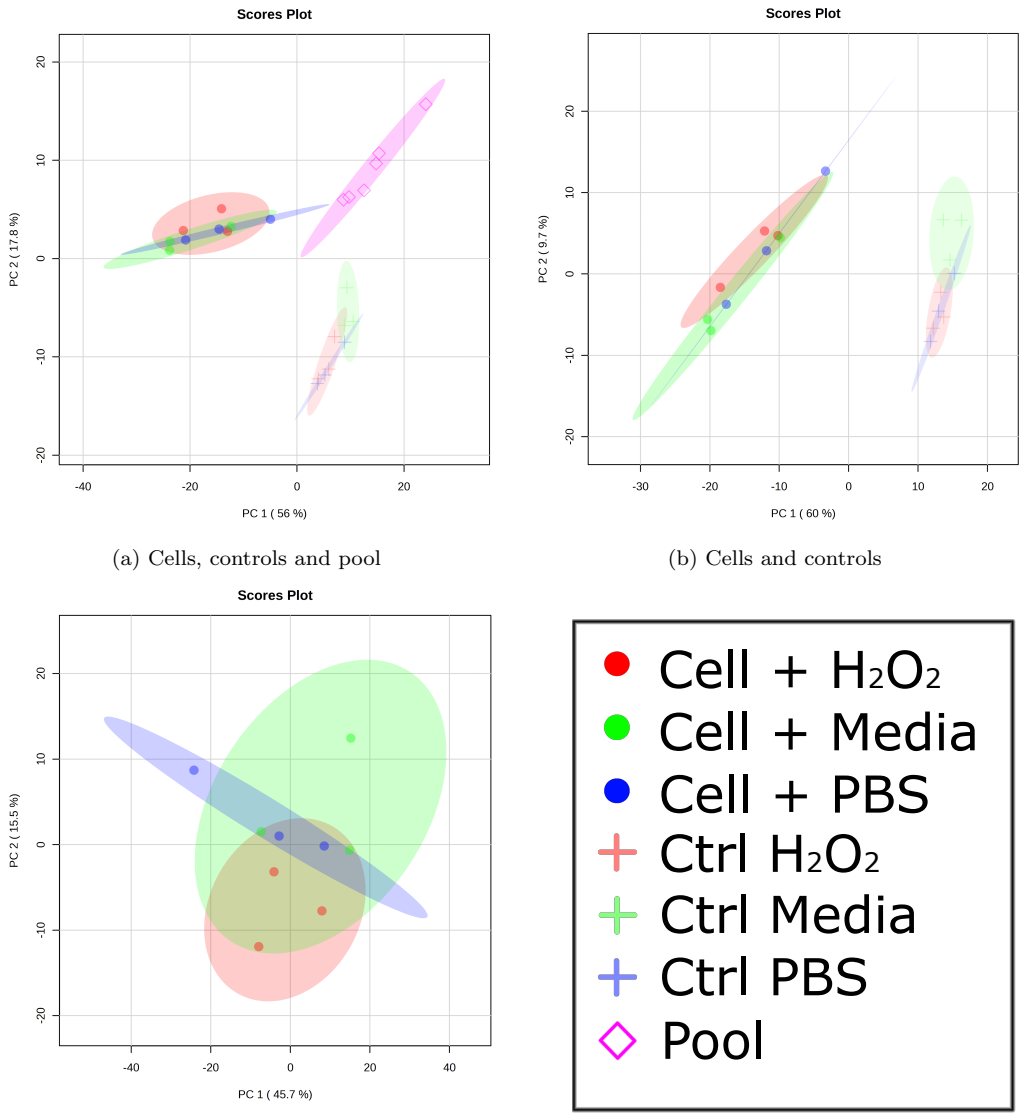


Figure 21: PCA of all the obtained data for all cell lines

4.4.3 Caco2

Figure 22 displays the PCA results obtained for various comparisons involving cell experiments, controls and the prepared pool of the Caco2 cell line. This analysis was performed using all the 360 detected chemical compounds. This figure demonstrates that the clustering patterns of the cell experiments differ from those of the controls and the pool. However, as shown in Figure 22c, the clustering among the 3 experimental conditions used for the cell line does not exhibit clear separation. Consequently, it becomes necessary to filter the detected compound to enhance the differentiation between the oxidation of the Caco2 cell line and the other conditions.

To identify significant VOCs, ANOVA was employed. The F ratio was computed for all detected chemical compounds, under all the different conditions as shown in Figure 23. Considering the equation 6, the F critical value depends on 2 factors: N, the total number of observational values (number of experiments) and K, the number of groups. Consequently, when experiments or a group of experiments are excluded, the values of N and K values change, leading to different F critical values for the different conditions used to build the PCA. This results in a different amount of significant VOCs used to build the PCA depicted in Figure 23. Specifically, 308 significant VOCs were identified for the PCA involving cells, controls and the pool. The PCA comparing cells and controls employed 255 VOCs, while the PCA plot for the cell conditions alone utilized only 6 VOCs. Figure 23c indicates a separation between the oxidation of the cells and the two other conditions.



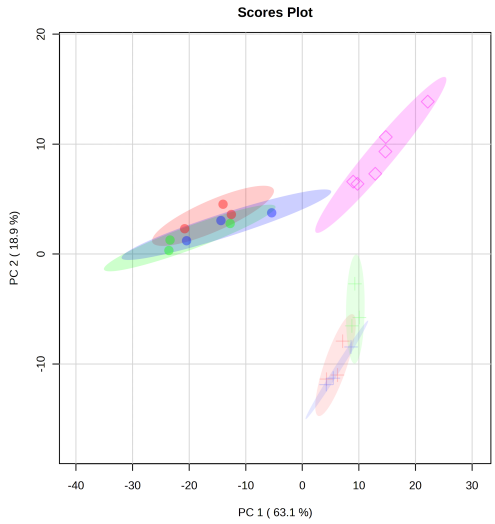
(a) Cells, controls and pool

(b) Cells and controls

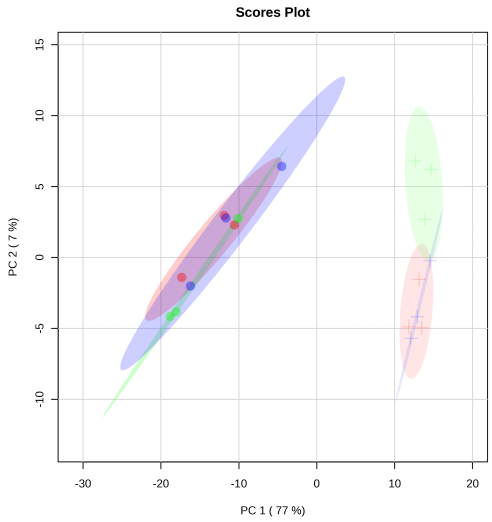
(c) Cells

(d) Legend

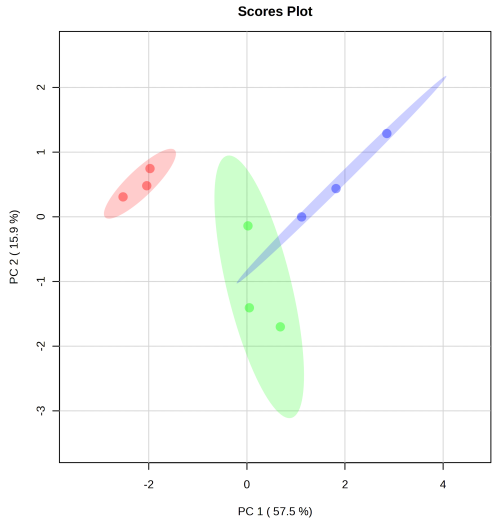
Figure 22: PCA of all of the detected chemical compounds in the Caco2 cell line. (a) PCA plot showing the distribution of cells, controls and the pool (sum=73.8%). (b) PCA plot showing the distribution of the cell samples with the controls (sum=69.7%). (c) PCA plot showing the different cell conditions (sum=61.2%). (d) Color code legend.



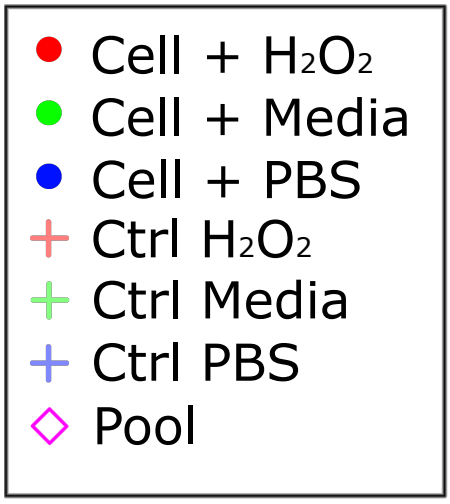
(a) Cells, controls and pool



(b) Cells and controls



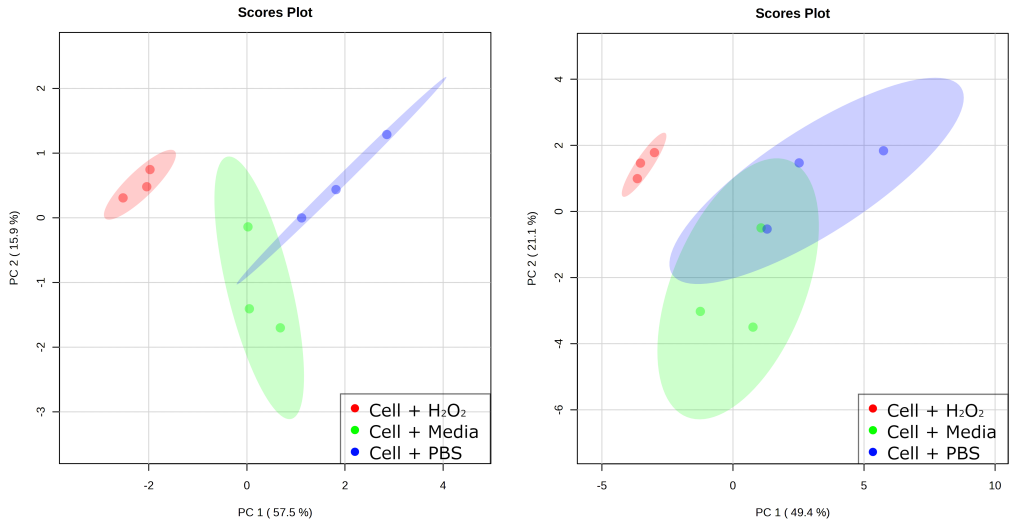
(c) Cells



(d) Legend

Figure 23: PCA of the filter detected chemical compounds in the Caco2 cell line determined by variance analysis. (a) PCA plot showing the distribution of cells, controls and the pool build using a total of 308 VOCs (sum=82%). (b) PCA plot showing the distribution of the cell samples with the controls build using a total of 255 VOCs (sum=84%). (c) PCA plot showing the different cell conditions build using a total of 6 VOCs (sum=73.4%). (d) Color code legend.

To be able to grasp potential biomarkers, the top 20 significant VOCs were determined using ANOVA. Figure 24 compares the PCA obtained using 6 and 20 significant VOCs. The distinction between oxidized cells and controls is still achieved using 20 significant VOCs.



(a) PCA using 6 VOCs

(b) PCA using 20 VOCs

Figure 24: PCA plots obtain for the Caco2 cell line using 6 significant VOCs, (a) and 20 significant VOCs (b). Total variance using the two first principal components accounts for 73.4% (a) and 70.5%.

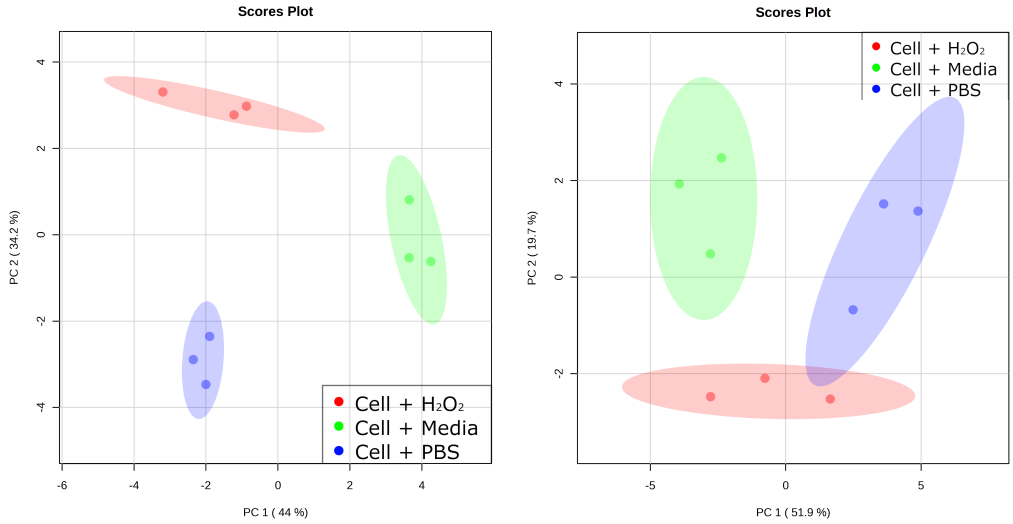
Table 3 gives a list of the 20 identified VOCs using GC Image and ChromaTOF. The table includes all first and second dimension retention time, similarity and similarity reverse match, probability, and Kovats index computed with alkane retention time. Match and reverse match was given by comparison of the mass spectrum with the NIST mass spectra library.

Compound (Caco2)	Rt1 (min)	Rt2 (s)	Similarity	Reverse	Probability (%)	Computed Kovats index	Literature Kovats index
Hexanal	8.7078	2.524	878	892	84.9	/	/
2-Heptanone, 4-methyl-	13.3742	2.527	846	917	68.9	935	/
Ketone	15.4990	2.617	715	715	10.7	998	/
Thiol	16.7490	2.140	855	864	17.6	1035	/
Acetic acid, 2-ethylhexyl ester	20.2488	2.448	830	864	16.0	1144	1144
2-Decanone	21.7070	2.585	849	870	55.4	1191	1184
Decanal	22.1653	2.580	866	870	21.6	1206	1186
Naphthalene, 2-methyl-	25.3318	3.535	888	888	61.8	1315	1312
Imidazole	25.4985	3.195	756	770	21.0	1321	/
Alcohol	28.6232	2.420	800	847	6.5	1438	/
Glutaric acid, butyl ethyl ester	29.3315	2.893	834	834	29.4	1465	/
Benzene alcohol	30.4982	3.591	737	754	73.3	1510	/
Ester	32.4980	2.612	889	898	76.7	1592	/
Glutaric acid, butyl isobutyl ester	33.8728	2.479	867	871	17.8	1660	/
Alcohol	34.0395	2.063	856	856	10.1	1668	/
Ester	36.7893	2.393	856	862	54.9	1837	/
Calamenene	36.9977	2.344	697	760	60.4	1848	/
Hydrocarbon	37.0810	2.386	728	735	21.5	1852	/
Aromatics	37.2060	2.499	660	711	14.2	1858	/
Dibutyl phthalate	37.2893	2.397	835	853	37.5	1862	/

Table 3: Top20 detected VOCs using ANOVA for the oxidized Caco2 cell line. Rt1 indicates the first dimension retention time. Rt2 indicated the second dimension retention time. Kovats index where computed based on known alkane retention time ranging from C_8 to C_{20}.

4.4.4 HCT116 and HT29

As the statistical analysis for the HCT116 and HT29 cell lines follows the same guideline as the analysis performed for the Caco2 cell line in section 4.4.3, the same data filtering procedure was followed. Additional information considering the separation with or without filtration can be found in the Appendices section. Figure 25 depicted the PCA built using the top 20 significant VOCs for the HCT116 and HT29 cell lines.



(a) HCT116 (b) HT29

Figure 25: PCA analysis of the 20 VOCs for HCT116 (a) and HT29 (b) cell lines. Total variance using the two first principal components accounts for 78.2% (a) and 71.6%.

A list of the top 20 VOCs determined using ANOVA is described in the Appendices. The same parameters described for the Caco2 cell line such as retention time, similarity and similarity reverse match, probability, and Kovats index are indicated. Match and reverse match was given by comparison of the mass spectrum with the NIST mass spectra library.

5 Discussion

The following section presents a comprehensive discussion of the obtained results, encompassing recommendations for experimental procedures and data analysis.

5.1 Cell culture

Cell culture may not be a common experiment for researchers in the chemical field, but it plays a crucial role for anyone seeking insight into the various steps reacquired to conduct such culture. While cellular culture can be inherently challenging, with numerous considerations such as avoiding contamination, constant ethanol washing, and precise and rapid movement to prevent cellular damage, it remains an intriguing experience for researchers. The result from this work also highlights three major points to be considered when conducting culture conditions for an experiment:

- Careful attention should be given to the transportation and storage of cellular lines. Tracking the conditions used, the experimenter involved or the temperature employed during transportation is essential to avoid any issues, as experienced in this work

- Experimental design must be thoughtfully planned to conduct as many experiments or samplings as needed. As demonstrated in section 4.4.1, data normalization by PQN significantly impacts the quality of separation for the 3 conditions. Therefore, it was necessary for this experiment to perform a pool of different conditions used with cells and controls. Without this experiment, pool QC would not be accessible, and optimal data separation could not be achieved. It is crucial to consider all aspects of the experiment for metabolomic studies, from data acquisition to quality control and quality assurance, as recommended by the literature [65]. These procedures ensure that the acquired data can be reproducible and ensure confidence in the results

- Thorough documentation of all the steps performed during an experiment is essential. This is especially crucial for cellular cultures where a vast number of different chemical compounds are used, and errors can quickly occur. Rigorously recording all actions helps ensure data integrity and facilitates troubleshooting if any issues arise

5.2 Bi-dimensional GC analysis

5.2.1 Instrument maintenance

Untargeted metabolomic analysis using GC×GC harbors an impressive resolution power that contributes to the understanding of metabolomic profiling for inflammatory mechanisms. However, performing a GC×GC experiment is challenging. In this work, a dual-stage quad-jet cryogenic modulator was utilized to perform the analysis. These modulators required a significant amount of liquid nitrogen for the cooling phase of the column, resulting in higher costs for device utilization and necessitating constant attention from the experimenter to maintain the liquid nitrogen supply. Following the manufacturer's recommendations, regular maintenance is essential and column change can be a very challenging task, especially for researchers without prior experience.

5.2.2 Analytical stability assessment

Adherence to analytical good practices is vital to ensure the reliability of the obtained data. Quality assessment and control procedures should be conducted before, during and after experiments to prevent the analysis from generating misleading data or instrumental inaccuracy. The use of tools such as a QC chart, as described in section 4.3, demonstrates the importance of such practice. Visual inspection of Figure 19 depicted that the measured area for standard injections is consistent across the analysis for almost all injections. Furthermore, this figure illustrates that the standard injected on the 8th of December presents an over-expressed area for cyclohexanone. Similar over or under-expression is found for 12 chemical compounds among the 20 compounds present in the mixture. This indicates a potential issue with the instrument during that specific day. Further investigations for this hypothesis are required to assess the instrument's stability.

5.2.3 Data analysis

Data analysis for metabolomic experiments, while seemingly straightforward in this work, is a time-consuming task that can take days before any meaningful interpretation. The large volume of data generated requires robust

software and hardware for the application of statistical treatments. Common tools such as Excel may take up to a minute to perform a computation and may frequently crash if the hardware is not sufficiently capable of the computation. Through this work, the need for specific skills in data analysis, such as programming using R studio, is highly emphasized. Researchers intending to conduct metabolomic studies involving substantial data sets are advised to develop such skills to expedite the data analysis process and enable the discovery of significant results. Additionally, an excellent understanding of statistics and statistical tools is essential, as some employed in this work may not have been covered in traditional chemical courses for undergraduate students.

5.3 Oxidation

Based on the result obtained from the viability test for the cellular oxidation in section 4.2.2, these findings align with the results discussed in section 1.4, providing additional support for the suitability of the selected oxidation conditions proposed by Zanella *et al.*[20]. The viability of 90% indicates that cellular necrosis was not induced by oxidation. Furthermore, the successful separation between oxidized cell groups and control groups is also achieved using the top 20 VOCs.

The identification of chemical compounds in untargeted metabolomic studies remains a challenge. Although significant VOCs show similarity and similarity reverse match higher than 800 for a majority of the compounds, they often present low probability and incorrect identification. Despite the computation of the Kovats index, further verification is still required to achieve proper identification. One approach to achieve this goal is by injection of the corresponding analytical standards. By using analytical standards with the exact GC parameters used for the sample, complete identification of significant VOCs becomes possible. The different table encompassing the top 20 for the different cellular line also highlight that it was not always possible to identify the VOC using the sample injected and collected with the instrument. Identification for several compounds was only possible using the template (a cumulative image of all the collected samples) and not a specific sample corresponding to the cellular line. The generation of a general template using all the samples of all the cell lines may induce misleading data. Identification using a specific template generated for a specific cell line is

therefore recommended but not pursued for this work.

Among all the cellular lines, the list of significant VOCs reveals 2 particular compounds: hexanal and glutaric acid, butyl isobutyl ester. These 2 compounds were previously highlighted as potential biomarkers in the foundational study of this work[20], particularly for the A549 epithelial cell line. This result suggests that the inflammation mechanism for IDB may induce the same metabolic changes as inflammation in A549 cells. Nevertheless, it is premature to confirm such a hypothesis and further studies need to be conducted to elucidate the link between inflammation in gastrointestinal cells and inflammation in A549 cells.

6 Conclusion

Given how Inflammatory Bowel Disease substantially impacts a significant portion of the population, a comprehensive understanding of its underlying mechanisms is crucial to enhance diagnostic and treatment strategies to ultimately improve the quality of life for affected individuals. Given the ongoing nature of research into these diseases, clinical investigations must be built upon foundational studies to systematically comprehend the intricate inflammation mechanisms under controlled parameters. This study is dedicated to achieving this objective by developing an analytical framework for metabolomics profiling of *in vitro* gastrointestinal cells using multidimensional gas chromatography.

This work seeks to employ the same cultures and oxidation conditions previously developed and documented in a prior publication[20], extending their application to distinct cell lines originating from the gastrointestinal tract. In pursuit of this objective, the description of the materials and method used to culture three cellular lines and the assessment of the HCT116 cell line viability after oxidation is undertaken. This evaluation serves to verify the applicability of the oxidation conditions without inducing necrosis. The observed viability of 90% after a 1h exposure to 0.1 mM of H_2O_2 underscores the viability of these conditions, substantiating their adoption for the oxidation of the three gastrointestinal cell lines, namely HCT116, Caco2, and HT29. Following the application of the outlined procedures, these cell lines are cultured and subjected to oxidative treatment utilizing the prescribed chemical agent H_2O_2. A parallel protocol is executed for control samples, with a meticulous collection of both global and specific quality-controlled sample pools to enhance data separation. Collected samples undergo analysis using GC×GC-TOFMS to extract VOCs fingerprint, facilitated through SPME sampling. The resultant dataset is collected, filtered and categorized in correspondence with each cell line, a necessity given the inherent complexity of the amassed information.

To fortify data segregation and analysis, ANOVA and multivariate statistical analysis are applied, leveraging a PQN normalization approach. This underscores the pivotal role of quality control assessment throughout the experimental procedure. Employing the top 20 significant VOCs unique to each cell line, this study highlights 2 specific biomarkers, namely hexanal and

glutaric acid, butyl isobutyl ester. These biomarkers hold promise in offering valuable insights into the intricate mechanisms governing inflammation, as they were already highlighted in previous studies. However, it is incumbent upon future research endeavors to corroborate these findings and affirm the analytical identification of these two compounds.

7 Perspective

Several major steps are identified that hold the potential to improve the results of this work:

- **Replication and validation**: Replication of the results from this work to confirm their validity. Such investigations, employing identical or distinct GC×GC-TOFMS parameters such as same column parameters or different instruments, can confirm the detection of the chemical compounds identified as potential biomarkers. The use of a different column set is hypothesized to improve the quality of separation for the cell lines and improve identification

- **Methodological Diversification**: Alternating methodologies, such as derivatization or lipidomics may help to undercover other chemical biomarkers that were not detected using the method used for this work. In particular, lipidomics was used in cohorts with metabolomic in a previous study from OBiAChem [66] that helped to gain insight into the metabolites and lipids originating human serum affected with several forms of colorectal cancer. Such methods can be translated for this work to uncover a broader range of molecules that can be classified as potential biomarkers of inflammation

- **Enhanced Analytical Identification**: Analytical identification becomes critical for the identification of the biomarkers found in this study. Further data analysis using a specific template generated for each cell line may improve the identification of specific biomarkers as well as using reference materials to improve the quality of the identification. The use of comprehensive two-dimensional gas chromatography coupled with high-resolution time-of-flight mass spectrometry can further improve the identification of a selected range of metabolites

- **Pyrolysis experiment**: In culmination, a project of research involving the analysis of the metabolites resulting from the pyrolysis of the oxidized HCT116 cells is planned, which serves as the cornerstone of this study. This novel approach aspires to give a broader understanding of the complex mechanisms lying behind IBD, using a new approach to unravel potential discoveries

Abbreviations

DAMPs: Damage-Associated Molecular Patterns
DMEM: Dulbecco's Modified Eagle Medium
DPBS: Dulbecco's Phosphate-Buffered Saline
FBS: Fetal Bovine Serum
GC: Gas Chromatography
GC×GC: Bi-dimensional Gas Chromatography
HETP or H: Height Equivalent to a Theoretical Plate
HS: Headspace
IBD: Inflammatory Bowel Disease
IECs: Intestinal Epithelial Cells
LLE: Liquid-Liquid Extraction
MS: Mass Spectrometry
m/z : Mass-to-charge ration
N: Number of theoretical plates
NMR: Nuclear Magnetic Resonance
One-way ANOVA: One-way Analysis Of Variance
PBS++: Phosphate Buffered Saline without calcium or magnesium
PCA: Principal Component Analysis
PQN: Probabilistic quotient normalization
PRRs: Pattern Recognition Receptors
QC: Quality Control
RSD: Relative Standard Deviation
SBSE: Stir Bar Sorptive Extraction
SPME: Solid Phase Micro Extraction
TOF-MS: Time-Of-Flight - Mass Spectrometry
VOC: Volatile Organic Compounds

List of Figures

List of Tables

References

[1] Institute for Quality and Efficiency in Health Care. "What is an inflammation?" (Feb. 2018), [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK279298/#:~:text=Very%20generally%20speaking%2C%20inflammation%20is,a%20splinter%20in%20your%20finger. (visited on 05/07/2023).

[2] R. Pahwa, A. Goyal, P. Bansal, and I. Jialal, "Chronic inflammation," 2018.

[3] L. Chen, H. Deng, H. Cui, *et al.*, "Inflammatory responses and inflammation-associated diseases in organs," *Oncotarget*, vol. 9, no. 6, p. 7204, 2018.

[4] P. A. Ward, "Acute and chronic inflammation," *Fundamentals of inflammation*, vol. 3, pp. 1–16, 2010.

[5] S. Alatab, S. G. Sepanlou, K. Ikuta, *et al.*, "The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: A systematic analysis for the global burden of disease study 2017," *The Lancet gastroenterology & hepatology*, vol. 5, no. 1, pp. 17–30, 2020.

[6] D. Furman, J. Campisi, E. Verdin, *et al.*, "Chronic inflammation in the etiology of disease across the life span," *Nature medicine*, vol. 25, no. 12, pp. 1822–1832, 2019.

[7] NHS controls. "Inflammatory bowel disease." (2023), [Online]. Available: https://www.nhs.uk/conditions/inflammatory-bowel-disease/ (visited on 06/07/2023).

[8] Centers for Disease Control and Prevention. "What is inflammatory bowel disease (ibd)?" (2022), [Online]. Available: https://www.cdc.gov/ibd/what-is-IBD.htm#:~:text=Inflammatory%20bowel%20disease%20(IBD)%20is,damage%20to%20the%20GI%20tract. (visited on 06/07/2023).

[9] N. R. C. (C. on Methods of Producing Monoclonal Antibodies *et al.*, "Summary of advantages and disadvantages of in vitro and in vivo methods," *Monoclonal antibody production*, pp. 22–24, 1999.

[10] M. Nickien, A. Heuijerjans, K. Ito, and C. C. van Donkelaar, "Comparison between in vitro and in vivo cartilage overloading studies based on a systematic literature review," *Journal of Orthopaedic Research®*, vol. 36, no. 8, pp. 2076–2086, 2018.

[11] O. Graudejus, R. Ponce Wong, N. Varghese, S. Wagner, and B. Morrison, "Bridging the gap between in vivo and in vitro research: Reproducing in vitro the mechanical and electrical environment of cells in vivo," *Frontiers in Cellular Neuroscience*, vol. 12, 2018.

[12] I. Samoilă, S. Dinescu, and M. Costache, "Interplay between cellular and molecular mechanisms underlying inflammatory bowel diseases development—a focus on ulcerative colitis," *Cells*, vol. 9, no. 7, p. 1647, 2020.

[13] L. Abdulkhaleq, M. Assi, R. Abdullah, M. Zamri-Saad, Y. Taufiq-Yap, and M. Hezmee, "The crucial roles of inflammatory mediators in inflammation: A review," *Veterinary world*, vol. 11, no. 5, p. 627, 2018.

[14] G. S. Jutley and S. P. Young, "Metabolomics to identify biomarkers and as a predictive tool in inflammatory diseases," *Best practice & research Clinical rheumatology*, vol. 29, no. 6, pp. 770–782, 2015.

[15] M. Kumar, M. Garand, and S. Al Khodor, "Integrating omics for a better understanding of inflammatory bowel disease: A step towards personalized medicine," *Journal of translational medicine*, vol. 17, pp. 1–13, 2019.

[16] M. Vailati-Riboni, V. Palombo, and J. J. Loor, "What are omics sciences?" *Periparturient diseases of dairy cows: a systems biology approach*, pp. 1–7, 2017.

[17] E. Clay, S. Kapoor, R. Bayley, G. R. Wallace, S. P. Young, M. Fitzpatrick, *et al.*, "Metabolomics in the analysis of inflammatory diseases," *Metabolomics*, 2012.

[18] E. Riekeberg and R. Powers, "New frontiers in metabolomics: From measurement to insight," *F1000Research*, vol. 6, 2017.

[19] C. B. Clish, "Metabolomics: An emerging but powerful tool for precision medicine," *Molecular Case Studies*, vol. 1, no. 1, a000588, 2015.

[20] D. Zanella, M. Henket, F. Schleich, *et al.*, "Comparison of the effect of chemically and biologically induced inflammation on the volatile metabolite production of lung epithelial cells by gc×gc-tofms," *Analyst*, vol. 145, pp. 5148–5157, 15 2020. DOI: 10.1039/D0AN00720J. [Online]. Available: http://dx.doi.org/10.1039/D0AN00720J.

[21] Laboratory for nanomaterial-based devices. "Volatolomics." (2015), [Online]. Available: https://lnbd.technion.ac.il/research/electronic-skin/ (visited on 11/07/2023).

[22] K. A. Kouremenos, M. Johansson, and P. J. Marriott, "Advances in gas chromatographic methods for the identification of biomarkers in cancer," *Journal of Cancer*, vol. 3, p. 404, 2012.

[23] N. Di Giovanni, M.-A. Meuwis, E. Louis, and J.-F. Focant, "Untargeted serum metabolic profiling by comprehensive two-dimensional gas chromatography–high-resolution time-of-flight mass spectrometry," *Journal of proteome research*, vol. 19, no. 3, pp. 1013–1028, 2019.

[24] ATCC, *Htc116*, Accessed May 19, 2022. [Online]. Available: https://www.atcc.org/products/ccl-247.

[25] D. S. Goodsell, "The Molecular Perspective: The ras Oncogene," *The Oncologist*, vol. 4, no. 3, pp. 263–264, Jun. 1999, ISSN: 1083-7159. DOI: 10.1634/theoncologist.4-3-263. eprint: https://academic.oup.com/oncolo/article-pdf/4/3/263/41836063/oncolo_4_3_263.pdf. [Online]. Available: https://doi.org/10.1634/theoncologist.4-3-263.

[26] T. University and Harvard, *Oncogenes*, Accessed May 25, 2022, Dec. 2020. [Online]. Available: https://bio.libretexts.org/@go/page/5080.

[27] S. Giuriato and D. W. Felsher, "How cancers escape their oncogene habit," *Cell cycle*, vol. 2, no. 4, pp. 328–331, 2003.

[28] I. J. Hidalgo, T. J. Raub, and R. T. Borchardt, "Characterization of the human colon carcinoma cell line (caco-2) as a model system for intestinal epithelial permeability," *Gastroenterology*, vol. 96, no. 2, pp. 736–749, 1989.

[29] A. Le Bivic, M. Hirn, and H. Reggio, "Ht-29 cells are an in vitro model for the generation of cell polarity in epithelia during embryonic differentiation.," *Proceedings of the National Academy of Sciences*, vol. 85, no. 1, pp. 136–140, 1988.

[30] E. Perkins, "Volatile odor and flavor components formed in deep frying," pp. 51–56, Dec. 2007. DOI: 10.1016/B978-1-893997-92-9.50010-4.

[31] N. H. Snow, *Basic Multidimensional Gas Chromatography*. Academic Press, 2020.

[32] H. M. McNair, J. M. Miller, and N. H. Snow, *Basic gas chromatography*. John Wiley & Sons, 2019.

[33] D. A. Skoog and D. M. West, *Chimie analytique*. De Boeck Supérieur, 2015.

[34] A. J. Martin and R. L. Synge, "A new form of chromatogram employing two liquid phases: A theory of chromatography. 2. application to the micro-determination of the higher monoamino-acids in proteins," *Biochemical Journal*, vol. 35, no. 12, p. 1358, 1941.

[35] E. Stauffer, J. A. Dolan, and R. Newman, "Chapter 8 - gas chromatography and gas chromatography—mass spectrometry," in *Fire Debris Analysis*, E. Stauffer, J. A. Dolan, and R. Newman, Eds., Burlington: Academic Press, 2008, pp. 235–293, ISBN: 978-0-12-663971-1. DOI: https://doi.org/10.1016/B978-012663971-1.50012-9. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780126639711500129.

[36] D. Rusjan, "Aromas in grape and wine," in Oct. 2010, pp. 411–442. DOI: 10.1007/978-90-481-9283-0_29.

[37] J. de Zeeuw and J. de Zeeuw, "Impact of GC Parameters on The Separation : Stationary phase," *Separation Science*, vol. 6, no. 4, pp. 8–13, 2014.

[38] J. de Zeeuw and J. de Zeeuw, "Impact of GC Parameters on The Separation : Internal diameter," *Separation Science*, vol. 6, no. 4, pp. 8–13, 2014.

[39] J. Zeeuw, "Impact of GC Parameters on The Separation : Column length," *Separation Science*, vol. 6, no. 4, pp. 8–13, 2014.

[40] J. de Zeeuw and J. de Zeeuw, "Impact of GC Parameters on The Separation : Film thickness," *Separation Science*, vol. 6, no. 4, pp. 8–13, 2014.

[41] J. de Zeeuw, "Impact of GC Parameters on The Separation : Column temperature," *Separation Science*, vol. 6, no. 4, pp. 8–13, 2014. [Online]. Available: www.sepscience.com.

[42] J. de Zeeuw and J. de Zeeuw, "Impact of GC Parameters on The Separation : Choice of Carrier Gas and Liner Velocity," *Separation Science*, vol. 6, no. 4, pp. 8–13, 2014.

[43] A. L. Rockwood, M. M. Kushnir, and N. J. Clarke, "2 - mass spectrometry," in *Principles and Applications of Clinical Mass Spectrometry*, N. Rifai, A. R. Horvath, and C. T. Wittwer, Eds., Elsevier, 2018, pp. 33–65, ISBN: 978-0-12-816063-3. DOI: https://doi.org/10.1016/B978-0-12-816063-3.00002-5. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128160633000025.

[44] D. J. Kiemle, R. M. Silverstein, and F. X. Webster, *Identification spectrométrique de composés organiques-3ème édition*. De Boeck Superieur, 2016.

[45] LECO, *Flux™: Modulateur de flux gcxgc*, Accessed March 14, 2022. [Online]. Available: https://fr.leco.com/product/flux.

[46] T. Górecki, J. Harynuk, and O. Panić, "The evolution of comprehensive two-dimensional gas chromatography (gc× gc)," *Journal of separation science*, vol. 27, no. 5-6, pp. 359–379, 2004.

[47] P. J. Marriott and R. M. Kinghorn, "Longitudinally modulated cryogenic system. a generally applicable approach to solute trapping and mobilization in gas chromatography," *Analytical Chemistry*, vol. 69, no. 13, pp. 2582–2588, 1997, PMID: 21639393. DOI: 10.1021/ac961310w. eprint: https://doi.org/10.1021/ac961310w. [Online]. Available: https://doi.org/10.1021/ac961310w.

[48] E. B. Ledford Jr and C. Billesbach, "Jet-cooled thermal modulator for comprehensive multidimensional gas chromatography," *Journal of High Resolution Chromatography*, vol. 23, no. 3, pp. 202–204, 2000.

[49] L. Ramos, *Comprehensive two dimensional gas chromatography*. Elsevier, 2009.

[50] J.-F. Focant, A. Sjödin, and D. G. Patterson Jr, "Qualitative evaluation of thermal desorption-programmable temperature vaporization-comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry for the analysis of selected halogenated contaminants," *Journal of Chromatography A*, vol. 1019, no. 1-2, pp. 143–156, 2003.

[51] P. Andrew Tipler, *An introduction to headspace sampling in gas chromatography, fundamentals and theory*, Accessed September 21, 2022, 2013. [Online]. Available: `https://resources.perkinelmer.com/corporate/pdfs/downloads/gde_intro_to_headspace.pdf`.

[52] B. Kolb and L. S. Ettre, *Static headspace-gas chromatography: theory and practice*. John Wiley & Sons, 2006.

[53] S. E. Reichenbach, X. Tian, Q. Tao, E. B. Ledford Jr, Z. Wu, and O. Fiehn, "Informatics for cross-sample analysis with comprehensive two-dimensional gas chromatography and high-resolution mass spectrometry (gcxgc–hrms)," *Talanta*, vol. 83, no. 4, pp. 1279–1288, 2011.

[54] I. Gertsman and B. A. Barshop, "Promises and pitfalls of untargeted metabolomics," *Journal of inherited metabolic disease*, vol. 41, pp. 355–366, 2018.

[55] C. Science and E. (D. of the University of Nebraska, *Gc image gcxgc edition users' guide*, Accessed June 22, 2023, 2023. [Online]. Available: `https://www.gcimage.com/gcxgc/usersguide/index.html`.

[56] P. Cuevas-Delgado, D. Dudzik, V. Miguel, S. Lamas, and C. Barbas, "Data-dependent normalization strategies for untargeted metabolomics—a case study," *Analytical and Bioanalytical Chemistry*, vol. 412, pp. 6391–6405, 2020.

[57] B. Li, J. Tang, Q. Yang, *et al.*, "Performance evaluation and online realization of data-driven normalization methods used in lc/ms based untargeted metabolomics analysis," *Scientific reports*, vol. 6, no. 1, p. 38 881, 2016.

[58] J. Chen, P. Zhang, M. Lv, *et al.*, "Influences of normalization method on biomarker discovery in gas chromatography–mass spectrometry-based untargeted metabolomics: What should be considered?" *Analytical chemistry*, vol. 89, no. 10, pp. 5342–5348, 2017.

[59] F. Dieterle, A. Ross, G. Schlotterbeck, and H. Senn, "Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. application in 1h nmr metabonomics," *Analytical chemistry*, vol. 78, no. 13, pp. 4281–4290, 2006.

[60] T. K. Kim, "Understanding one-way anova using conceptual figures," *Korean journal of anesthesiology*, vol. 70, no. 1, pp. 22–26, 2017.

[61] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20 150 202, 2016. DOI: 10.1098/rsta.2015.0202.

[62] L. I. Smith, "A tutorial on principal components analysis," 2002.

[63] W. R. Group, *Metaboanalyst5.0 - user-friendly, streamlined metabolomics data analysis*, Accessed May, 2023, 2023. [Online]. Available: https://www.metaboanalyst.ca/docs/About.xhtml.

[64] T. L. Riss, R. A. Moravec, A. L. Niles, *et al.*, "Cell viability assays," *Assay Guidance Manual [Internet]*, 2016.

[65] J. A. Kirwan, H. Gika, R. D. Beger, *et al.*, "Quality assurance and quality control reporting in untargeted metabolic phenotyping: Mqacc recommendations for analytical quality management," *Metabolomics*, vol. 18, no. 9, p. 70, 2022.

[66] K. Bhatt, T. Orlando, M.-A. Meuwis, E. Louis, P.-H. Stefanuto, and J.-F. Focant, "Comprehensive insight into colorectal cancer metabolites and lipids for human serum: A proof-of-concept study," *International Journal of Molecular Sciences*, vol. 24, no. 11, p. 9614, 2023.

Appendices

Measured area for 4 consecutive standard injections

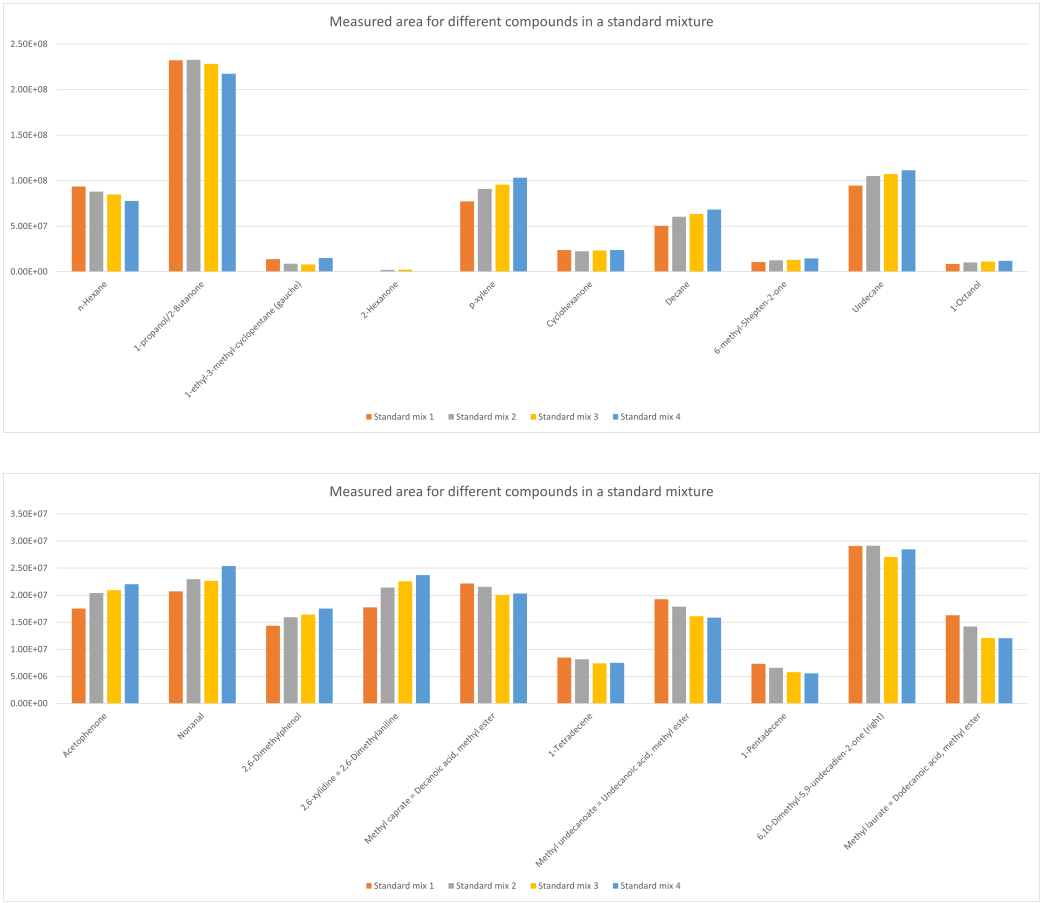
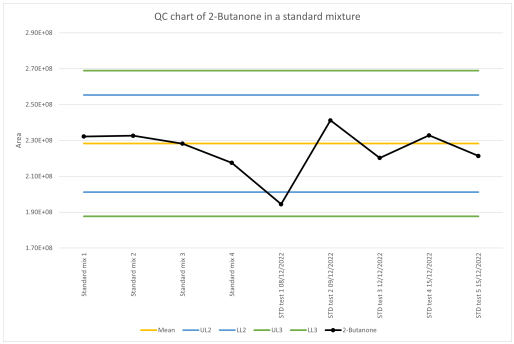
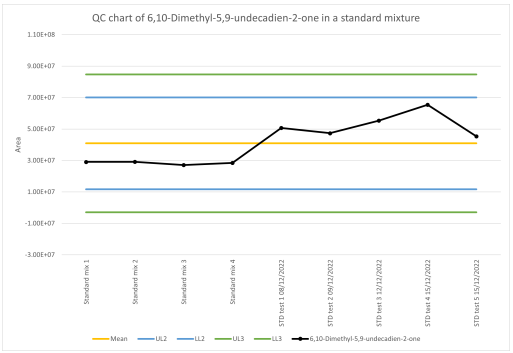


Figure 26: Normalized measured standard area for 4 consecutive injections. The upper figure presents the first half of the measured standards, whereas the lower figure presents the other standards. Areas were normalized by PQN.

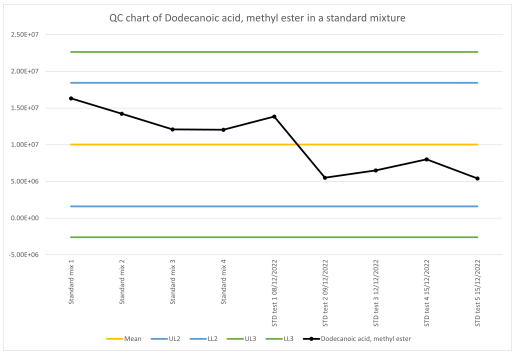
QC chart



(a) 2-butanone



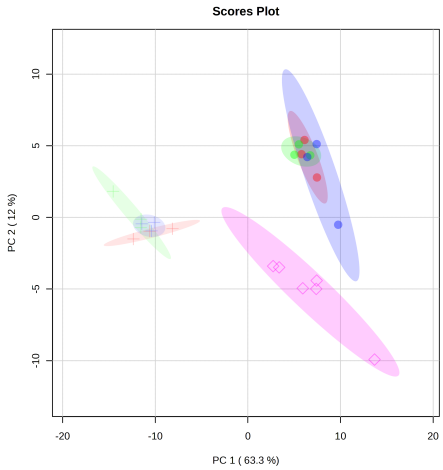
(b) 6,10-dimethyl-5,9-undecadien-2-one



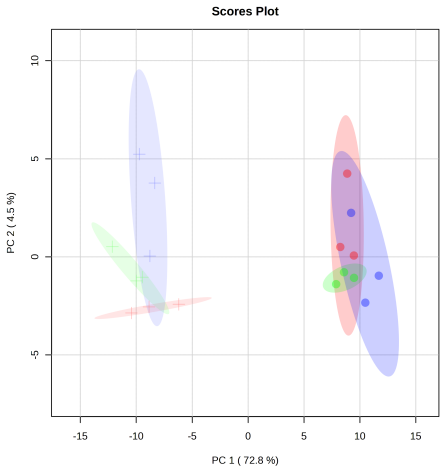
(c) Dodecanoic acid

Figure 27: QC chart built using different compounds found in the standard mixture. (a) QC chart for 2-butanone. (b) QC chart for 6,10-dimethyl-5,9-undecadien-2-one. (c) QC chart for dodecanoic acid. LL = lower limit. UL = Upper limit.

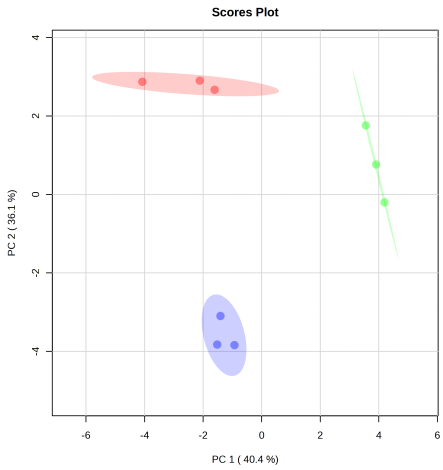
Oxidation of HCT116



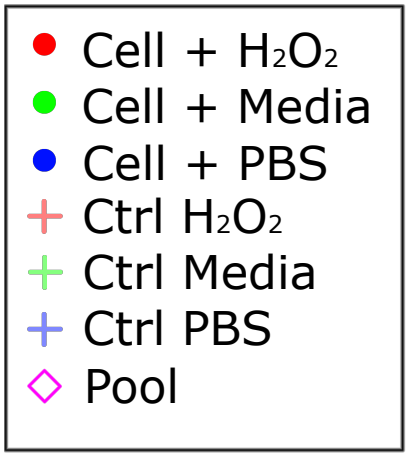
(a) Cells, controls and pool



(b) Cells and controls



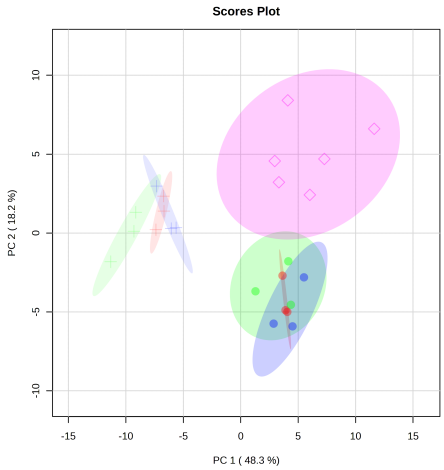
(c) Cells



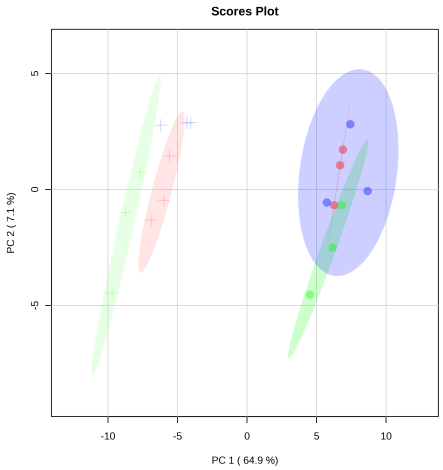
(d) Legend

Figure 28: PCA of the filter detected chemical compounds in the HCT116 cell line determined by variance analysis. (a) PCA plot showing the distribution of cells, controls and the pool build using a total of 132 VOCs (sum=75.3%). (b) PCA plot showing the distribution of the cell samples with the controls build using a total of 129 VOCs (sum=77.3%). (c) PCA plot showing the different cell conditions build using a total of 23 VOCs (sum=76.5%). (d) Color code legend.

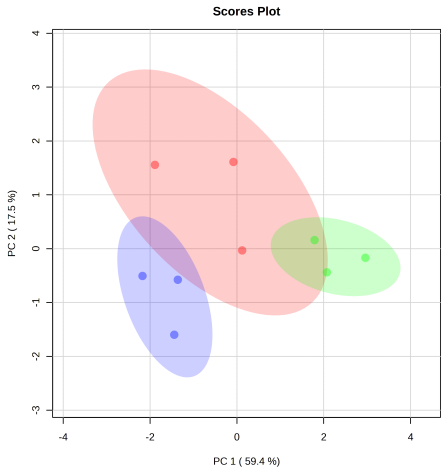
Oxidation of HT29



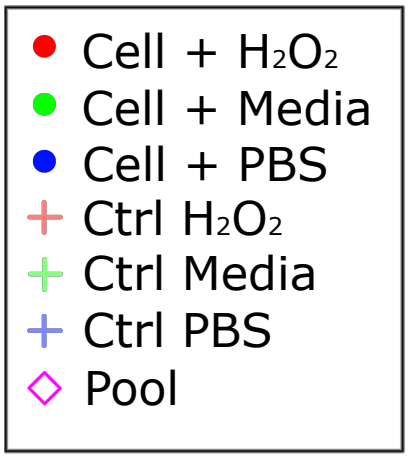
(a) Cells, controls and pool



(b) Cells and controls



(c) Cells



(d) Legend

Figure 29: PCA of the filter detected chemical compounds in the HT29 cell line determined by variance analysis. (a) PCA plot showing the distribution of cells, controls and the pool build using a total of 87 VOCs (sum=66.5%). (b) PCA plot showing the distribution of the cell samples with the controls build using a total of 74 VOCs (sum=72%). (c) PCA plot showing the different cell conditions build using a total of 6 VOCs (sum=76.9%). (d) Color code legend.

Top20 HCT116

Compound	Rt1 (min)	Rt2 (s)	Similarity	Reverse	Probability (%)	Computed Kovats index	Literature Kovats index
Methylene chloride	2.6248	1.93	940	940	98.7	/	/
3-Heptanone	11.5827	2.558	711	837	35.4	883	887
Ketone	15.0408	2.569	794	794	22.1	984	/
Hydrocarbon	15.7490	2.041	847	868	31.1	1005	/
Hydrocarbon	16.7073	2.091	814	814	6.5	1034	/
Cyclohexanone	26.4150	2.854	698	707	13.6	1355	/
Hydrocarbon	31.4980	2.128	832	857	14.9	1551	/
Amide	31.8730	2.717	800	814	29.5	1567	/
Ketone	32.6230	2.55	756	775	22.4	1597	/
Hydrocarbon	32.7063	2.145	890	911	15.4	1601	/
Glutaric acid, butyl isobutyl ester	33.8728	2.479	831	849	13.7	1660	/
Benzaldehyde	35.7478	2.517	781	796	82.9	1779	/
Oxypentadecane	15.1250	1.325	/	/	/	/	/
Oxypentadecane	23.0000	1.225	/	/	/	/	/
Phenol	31.5000	1.550	/	/	/	/	/
Nitrobenzoic acid	32.667	1.710	/	/	/	/	/
No match, close to glutaric acid, butyl isobutyl ester	34.0830	1.485	/	/	/	/	/
No match, close to glutaric acid, butyl isobutyl ester	34.1670	1.305	/	/	/	/	/
Gentamicin a	34.2080	1.210	/	/	/	/	/
Ergosta	35.7920	1.090	/	/	/	/	/

Table 4: Top20 detected VOCs using ANOVA for the oxidized HCT116 cell line. Rt1 indicates the first dimension retention time. Rt2 indicated the second dimension retention time. Kovats index where computed based on known alkane retention time ranging from C_8 to C_{20}. Rows in orange present the compound detected only in the template using GC-Image. Uncolored rows depicted the compounds detected in the template using GC-Image and in the samples using ChromaTOF.

Top20 HT29

Compound	Rt1 (min)	Rt2 (s)	Similarity	Reverse	Probability (%)	Computed Kovats index	Literature Kovats index
1-Butanol	4.7498	2.203	857	867	41.5	/	/
2,4-Dimethyl-1-heptene	10.1243	2.054	882	882	69.8	840	850
Benzene, (1,1-dimethylpropyl)-	18.8738	2.723	842	847	62.6	1100	1096
Etymemazine	22.6653	2.087	826	921	78	1223	/
Alcohol	24.0818	2.545	756	785	16.2	1271	/
Ester	24.1235	3.143	843	865	76.2	1273	/
Alcohol	24.1235	2.783	772	773	43.4	1273	/
Nonadecane	24.4568	2.043	840	855	11.1	1284	/
Hydrocarbon	24.8318	2.043	845	866	8.2	1297	/
Cyclopentane	27.7483	2.235	800	823	24.7	1404	/
Ketone	28.9565	2.791	888	908	56.3	1450	/
Hexadecane	33.7062	1.959	872	881	15.6	1652	/
Hydrocarbon	33.9562	1.989	811	831	7.8	1664	/
Ester	37.7477	2.553	932	940	61.1	1885	/
Oxypentadecane	19.5000	1.375	/	/	/	/	/
Nitrobenzoic acid	24.1250	1.775	/	/	/	/	/
Imidazole	29.0000	1.225	/	/	/	/	/
Imidazole	29.0000	1.270	/	/	/	/	/
Benzeneethanamine	29.0000	1.445	/	/	/	/	/
Iodohistidine	38.458	2.275	/	/	/	/	/

Table 5: Top20 detected VOCs using ANOVA for the oxidized HCT116 cell line. Rt1 indicates the first dimension retention time. Rt2 indicated the second dimension retention time. Kovats index where computed based on known alkane retention time ranging from C_8 to C_{20}. Rows in orange present the compound detected only in the template using GC-Image. Uncolored rows depicted the compounds detected in the template using GC-Image and in the samples using ChromaTOF.