Comparative analytical work-up for reproducible metabolite derivatisation for gas chromatography-mass spectrometry

M Viviers

orcid.org 0000-0001-6134-8893

Dissertation accepted in partial fulfilment of the requirements for the degree *Master of Science in Biochemistry* at the North-West University

Supervisor: Mr PJ Jansen van Rensburg

Co-supervisor: Dr L Venter

Graduation May 2024

28803027

ABSTRACT

Gas chromatography-mass spectrometry (GC-MS) is one of the most popular analytical instruments used for metabolomics analyses. However, chemical derivatisation of the sample is generally required to increase the volatility of non-volatile and semi-volatile metabolites prior to GC-MS analyses. Various methods of chemical derivatisation exist, but only a few, well standardised techniques such as silylation, acylation, and alkylation are routinely employed. The method of choice is highly dependent on the nature of compounds analysed making it important to familiarise oneself with the different derivatisation methods to improve or develop analytical methods suitable for the samples at hand.

The current investigation sets out to compare five different derivatisation techniques, utilising the reagents N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA), BSTFA with methoxyamine hydrochloride, N-methyl-(trimethylsilyl)trifluoroacetamide (MSTFA), MSTFA with methoxyamine hydrochloride, and methyl chloroformate (MCF) on a set of key metabolites prior to GC-MS analyses. The samples of interest included a standard solution of 19 selected compounds and four internal standards (2-amino-3-bromo-5-methylbenzoic acid, 3-phenylbutyric acid, norleucine and nonadecanoic acid methyl ester), as well as biological quality control serum samples. In both instances, the samples were freeze-dried, derivatised and analysed using a GC-MS approach. The data obtained were processed followed by statistical comparisons.

Data obtained from analyses of the standard stock solution showed clear separation when subjected to principal component analyses, distinguishing between silylated and alkylated samples, suggesting different profiles for the same set of compounds. In terms of compound response and repeatability of each derivatisation method, MCF derivatisation resulted in the lowest average relative standard deviation (RSD) values for the majority of the analysed compounds. Additional comparisons showed that the use of silylation (BSTFA and MSTFA), without the methoximation step, resulted in the same number of compounds detected, the formation of multiple derivatives for the same compound and displayed similar repeatability amongst tested methods. Likewise, the use of methoximation prior to silylation showed no differences when using BSTFA or MSTFA. When alkylation (MCF) was used as a derivatisation method, mostly a single derivative was produced per compound, suggesting that alkylation prevents the formation of multiple derivates per compound. Stability of derivatives produced per derivatisation method over time, showed similar metabolite profiles when using MSTFA with methoxyamine hydrochloride solution (MeOX) or MCF derivatisation

methods. Analyses of the same batch of samples 84 hours post-preparation confirmed that the derivatisation method influenced the relative abundance of the compounds of interest, highlighting the need for method optimisation for the samples at hand. Analyses of two sets (groups A and B) of biological samples were used to evaluate and compare the derivatisation methods as proof of concept for metabolomics studies. In this comparison, BSTFA with MeOX was identified as superior compared to the other methods, as this method resulted in the highest number of statistically significant features detected between groups A and B.

This study confirms that no universal derivatising agent can produce satisfactory results for every compound in a sample, independent of its chemical class. Derivatisation reagents should be standardised for the compounds of interest within a specific metabolomics study, to find the method that ensures repeatability for the majority of the compounds present in a sample, while also considering compounds that cannot be reliably detected and using the outcome of that with caution.

Key words: Alkylation; Derivatisation; Gas chromatography-mass spectrometry; Metabolomics; Methoximation; Repeatability; Silylation; Stability.

ACKNOWLEDGEMENTS

I would like to express my gratitude towards the following individuals for their support and contributions to this study:

• Dr Leonie Venter, my co-supervisor, for all the effort and time that Dr has spent in answering questions and proofreading work. Thank you for all the guidance you have given me through the course of my study and throughout the writing of my dissertation.

• Mr Peet Jansen van Rensburg, my supervisor, was always willing to help me no matter the time. Thank you for always being actively part of my practical training. Thank you for the privilege of being able to learn from you.

• My friends, Cara Olivier, Armand Vorster, and Louis Mostert for all your support, encouragement, and care throughout my Master's degree.

• My family, for all the emotional and financial support, words of encouragement, and simply just for always being there. Thanks for giving me the privilege to follow my dreams.

And most of all, thanks to the Lord, who has blessed me with the ability to learn and people to learn from, for giving me strength and wisdom each day to pursue my ambitions and live life to the fullest.

TABLE OF CONTENTS

ABSTRACT 1
ACKNOWLEDGEMENTS 3
TABLE OF CONTENTS 4
LIST OF FIGURES AND TABLES 6
LIST OF ABBREVIATIONS 8

CHAPTER 1: INTRODUCTION 9
1.1 BACKGROUND AND MOTIVATION 9
1.1.1 Aim 10
1.1.2 Objectives 10

CHAPTER 2: LITERATURE OVERVIEW 11
2.1 INTRODUCTION TO METABOLOMICS 11
2.1.1 Gas chromatography-mass spectrometry as analytical tool 12
2.2 CHEMICAL DERIVATISATION 13
2.2.1 Chemical derivatisation techniques for GC-MS 13
2.2.1.1 Silylation 14
2.2.1.2 Alkylation 15
2.2.1.3 Combination derivatisation 16
2.2.2 Comparing GC-based derivatisation methods for untargeted metabolomics analysis 16
2.3 COMPOUND SPECIFIC DERIVATISATION 17
2.4 METABOLOMICS DATA ANALYSIS 26

CHAPTER 3: MATERIALS AND METHODS 28
3.1 INTRODUCTION 28
3.2 EXPERIMENTAL DESIGN 28
3.2.1 Ethical approval 31
3.2.2 Standard stock solution 31
3.2.2.1 Reagents and chemicals 31
3.2.2.2 Solutions 31
3.2.2.3 Internal standard preparation: 32
3.2.2.4 Derivatisation reagent preparation: 32

3.2.3 Derivatisation methods 33
3.2.3.1 Silylation 33
3.2.3.2 Combination derivatisation with methoximation and silylation 34
3.2.3.3 Alkylation 34
3.2.4 Biological samples 35
3.2.5 GC-MS analysis 35
3.2.6 GC-MS data processing 37
3.2.6.1 Standard stock solution 37
3.2.6.2 Biological samples 38
3.2.7 Statistical analyses 38
3.2.7.1 Standard stock solution 38
3.2.7.2 Biological samples 38

CHAPTER 4: RESULTS AND DISCUSSION 39
4.1 DATA OVERVIEW PCA SCORE PLOTS 39
4.2 REPEATABILITY DATA OVERVIEW 42
4.3 REPEATABILITY PER DERIVATISATION METHOD 46
4.3.1 BSTFA 46
4.3.2 BSTFA-MeOX 48
4.3.3 MSTFA 50
4.3.4 MSTFA-MeOX 51
4.3.5 MCF 53
4.4 DERIVATISATION METHOD COMPARISON PER COMPOUND 54
4.5 DERIVATISATION RESPONSE OVER TIME 62
4.6 SUMMARISED COMPOUND RESPONSE OVER TIME 66
4.7 INDIVIDUAL COMPOUND RESPONSE OVER TIME 68
4.8 BIOLOGICAL SAMPLES 73

CHAPTER 5: CONCLUSIONS AND FUTURE PROSPECTS 78
5.1 INTRODUCTION 78
5.2 CONCLUSIONS 78
5.3 FINAL REMARKS 80
5.4 FUTURE RECOMMENDATIONS 80
REFERENCES 81

LIST OF FIGURES AND TABLES

List of figures:

Figure 2.1: An overview scheme of silylation reactions using MSTFA (Villas-Bôas *et al.*, 2011).. 15
Figure 2.2: An overview of alkylation reactions using MCF. Adapted from Villas-Bôas *et al.* (2011).. 16
Figure 2.3: A general example of a methoximation reaction using methoxyamine hydrochloride dissolved in pyridine. Adapted from Lai and Fiehn (2016). 16

Figure 3.1: Summary of the general experimental design used to address the aim of this investigation, including: sample preparation, derivatisation, GC-MS analysis, data processing, compound identification and analytical comparison and interpretation............ 30

Figure 4.1: The PCA score plots obtained from 25 samples per derivatisation method analysed by GC-MS. Derivatised by: a) BSTFA; b) BSTFA-MeOX; c) MSFTA; d) MSTFA-MeOX; e) MCF; and f) PCA score plot of all tested methods (only compounds detected with all methods are shown). Batches 1 (●), 2 (X), 3 (■), 4 (▲), 5 (◆)..................................... 41
Figure 4.2: An overview of the average RSD distribution of the sample replicates derivatised by five methods over five days. 43
Figure 4.3: The average (±STEDV) RSD of BSTFA derivatised compounds over five analysed batches (n=25). 47
Figure 4.4: The average RSD (±STDEV) of BSTFA-MeOX derivatised compounds over five analysed batches (n=25). 49
Figure 4.5: The average RSD (±STDEV) of MSTFA derivatised compounds over five analysed batches (n=25). 50
Figure 4.6: The average (±STDEV) RSD of MSTFA-MeOX derivatised compounds over five analysed batches (n=25). 52
Figure 4.7: The average (±STDEV) RSD of MCF derivatised compounds over five analysed batches (n=25). 54
Figure 4.8: The comparison of the average percentage RSD for each compound derivatised by BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF. A total of 25 samples were derivatised per method. 55
Figure 4.9: An overview of the distribution of the average RSD values for all derivatives of the sample replicates of each derivatisation method a) BSTFA; b) BSTFA-MeOX; c) MSFTA; d) MSTFA-MeOX; e) MCF. This response was measured at five time periods (T1: 0h; T2: 12h; T3: 36h; T4: 60h; T5: 84h) as summarised by the PCA score plots and boxplots. 65

Figure 4.10: Average normalised abundance per derivative (as determined from the analyses of five samples) measured over time derivatised by BSTFA (●), BSTFA-MeOX (■), MSTFA (▲), MSTFA-MeOX (▼) and MCF (♦) methods to infer stability. 67
Figure 4.11: Compound stability (average normalised abundance) measured over time derivatised by BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF. 1(●): 0 h; 2(▲): 12 h; 3(♦): 36 h; 4(▼): 60 h; 5(■): 84 h. 69
Figure 4.12: The PCA score plots obtained from analyses of biological samples derivatised with: a) BSTFA; b) BSTFA-MeOX; c) MSFTA; d) MSTFA-MeOX; e) MCF; Group A (■), Group B (●), n=5 per group. 74

List of tables:

Table 2.1: The pre-selected compounds utilised in this study encompassing amino acids and small organic acids by gas chromatography-mass spectrometry analyses. 19
Table 2.2: Derivatisation conditions of internal standards utilised in this study. 25

Table 3.1: Retention times and targeted ions for TMS derivatives used in selective ion monitoring mode.. 36
Table 3.2: Retention times and targeted ions for MCF derivatives used in selective ion monitoring mode.. 37

Table 4.1: BSTFA derivatisation overview of compounds detected, the retention times, average peak area and average RSD (± standard deviation). Compounds with RSD < 20% are highlighted in green, compounds with RSD > 50% and highlighted in red. Compounds are listed in order of detected retention times. 43
Table 4.2: MSTFA derivatisation overview of compounds detected, the retention times, average peak area and average RSD (± standard deviation). Compounds with RSD < 20% are highlighted in green, compounds with RSD > 50% and highlighted in red. Compounds are listed in order of detected retention times. 44
Table 4.3: MCF derivatisation overview of compounds detected, the retention times, average peak area and average RSD (± standard deviation). Compounds with RSD < 20% are highlighted in green, compounds with RSD > 50% and highlighted in red. Compounds are listed in order of detected retention times. 45
Table 4.4: The comparison of the average percentage RSD for each identified compound derivatised by BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF. ▲ Highest RSD; ▼ Lowest RSD; ♦ Between lowest and highest values; - Not detected. Compounds are listed in alphabetical order. 60
Table 4.5: The comparison of the total, unknown and statistically significant features detected, derivatised by BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF. 76

LIST OF ABBREVIATIONS

Abbreviation	Definition	Abbreviation	Definition
ABMBA	2-Amino-3-bromo-5-methylbenzoic acid	MeOX	Methoxyamine hydrochloride solution
AMDIS	Automated Mass spectral Deconvolution and Identification System	MSTFA	N-methyl-(trimethylsilyl) trifluoroacetamide
BSTFA	N, O-bis(trimethylsilyl) trifluoroacetamide	m/z	Mass-to-charge-ratio
C19:0 ME	Nonadecanoic acid methyl ester	NIST	National Institute of Standards and Technology
CV	Coefficient of variance	PCA	Principal component analysis
ECF	Ethyl chloroformate	QC	Quality control
ERNDIM	European Research Network for evaluation and improvement of screening, Diagnosis, and treatment of Inherited disorders of Metabolism	RSD	Relative standard deviation
GC	Gas chromatography	SIM	Selective ion monitoring
IS	Internal standard	TMCS	Trimethylchlorosilane
MS	Mass spectrometry	TMS	Trimethylsilyl
MCF	Methyl chloroformate		

Chemical formula	Definition	Chemical formula	Definition
$NaHCO_3$	Sodium bicarbonate	Na_2SO_4	Sodium sulphate
NaOH	Sodium hydroxide		

Units	Definition	Units	Definition
˚C	Degrees Celsius	µL	Microlitre
g	Gravitational force	mL	Millilitre
h	Hours	min	Minutes
m	Meter	mg	Milligram
µm	Micrometre	sec	Seconds

Web servers
AMDIS: www.amdis.net; NIST: https://www.nist.gov/, MetaboAnalyst: https://www.metaboanalyst.ca/

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

Investigating the metabolome has become an important part of many research areas (Parkinson, 2012; Lkhagva *et al.*, 2020). In most cases the selection of an appropriate sample preparation method is directly linked to the quality outcome of the metabolomics analyses (Wang *et al.*, 2017). Gas chromatography coupled to mass spectrometry (GC-MS) is one of the most popular hyphenated analytical techniques used to perform metabolomics analyses, given its high separation efficiency of complex biological samples and cost-effectiveness (Villas-Bôas *et al.*, 2005; Dayalan *et al.*, 2019; Feizi *et al.*, 2021; Putri *et al.*, 2022). However, only volatile compounds or those that can be made volatile can be analysed by GC-MS, therefore chemical derivatisation is an important step of the sample preparation process (Knapp, 1979; Smart *et al.*, 2010; Villas-Bôas *et al.*, 2011; Parkinson, 2012; Bhumireddy *et al.*, 2021). This process aims to improve the chromatographic properties of the analytes by reducing their polarity and increasing their thermal stability and volatility (Schummer *et al.*, 2009; Orata, 2012; Lin *et al.*, 2016; Khodadadi & Pourfarzam, 2020). Currently, great variability exists in literature in terms of sample derivatisation methods for untargeted metabolomics. Various methods for chemical derivatisation exist, but only a few, such as silylation, acylation, and alkylation are routinely employed (Villas-Bôas *et al.*, 2005; Khodadadi & Pourfarzam, 2020). The analytical technique, as well as the nature of the target compounds determine the choice of method (Parkinson, 2012).

Silylation is a commonly used GC-MS derivatisation technique that converts compounds into a more suitable derivative (Schummer *et al.*, 2009; Valdez & Leif, 2021). Silyl reagents are suitable for a wide variety of functional groups and are generally safe and easy to use (Glicksberg & Kerrigan, 2020; Mojsak *et al.*, 2020). Reagents such as N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) or N-methyl-(trimethylsilyl)trifluoroacetamide (MSTFA) are some of the most commonly used silylating reagents (Glicksberg & Kerrigan, 2020; Khodadadi & Pourfarzam, 2020). Then again, alkylation can be utilised as a derivatisation method, using methyl chloroformate (MCF) as a derivatising reagent (Villas-Bôas *et al.*, 2011). Similarly to silylation, an active hydrogen is replaced, but with an alkyl or aryl group (Parkinson, 2012; Glicksberg & Kerrigan, 2020). Combined derivatisation is often employed where two types of derivatisation methods can enhance the chromatographic separation or peak shape. Mostly employed is a methoximation step, prior to silylation (Marcos & Pozo, 2015), using a solution such as methoxyamine hydrochloride (MeOX) in pyridine, typically to prevent the formation of cyclic structures and to reduce the number of possible stereoisomers per compound, while also increasing detection (Fritsche-Guenther *et al.*, 2021).

Choosing the most suitable derivatisation reagent is not always a simple task, as the formation of multiple derivatives, the stability of compounds, the incomplete derivatisation and the degradation of compounds greatly differs between metabolites and samples. For the largest part metabolomics studies in literature have concentrated on the application of a single derivatisation method or reagent, and neglected investigations into the comparison between techniques relating to specific compounds present in the sample of interest. As metabolomics research necessitates the analysis of a large number of samples, a well-chosen optimised derivatisation procedure is required to ensure reproducible results. Ideally, a derivatisation procedure should be rapid, quantitative and produce minimal by-products. Keeping in mind that no universally accepted derivatisation protocol exists for GC-MS metabolomics (Moros *et al.*, 2017) this investigation supports new insight into the response of selected metabolites, derivatised using different reagents. The reagents of choice for this study included BSTFA and MSTFA with and without MeOX, as these are the most commonly used silylation methods from literature. Additionally, an alkylation reaction was included using MCF, to determine if there is a benefit of switching to a less popular reagent. It is envised that the findings from this study will facalitate decision making in future studies aimed at GC-MS metabolomics analyses.

1.1.1 Aim

The aim of this investigation was to compare pre-selected derivatisation techniques on a set of key metabolites including a range of amino acids and small organic acids as a precursor of gas chromatography-mass spectrometry metabolomics analyses.

1.1.2 Objectives

Considering the above-mentioned aim, the objectives of the study include:

1. A literature study of the current derivatisation techniques applied for untargeted metabolomics research incorporating specific metabolite classes.
2. Preparation of individually selected compounds representative of different functional groups and corresponding detection of the specific trimethylsilyl (TMS) and methyl chloroformate (MCF) derivatives.
3. Derivatising, analysing and comparing pre-aliquoted sample stock solutions using BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF derivatisation techniques.
4. Determination of derivative stability over time between derivatisation techniques by reanalysing the same batches.
5. Derivatising, analysing and comparing two sets of biological samples using BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF derivatisation techniques.

CHAPTER 2: LITERATURE REVIEW

2.1 INTRODUCTION TO METABOLOMICS

The metabolome was first defined in 1998 as all the metabolites present in any given sample (Oliver *et al.*, 1998; Gao *et al.*, 2010; Burgess *et al.*, 2014; Tan *et al.*, 2016; Adamski, 2020; Muthubharathi *et al.*, 2021). Metabolomics, which is the study of the metabolome, is the ultimate point of the omics cascade (Roessner & Bowne, 2009; Pinu *et al.*, 2017). It aims to quantitatively and qualitatively analyse a comprehensive group of metabolic compounds in a biological system (Kaspar *et al.*, 2008; Manchester & Anand, 2017; Pinu *et al.*, 2017; Miggiels *et al.*, 2019; Villaret-Cazadamont *et al.*, 2020; Ashrafian *et al.*, 2021). This is a rapidly emerging field and has been applied to a diverse range of research areas, such as medicine, agriculture, food science and environmental science (Pinu *et al.*, 2017; Dayalan *et al.*, 2019; Adamski, 2020; Zeki Ö *et al.*, 2020).

Depending on the research question and context of the study, metabolomics can be executed in an untargeted or targeted manner (Zheng *et al.*, 2018; Pereira Braga & Adamec, 2019; Wang *et al.*, 2019; Engel *et al.*, 2020). In an untargeted approach, the goal is to measure as many metabolites as possible, hence it focuses on the comprehensive characterisation of the metabolome in an unbiased fashion (Johnson *et al.*, 2016; Wang *et al.*, 2019; Villaret-Cazadamont *et al.*, 2020). This type of approach can be challenging due to the resulting complex data set and identification or characterisation of unknown metabolites (Johnson *et al.*, 2016; Pereira Braga & Adamec, 2019). In contrast, a targeted approach is used to detect and quantify a single or a panel of specific metabolic compounds that are pre-selected (León *et al.*, 2013; Pereira Braga & Adamec, 2019; Wang *et al.*, 2019). However, to predict the unknown concentration of a target compound, a calibration curve has to be established through the use of external and/or internal calibrators (Rimayi *et al.*, 2015; Tumanov *et al.*, 2016; Pino *et al.*, 2018). External calibrators or standards are separate from the test sample, while internal calibrators or standards are spiked into the same test sample (Pino *et al.*, 2018). Consequently, a more sensitive and precise analysis can be achieved (Johnson *et al.*, 2016).

Metabolic compounds possess a vast variety of physical and chemical properties (Dayalan *et al.*, 2019). Therefore, suitable analytical platforms are required to analyse a collection of metabolites simultaneously (Roessner & Bowne, 2009; Engel *et al.*, 2020; Zeki Ö *et al.*, 2020; Paiva *et al.*, 2021). This includes nuclear magnetic resonance (NMR) and mass spectrometry (MS)-based platforms (Dayalan *et al.*, 2019; Ashrafian *et al.*, 2021). Hyphenated mass spectrometers are typically coupled to chromatography separation techniques, such as liquid chromatography (LC) or gas chromatography (GC) (Manchester & Anand, 2017; Dayalan *et al.*, 2019; Miggiels *et al.*, 2019; Adamski, 2020; Ashrafian *et al.*, 2021; Muthubharathi *et al.*,

2021). Commonly, LC-MS is used for the analyses of non-volatile metabolites, typically, without the application of chemical derivatisation. Chemical derivatisation can have other functions when used with LC-MS, such as improving the extraction efficiency or the detectability property of a compound. Then again, GC-MS is one of the most popular analytical platforms used in metabolomics studies. However, chemical derivatisation of the sample is generally required to increase the volatility of non-volatile metabolites (Parkinson, 2012).

2.1.1 Gas chromatography-mass spectrometry as analytical tool

The combination of two powerful analytical tools, GC and MS, is often seen as a "gold standard" platform in laboratories (Kyle, 2017). Hence, GC-MS continues to play an important role in metabolomics studies (Maštovská & Lehotay, 2003; Santos & Galceran, 2003; Zhang, 2012a; Putri *et al.*, 2022). Gas chromatography is an outstanding separating technique (Pasikanti *et al.*, 2008; Tan *et al.*, 2016) based on the partitioning behaviour of the specific compounds between a gas phase and liquid phase when they move through the GC capillary column at a given temperature (Bartle & Myers, 2002; Pereira Braga & Adamec, 2019). Provided that mostly only volatile compounds can be analysed by GC, derivatisation needs to be employed to volatilise semi-volatile and non-volatile compounds (Zhang, 2012a; Pereira Braga & Adamec, 2019). Mass spectrometry is an analytical technique that aids in compound identification by providing information on the analysed compounds (Hites, 1997). A mass spectrometer consists of three integrated parts, including an ionisation source, mass analyser and detector (Pereira Braga & Adamec, 2019). Hard ionisation, such as electron impact, is commonly used with GC (Santos & Galceran, 2003; Burgess *et al.*, 2014). Consequently, this is a destructive analytical process that uses a lot of energy to produce gas phase ions by fragmentation of molecular species (Burgess *et al.*, 2014; Wishart, 2019; Ashrafian *et al.*, 2021). The produced ions are then separated by the mass analyser and measured according to their mass-to-charge [m/z] ratios by the detector (Tan *et al.*, 2016; Špánik & Machyňáková, 2018; Pereira Braga & Adamec, 2019; Ashrafian *et al.*, 2021). For untargeted analysis the detector may be set on full-scan mode, while selective ion monitoring (SIM) can be used for targeted analysis. By applying SIM, the analytical sensitivity can be increased since the instrument is set to acquire only a number of specific ions per second at a specific retention time (Pino *et al.*, 2018).

Reproducibility, peak resolution, high sensitivity, robustness and cost-effectiveness of GC-MS, has made it a very attractive and popular analytical technique (Santos & Galceran, 2003; Lisec *et al.*, 2006; Dettmer *et al.*, 2007; Pasikanti *et al.*, 2008; Burgess *et al.*, 2014; Pinu *et al.*, 2017; Dayalan *et al.*, 2019; Miyagawa & Bamba, 2019; Feizi *et al.*, 2021; Putri *et al.*, 2022). Furthermore, as mentioned, the mass spectra produced by EI fragmentation facilitate the

identification of compounds by comparing it to spectral libraries (Santos & Galceran, 2003; Pasikanti *et al.*, 2008; Ashrafian *et al.*, 2021). The most commonly used database is the National Institute of Standards and Technology (NIST) (Burgess *et al.*, 2014; Putri *et al.*, 2022). However, since derivatisation is required to make analysis of semi- and non-volatile compounds possible, sample preparation for GC-MS analyses may be time-consuming (Pasikanti *et al.*, 2008; Burgess *et al.*, 2014; Kyle, 2017; Dayalan *et al.*, 2019; Miyagawa & Bamba, 2019; Pereira Braga & Adamec, 2019; Putri *et al.*, 2022). Moreover, GC-MS is limited to low-molecular weight metabolites (Tan *et al.*, 2016). Hence, to increase the metabolite coverage, GC-MS may have to be used in conjunction with complementary techniques (Lisec *et al.*, 2006; Dettmer *et al.*, 2007; Feizi *et al.*, 2021). Yet, GC-MS remains to be an exceptional technique despite some limitations.

2.2 CHEMICAL DERIVATISATION

Sample preparation for GC-MS analyses typically includes chemical derivatisation (Wilson & Walker, 2005; Parkinson, 2012; Su *et al.*, 2017; Vilbaste *et al.*, 2020). Derivatisation can be defined as a reaction that primarily changes the chemical structure of an analyte, ultimately aiming to enhance the volatility and gas chromatographic properties of the analyte (Lord & Pfannkoch, 2012; Orata, 2012; Kyle, 2017; El-Maghrabey *et al.*, 2020; Paiva *et al.*, 2021). These properties include: increased separation, detection, and thermal stability (Parkinson, 2012; Řezanka *et al.*, 2016; Moldoveanu & David, 2018; Munir & Badri, 2020). This is advantageous, given that many analytes do not possess the chemical or structural properties necessary for GC analysis (Sajid & Płotka-Wasylka, 2018). However, derivatisation adds an extra step to sample preparation which can often be regarded as a source of variation (Rompa *et al.*, 2003; Xu, 2009; Ferreira *et al.*, 2013; Poole, 2013; Moldoveanu & David, 2018). Additionally, it could lead to analyte losses, the occurrence of side products and the formation of undesired artefacts or decomposition products (Marcos & Pozo, 2015; Moldoveanu & David, 2018; Sajid & Płotka-Wasylka, 2018; Glicksberg & Kerrigan, 2020; Munir & Badri, 2020). Nonetheless, the benefits, such as improved analyte detection, stability and chromatographic behaviour, obtained by derivatisation overshadows the drawbacks (Han *et al.*, 2011; Moldoveanu & David, 2018; Kranenburg *et al.*, 2020; Munir & Badri, 2020).

2.2.1 Chemical derivatisation techniques for GC-MS

Various methods of chemical derivatisation exist, but only a few well-standardised techniques such as silylation, acylation, and alkylation are routinely employed (Villas-Bôas *et al.*, 2005; Moldoveanu & David, 2018; El-Maghrabey *et al.*, 2020; Khodadadi & Pourfarzam, 2020). Using multiple derivatisation methods can be beneficial, because having more than one derivatised version of an analyte supports strong, evidence of its presence in a matrix (Valdez & Leif,

2021). Yet, the choice of the method is highly dependent on the analytical technique and the nature of compounds analysed (Parkinson, 2012). Since derivatisation is often a necessary sample preparation step prior to GC-MS analysis, it is important to familiarise oneself with the different derivatisation methods in order to improve or develop analytical methods suitable for the samples at hand (Orata, 2012).

2.2.1.1 Silylation

The most frequently used silylation method, namely trimethylsilylation, is based on the principle that trimethylsilyl (TMS) derivatives are formed when the TMS group replaces an acidic hydrogen on a polar analyte (Figure 2.1), which converts it into a suitable product for analysis by GC-MS (Schummer *et al.*, 2009; Moldoveanu & David, 2018; Valdez & Leif, 2021). The prevalent use of silyl reagents are due to the reagents suitability for a wide variety of functional groups, and it is also user-friendly (Glicksberg & Kerrigan, 2020; Mojsak *et al.*, 2020). This method usually does not require a purification step after derivatisation and can be directly injected into the GC-MS (Schummer *et al.*, 2009; Parkinson, 2012). Additionally, the silylating group tag can aid in structural identification by facilitating the fragmentation patterns (Parkinson, 2012). However, it should be taken into consideration that these reagents are moisture sensitive, require anhydrous reaction conditions and likely heating (Lin *et al.*, 2008; Villas-Bôas *et al.*, 2011; Parkinson, 2012; Glicksberg & Kerrigan, 2020; Mojsak *et al.*, 2020; Bhumireddy *et al.*, 2021). Generally, silylation methods are regarded as methods with lower reproducibility, since the derivatives produced are relatively unstable. Also, the excess or unreacted silylated reagent and non-derivatised involatile compounds may be detrimental to the GC-capillary column when directly injected into it (Villas-Bôas *et al.*, 2011; Glicksberg & Kerrigan, 2020; Mojsak *et al.*, 2020). Yet, silylation remains one of the most commonly used GC-MS derivatisation methods (Halket *et al.*, 1999; Fiehn, 2000; Halket & Zaikin, 2003; Gullberg *et al.*, 2004; Yegles *et al.*, 2004; Villas-Bôas *et al.*, 2005; Villas-Bôas *et al.*, 2006; Villas-Bôas *et al.*, 2007; Roessner & Bowne, 2009; Mojsak *et al.*, 2020; Valdez & Leif, 2021). Reagents such as N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) or N-methyl-(trimethylsilyl)trifluoroacetamide (MSTFA) are some of the most commonly used silylating reagents (Glicksberg & Kerrigan, 2020; Khodadadi & Pourfarzam, 2020). Of these two reagents, MSTFA is more reactive, resulting in a faster and more complete derivatisation reaction (Parkinson, 2012; Zhu *et al.*, 2017). Moreover, MSTFA is more volatile than BSTFA (Khodadadi & Pourfarzam, 2020; Mojsak *et al.*, 2020). Trimethylchlorosilane (TMCS) can also be added to silylation reagents to act as a catalyst (Zhao *et al.*, 2017; Mojsak *et al.*, 2020; Bhumireddy *et al.*, 2021).

Figure 2.1: An overview scheme of silylation reactions using MSTFA (Villas-Bôas *et al.*, 2011).

2.2.1.2 Alkylation

During alkylation, an active hydrogen is replaced with an alkyl or aryl group (Figure 2.2), instead of a trimethylsilyl group, as is experienced when using silylation. Therefore, alkylation can be used as an alternative derivatisation method (Parkinson, 2012; Glicksberg & Kerrigan, 2020). The derivatives produced by this reaction have much greater stability than TMS derivatives, making it possible to store derivatised samples for up to 30 days (Bhumireddy *et al.*, 2021), which results in higher experimental reproducibility (Parkinson, 2012; Bhumireddy *et al.*, 2021; Casas & Matamoros, 2021). Moreover, this method is fast and uses mild reaction conditions where a heating step is excluded (Wang *et al.*, 2011). Alkylation also creates more favourable conditions for polar compounds than silylation, since the reaction can be performed in the presence of water (Moldoveanu & David, 2018; Kim *et al.*, 2019). Nevertheless, alkylation is a multi-step procedure, which could add additional variation between samples. Alkylating reagents may also be partially biased or limited towards certain functional groups, therefore idealistically it should be used in combination with silylation derivatisation according to Villas-Bôas *et al.* (2011). One popular reagent for alkylation is methyl chloroformate (MCF) (Smart *et al.*, 2010; Tumanov *et al.*, 2016; Casas & Matamoros, 2021). Methyl chloroformate would be ideal for studying the central carbon metabolism since amino acid and organic acid derivatives produced are more stable than when silylation derivatisation is used. Villas-Bôas *et al.* (2011), showed that MCF derivatisation enabled the analysis of over a hundred amino and non-amino organic acids simultaneously.

Figure 2.2: An overview of alkylation reactions using MCF. Adapted from Villas-Bôas *et al.* (2011).

2.2.1.3 Combination derivatisation

A combination of derivatisation methods is also frequently implemented, for example, methoximation using methoxyamine hydrochloride solution (MeOX) before silylation is often added to protect alpha-keto acids and aldehydes against decarboxylation and keto-enol tautomerism (Figure 2.3). It also inhibits the cyclisation of open chain and cyclic structures of sugars, thereby preventing the appearance of multiple peaks belonging to each individual sugar compound (Pasikanti *et al.*, 2008; Miyagawa & Bamba, 2019). An advantage of this extra step includes the simplification of obtained chromatograms as it prevents the alpha-keto acids and aldehydes to interconvert to their structural isomers, which is known as keto-enol tautomerisation (Parkinson, 2012; Glicksberg & Kerrigan, 2020; Khodadadi & Pourfarzam, 2020). However, methoximation adds extra time to the sample preparation process and creates an extra step for potential variation.

Figure 2.3: A general example of a methoximation reaction using methoxyamine hydrochloride dissolved in pyridine. Adapted from Lai and Fiehn (2016).

2.2.2 Comparing GC-based derivatisation methods for untargeted metabolomics analysis

Several chemical derivatisation methods have been developed and compared over the years with the aim of untargeted metabolomic analyses in mind. In the late 20th century Roessner *et al.* (2000) and co-workers found that MSTFA compared to BSTFA, bis(trimethylsilyl)acetamide, trimethylsilyl imidazole and hexamethyl disilazane, gave the best results in terms of the

chemical compound range and production of by-products. Hence, MSTFA was used as derivatisation reagent for further investigation and facilitated the identification of 77 compounds in potato tuber extracts. Furthermore, Gao *et al.* (2009) illustrated that with the use of BSTFA, 133 compounds (such as carbohydrates, polyols, bile acids, and other nitrogen containing heterocyclic compounds) could be identified, while with the use of ethyl chloroformate (ECF) only 73 compounds (such as amino acids, phenolics, and carboxylic acids) could be recovered. Similarly, Villas-Bôas *et al.* (2011) compared silylation and alkylation techniques concluding that the use of MCF as derivatisation reagent is better for the analysis of organic acids, nucleotides and polyfunctional amines. In addition, alkylation derivatisation showed higher reproducibility and stability. Moreover, different silylation reagents were tested by Moros *et al.* (2017), which concluded that the use of MSTFA and MSTFA-TMCS detected six more compounds and resulted in higher detection signals and better repeatability. Interestingly, when methoximation prior to silylation was tested by López-Bascón *et al.* (2019), they found that silylation resulted in the identification of 123 compounds, while methoximation prior to silylation only resulted in the identification of 99 compounds. Likewise, Engel *et al.* (2020) demonstrated that methoximation prior to silylation leads to significantly lower signal responses of compounds, though the complexity of the derivatised products is reduced. Lastly, Bhumireddy *et al.* (2021) came to the conclusion that ECF derivatisation is the best choice compared to BSTFA and MSTFA, since their chromatograms had a higher amino acid abundance and also included fatty acids and tricarboxylic acid cycle metabolites. Given the wide variety of derivatisation reagents and the importance of this step in GC-MS analysis, investigating a suitable reagent could facilitate better decision making when a derivatisation method or reagent is chosen for a specific study. The reagents of choice for this study included BSTFA and MSTFA with and without MeOX, as these are the most commonly used silylation methods from literature. Additionally, an alkylation reaction was included using MCF, to determine if there is a benefit of switching to a less popular reagent.

2.3 COMPOUND SPECIFIC DERIVATISATION

Although the analytical technique influences the choice of derivatisation method, the nature of the compounds of interest also plays a major role in this decision (Parkinson, 2012). This is true, given that derivatisation methods have been accused of being biased towards certain functional groups (Engel *et al.*, 2020). For example, silylation is thought to produce more stable carbohydrate derivatives (Villas-Bôas *et al.*, 2005; Villas-Bôas *et al.*, 2006), while alkylation is better suited for the derivatisation of amino- and organic acids (Villas-Bôas *et al.*, 2003; Smart *et al.*, 2010; Villas-Bôas *et al.*, 2011).

Therefore, for the purpose of this study different derivatisation methods were investigated from literature focusing on GC-MS metabolomics investigations of specific, pre-selected compounds, encompassing amino acids and small organic acids. In brief, the compounds alanine, aspartic acid, citric acid, cystine, fumaric acid, glutamic acid, glutamine, glycine, 2-hydroxybutyric acid, isocitric acid, alpha(α)-ketoglutaric acid, lactic acid, lysine, methionine, coumaric acid, phenylalanine, succinic acid, tryptophan and valine, were chosen for this derivatisation comparison as demonstrated in Table 2.1. The following compounds, 2-amino-3-bromo-5-methylbenzoic acid (ABMBA), 3-phenylbutyric acid, nonadecanoic acid methyl ester (C19:0 ME) and norleucine were chosen as internal standards (see

Table 2.2). The choice of internal standards were influenced by: Willers *et al.* (2016), Venter *et al.* (2016) and Swenson and Northen (2019). The compound 3-phenylbutyric acid is absent in normal biological samples and co-elutes with few other compounds, making it well suited to indicate derivatisation efficiency (Reinecke *et al.*, 2012). The same is true for norleucine, yet at a different stage of the chromatographic separation. Methylated nonadecanoic acid will not change due to derivatisation conditions as this is already methylated and does not undergo silylation or methoximation (Fiehn, 2016). Lastly, ABMBA was added as an extra normalising factor to compensate for potential end-volume differences, but also for quality control purposes of each run.

Table 2.1: The pre-selected compounds utilised in this study encompassing amino acids and small organic acids by gas chromatography-mass spectrometry analyses.

Compound	Structure	Derivatisation reagent	MeOX	Derivatisation Condition	Reference
Alanine		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	
		BSTFA and MSTFA	With	25˚C for 16 hours (methoximation) 25˚C for 1 hour (silylation)	(Moros *et al.*, 2017)
Aspartic acid		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	
		BSTFA	Without	60˚C for 1 hour	(Bhumireddy *et al.*, 2021)
		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	
Citric acid		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	
		BSTFA and MSTFA	With	25˚C for 16 hours (methoximation) 25˚C for 1 hour (silylation)	(Moros *et al.*, 2017)
		MSTFA	Both	37˚C for 90 min (methoximation) 37˚C for 30 min (silylation)	(Engel *et al.*, 2020

Table 2.1: Continued.

Coumaric acid		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	
Cystine		BSTFA	Without	135˚C for 4 hours (silylation)	(Sobolevsky *et al.*, 2003)
		MSTFA	With	Room temperature for 16 hours (methoximation) Room temperature for 1 hour (silylation)	(Liu *et al.*, 2021)
Fumaric acid		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	
		BSTFA	Without	60˚C for 1 hour (silylation)	(Bhumireddy *et al.*, 2021)
		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	
Glutamic acid		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	
		BSTFA and MSTFA	With	25˚C for 16 hours (methoximation) 25˚C for 1 hour (silylation)	(Moros *et al.*, 2017)
		MSTFA	Both	37˚C for 90 min (methoximation) 37˚C for 30 min (silylation)	(Engel *et al.*, 2020)

Table 2.1: Continued.

Glutamine		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
Glutamine		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	(Villas-Bôas *et al.*, 2011)
Glutamine		MSTFA	Both	37˚C for 90 min (methoximation) 37˚C for 30 min (silylation)	(Engel *et al.*, 2020)
Glutamine		BSTFA	Without	60˚C for 1 hour (silylation)	(Bhumireddy *et al.*, 2021)
Glutamine		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Bhumireddy *et al.*, 2021)
Glycine		BSTFA and MSTFA	With	25˚C for 16 hours (methoximation) 25˚C for 1 hour (silylation)	(Moros *et al.*, 2017)
Glycine		MSTFA	Both	37˚C for 90 min (methoximation) 37˚C for 30 min (silylation)	(Engel *et al.*, 2020)
2-Hydroxybutyric acid		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
2-Hydroxybutyric acid		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	(Villas-Bôas *et al.*, 2011)
2-Hydroxybutyric acid		BSTFA	Without	60˚C for 1 hour (silylation)	(Bhumireddy *et al.*, 2021)
2-Hydroxybutyric acid		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Bhumireddy *et al.*, 2021)
Isocitric acid		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
Isocitric acid		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	(Villas-Bôas *et al.*, 2011)

Table 2.1: Continued.

α-Ketoglutaric acid		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	
		MSTFA	Both	37˚C for 90 min (methoximation) 37˚C for 30 min (silylation)	(Engel *et al.*, 2020)
Lactic acid		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	
		BSTFA and MSTFA	With	25˚C for 16 hours (methoximation) 25˚C for 1 hour (silylation)	(Moros *et al.*, 2017)
Lysine		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	
		MSTFA	Both	37˚C for 90 min (methoximation) 37˚C for 30 min (silylation)	(Engel *et al.*, 2020)
		BSTFA	With	60˚C for 1 hour (silylation)	(Bhumireddy *et al.*, 2021)
		MSTFA	Without	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	
Methionine		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	

Table 2.1: Continued.

Phenylalanine		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	
		BSTFA	Without	60˚C for 1 hour (silylation)	(Bhumireddy *et al.*, 2021)
		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	
Succinic acid		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	
		BSTFA and MSTFA	With	25˚C for 16 hours (methoximation) 25˚C for 1 hour (silylation)	(Moros *et al.*, 2017)
Tryptophan		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
		MCF	Without	60 seconds while vortexing at room temperature (alkylation)	
		BSTFA	With	60˚C for 60 min (methoximation) 60˚C for 60 min (silylation)	(Venter *et al.*, 2016)
		BSTFA	Without	60˚C for 1 hour (silylation)	(Bhumireddy *et al.*, 2021)
		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	

Table 2.1: Continued.

Valine		MSTFA	With	Microwave for 2.8 min incubation (methoximation) Microwave for 3.0 min incubation (silylation)	(Villas-Bôas *et al.*, 2011)
		MC	Without	60 seconds while vortexing at room temperature (alkylation)	
		BSTFA and MSTFA	With	25˚C for 16 hours (methoximation) 25˚C for 1 hour (silylation)	(Moros *et al.*, 2017)

Table 2.2: Derivatisation conditions of internal standards utilised in this study.

Compound	Structure	Derivatisation reagent	MeOX	Derivatisation Condition	Reference
2-Amino-3-bromo-5-methylbenzoic acid	Br, CH_3, O, OH, NH_2	MSTFA	With	30˚C for 90 min (methoximation) 37˚C for 30 min (silylation)	(Swenson *et al.*, 2015)
3-Phenylbutyric acid	O, OH, H_3C	BSTFA	With	60˚C for 60 min (methoximation) 60˚C for 60 min (silylation)	(Venter *et al.*, 2016)
Nonadecanoic acid methyl ester	H_3C, O, O, H_3C	BSTFA	With	60˚C for 60 min (methoximation) 60˚C for 60 min (silylation)	(Willers *et al.*, 2016)
Norleucine	H_2N, O, OH, H_3C	BSTFA	With	60˚C for 60 min (methoximation) 60˚C for 60 min (silylation)	(Venter *et al.*, 2016)

2.4 METABOLOMICS DATA ANALYSIS

Large amounts of complex data are generated during untargeted metabolomics studies, given that the goal of such an approach is to measure as many metabolites as possible (Alonso *et al.*, 2015; Johnson *et al.*, 2016; Wang *et al.*, 2019; Villaret-Cazadamont *et al.*, 2020). Hence, several data analysis tools are required to extract the relevant information from the raw data generated by the GC-MS analyses (Hendriks *et al.*, 2011; Li *et al.*, 2018). Raw data extraction consists of three core steps, namely, peak detection, deconvolution and alignment of all samples (Fiehn, 2000; Luies & Loots, 2016). For GC-MS data this can be carried out by using Automated Mass spectral Deconvolution and Identification System (AMDIS), while feature annotation or identification can be done by comparison to metabolomics libraries or databases (van den Berg *et al.*, 2006; Xia *et al.*, 2009; Bartel *et al.*, 2013; Schrimpe-Rutledge *et al.*, 2016). However, it should be noted that all features are not always metabolites, they can also be redundant signals, artifacts or contaminants (Schrimpe-Rutledge *et al.*, 2016; Peisl *et al.*, 2018; Sindelar & Patti, 2020).

After completion of raw data extraction, data pre-processing usually follows to remove as much of the technical variance as possible, ideally leaving only biological variation in the dataset that will be used to perform statistical analysis (van den Berg *et al.*, 2006; Hendriks *et al.*, 2011; Alonso *et al.*, 2015; Ren *et al.*, 2015). This consists of data normalisation, zero filtering and replacement of zeros and missing values (Lamichhane *et al.*, 2018). Data normalisation can be done by the use of an internal standard (IS), which adds to the elimination of unwanted variations caused by technical errors, and makes samples quantitatively more comparable (Xia *et al.*, 2009; Li *et al.*, 2016; Karaman, 2017). Furthermore, zero filtering aims to remove metabolites that were not detected in a sub-section of the samples in all the groups (Smuts *et al.*, 2013; Lamichhane *et al.*, 2018). The replacement of zeros or missing values is based on the belief that certain compounds are present in the sample, but undetected due to low abundance (Xia *et al.*, 2009; Schoeman & du Preez, 2012; Armitage *et al.*, 2015). Typically, missing value imputation is done by replacing the missing value with half of the minimum of the lowest non-missing value in the dataset (Chong *et al.*, 2019).

Foregoing statistical analysis, data pre-treatment methods are executed, including centering, scaling or transformation to allow for appropriateness and comparability of data (van den Berg *et al.*, 2006; Antonelli *et al.*, 2019). Centering compensates for differences in the offset between high and low abundant metabolite concentrations (van den Berg *et al.*, 2006). Scaling adjusts for differences in metabolite concentration levels that originated from differences in the average abundance of these metabolites and ultimately enlarges the significance of low abundant metabolites (van den Berg *et al.*, 2006; Hendriks *et al.*, 2011; Antonelli *et al.*, 2019).

Transformation is required to correct for heteroscedasticity (the unequal scatter of standard deviations of measurements around zero) and skewness of the data (van den Berg *et al.*, 2006; Xia *et al.*, 2009; Karaman, 2017; Antonelli *et al.*, 2019).

Once the complete set of metabolic features is generated, those that best explain the variance between groups can be obtained by performing univariate and/or multivariate statistical analyses, typically using an online webserver like MetaboAnalyst (5.0). The Student's *t*-test, or analysis of variance, can be applied as univariate analysis (Alonso *et al.*, 2015). This type of analysis gives a preliminary overview of features that may be statistically significant in discriminating between two or more groups (Xia *et al.*, 2009; Bartel *et al.*, 2013; Saccenti *et al.*, 2014; Lamichhane *et al.*, 2018; Chong *et al.*, 2019). Multivariate methods can be categorised into unsupervised and supervised methods. Unsupervised methods make correlations or detect patterns within the data without any information (sample labels) on the study samples (Alonso *et al.*, 2015). The most frequently used method is principal component analysis (PCA) (Wang *et al.*, 2017). This method is based on the linear transformation of metabolic features and aims to reduce dimensionality by summarising the data into principal components (Xia *et al.*, 2009; Smilde *et al.*, 2010; Saccenti *et al.*, 2014). This enables the visualisation of natural separation between groups if present (Jolliffe, 2005; van den Berg *et al.*, 2006; Loots *et al.*, 2013). The first component explains the maximum variance, while the variance explained reduces with each subsequent component (Alonso *et al.*, 2015; Lamichhane *et al.*, 2018). In contrast, supervised methods, such as partial least squares-discriminant analysis (PLS-DA), make use of sample labels to maximise the separation between groups and identify the metabolic features that carry most of the group separating information. (Bartel *et al.*, 2013; Grace & Hudson, 2017; Antonelli *et al.*, 2019)

Even when sample preparation conditions are kept constant and data processing is executed in the same manner for all batches, good repeatability is difficult to obtain when derivatising different batches, as the time between the completion of the derivatisation varies amongst metabolites. Low repeatability is often attributed to incomplete derivatisation (Erarpat *et al.*, 2020). Resultantly variation in peak areas occur due to differences in derivatisation efficiency and stability affecting the overall repeatability of the method at hand (Miyagawa & Bamba, 2019). Repeatability and peak response are used as main measures of derivatisation performance in this study as will be discussed in the succeeding chapters.

CHAPTER 3: MATERIALS AND METHODS

3.1 INTRODUCTION

The choice of sample matrix during a metabolomics investigation is very important (Griffin, 2006), since different matrices contain different types of metabolites and will therefore affect the results obtained. Hence, to assess the derivatisation of specific compounds by alkylation and silylation derivatisation methods, a standard stock solution with selected representative compounds from different functional groups was prepared, pre-aliquoted, derivatised and analysed via GC-MS. The initial use of standards as sample of interest helps to exclude possible interference with the derivatisation procedure or analysis and aims to make alkylation and silylation derivatisation more comparable than when using a matrix (Koek *et al.*, 2006).

3.2 EXPERIMENTAL DESIGN

An overview of the experimental approach followed in this study is outlined in Figure 3.1. This study consisted of three main parts, **part A**: Preparing of individual compounds, **part B**: Preparing of a stock solution containing all the pre-selected compounds and **part C**: analysing of pre-extracted biological samples. In brief each part will be described below.

Part A: The individual pre-selected compounds (section 3.2.2.2) were prepared in 1 mM solutions followed by a drying step. Each compound was derivatised via the five selected derivatisation methods (BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF), followed by GC-MS analyses. For both TMS and MCF derivatisation methods the identity of the compound was confirmed, and a corresponding retention time was assigned. The information obtained within part A was used to standardise the methods that were used to analyse the samples of parts B and C.

Part B: A stock solution was prepared from the pre-selected individually prepared compounds, by transferring them all together. Next the mixed stock solution which was aliquoted by a Hamilton automated liquid handler in sub-sections of 200 µL for MCF evaluations and 100 µL for TMS evaluations. The samples were frozen and freeze-dried, resulting in replicate samples with a starting concentration of 1 mM (0.05 mM on column concentration). For each derivatisation method (BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF), five replicates were prepared, and analysed via GC-MS. This was repeated over five days, resulting in a total of 25 samples analysed per derivatisation method. Additionally, the last batch of each method, was used to determine compound stability once derivatised. This batch (of five samples) was re-injected at 12, 36, 60 and 84 hours after derivatisation of the first batch was completed (time 0). Following completion of the analyses, data processing was

conducted. Next, the data were normalised using the internal standard 3-phenylbutyric acid, followed by statistical analyses and technical comparisons.

Part C: Pre-extracted ERNDIM quality control (QC) serum samples of two cohorts (Group A, n=5, Group B, n =5) were derivatised using the five different derivatisation methods (BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF). The data were processed via in-house methods resulting in a data matrix that was subjected to statistical analyses.

.

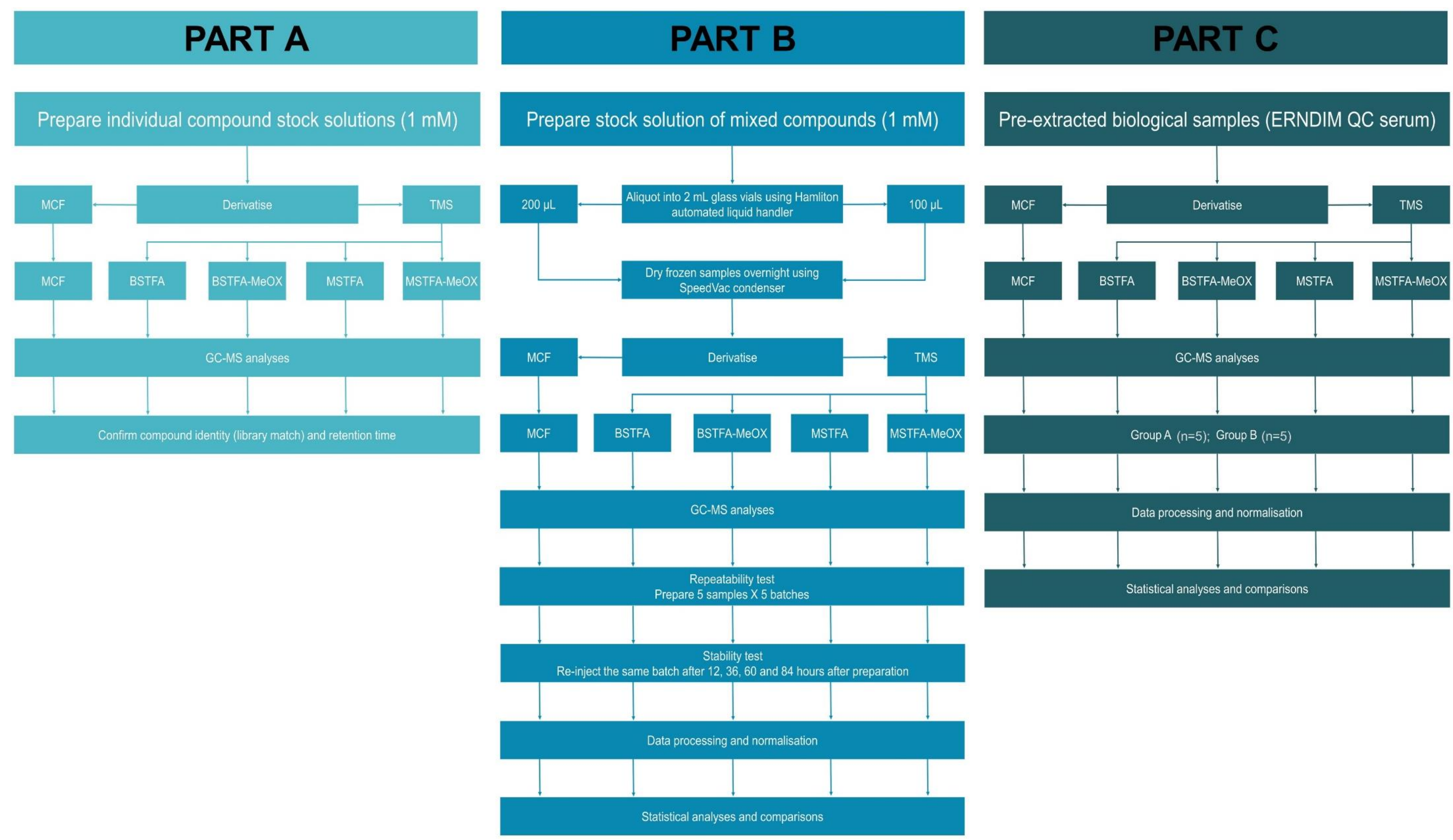


Figure 3.1: Summary of the general experimental design used to address the aim of this investigation, including: sample preparation, derivatisation, GC-MS analysis, data processing, compound identification and analytical comparison and interpretation.

3.2.1 Ethical approval

This investigation was conducted according to the Declaration of Helsinki and International Conference on Harmonisation Guidelines. Ethical approval for this investigation was obtained from the Research Ethics Committees of North-West University, South Africa (NWU-00016-23-A1).

3.2.2 Standard stock solution

3.2.2.1 Reagents and chemicals

Stock solution: Alanine (CAS RN® 56-41-7), 2-amino-3-bromo-5-methylbenzoic acid (ABMBA) (CAS RN® 13091-43-5), aspartic acid (CAS RN® 56-84-8), N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) with 1% trimethylchlorosilane (TMCS) (CAS RN® 25561-30-2), citric acid (CAS RN® 68-04-2), cystine (CAS RN® 56-89-3), fumaric acid (CAS RN® 110-17-8), glutamic acid (CAS RN® 86-0), glutamine (CAS RN® 56-85-9), glycine (CAS RN® 56-40-6), 2-hydroxybutyric acid (CAS RN® 19054-57-0), isocitric acid (CAS RN® 320-77-4), α-ketoglutaric acid (CAS RN® 328-50-7), lactic acid (CAS RN® 50-21-5), lysine (CAS RN® 56-87-1), methionine (CAS RN® 63-68-3), methoxyamine hydrochloride (CAS RN® 61-16-5), N-methyl-(trimethylsilyl)trifluoroacetamide (MSTFA) (CAS RN® 24589-78-4), methyl chloroformate (MCF) (CAS RN® 79-22-1), nonadecanoic acid methyl ester (C19:0 ME) (CAS RN® 1731-94-8), norleucine (CAS RN® 327-57-1), coumaric acid (CAS RN® 501-98-4), phenylalanine (CAS RN® 63-91-2), 3-phenylbutyric acid (CAS RN® 4593-90-2), pyridine (CAS RN® 110-86-1), sodium bicarbonate ($NaHCO_3$) (CAS RN® 144-55-8), sodium hydroxide (NaOH) (CAS RN® 1310-73-2),sodium sulphate (Na_2SO_4) (CAS RN® 7757-82-6), succinic acid (CAS RN® 150-90-3), tryptophan (CAS RN® 73-22-3) and valine (CAS RN® 72-18-4) were purchased from Sigma Aldrich (Kempton Park, Gauteng, SA). Honeywell (Burdick & Jackson) solvents including, chloroform (CAS RN® 67-66-3), methanol (CAS RN® 67-56-1) and water (CAS RN® 7732-18-5) were purchased from Anatech Instruments (Pty) Ltd (Johannesburg, South Africa).

3.2.2.2 Solutions

Preparing standard stock solution:

The compounds: Alanine (89.09 g/mol), 2-amino-3-bromo-5-methylbenzoic acid (ABMBA) (230.06 g/mol), aspartic acid (133.10 g/mol), citric acid (192.12 g/mol), cystine (240.30 g/mol), fumaric acid (116.07 g/mol), glutamic acid (147.13 g/mol), glutamine (146.14 g/mol), glycine (75.07 g/mol), 2-hydroxybutyric acid (104.11 g/mol), isocitric acid (192.12 g/mol), α-ketoglutaric acid (146.11 g/mol), lactic acid (90.08 g/mol), lysine (146.19 g/mol), methionine

(149.21 g/mol), norleucine (131.18 g/mol), coumaric acid (164.16 g/mol), phenylalanine (165.19 g/mol), 3-phenylbutyric acid (164.20 g/mol), succinic acid (118.09 g/mol), tryptophan (204.23 g/mol) and valine (117.15 g/mol) were weighed to closest decimal and prepared separately, by dissolving each in 3 mL of milli-Q water in a polypropylene tube (5 mL) (Brand). The stock solution was prepared by quantitatively transferring all the individually prepared compounds to a 1 000 mL volumetric flask, whereafter the volumetric flask was filled up to a 1 000 mL with, milli-Q water, resulting in a final concentration of approximately 1 mM. These compounds were prepared in sufficient amounts to ensure that enough was available for all relevant analytical applications throughout the study. The stock solution was aliquoted using a Hamilton automated liquid handler, where 100 µL of the solution was transferred to GC vials for use when performing TMS derivatisation and 200 µL of the solution was transferred to GC vials for use when performing MCF derivatisation. The vials with the pre-aliquoted volumes were freeze-dried overnight and stored at -80˚C until used. The remaining standard stock solution was stored at -80˚C when not in use.

3.2.2.3 Internal standard preparation:

Nonadecanoic acid methyl ester

TMS: Nonadecanoic acid methyl ester (C19:0 ME) for TMS derivatisation was prepared by dissolving 34.94 mg C19:0 ME in 25 mL of 2,2,4-trimethylpentane. Before TMS derivatisation, 200 µL of the C19:0 ME solution was added to a 1 000 µL of BSTFA or MSTFA reagent and added to the sample of interest (Willers *et al.*, 2016).

MCF: Nonadecanoic acid methyl ester (C19:0 ME) for MCF derivatisation was prepared by dissolving 2.92 mg C19:0 ME in 25 mL chloroform, which was added to the sample of interest during the last step of the MCF derivatisation procedure.

3.2.2.4 Derivatisation reagent preparation:

Methoxyamine hydrochloride solution

The methoximation reagent was prepared by dissolving 20 mg of methoxyamine hydrochloride in 1 mL of pyridine. The mixture was vortexed for 30 s and incubated for 15 min at 60˚C to help dissolve the methoxyamine hydrochloride (Fiehn, 2016). This was prepared daily for use when TMS derivatisation was performed.

Sodium hydroxide

The sodium hydroxide (NaOH) solution was prepared by dissolving 40 g of NaOH in 1 000 mL milli-Q water. The same solution was used for the duration of the experiment and stored at room temperature (Smart *et al.*, 2010).

Sodium bicarbonate

The sodium bicarbonate ($NaHCO_3$) solution was prepared by dissolving 4 g of $NaHCO_3$ in 1 000 mL milli-Q water. The same solution was used for the duration of the experiment and stored at room temperature (Smart *et al.*, 2010).

3.2.3 Derivatisation methods

BSTFA with 1% TMCS (purchased as a mixture) will further be referred to as BSTFA.

3.2.3.1 Silylation

N, O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA)

The freeze-dried samples containing the aliquoted stock solution (section 3.2.2.2) and the biological samples (section 3.2.4) were removed from the freezer 10 min prior to silylation and left on the bench to reach room temperature. A volume of 100 µL of pyridine and next 100 µL BSTFA were added to each sample and capped, vortexed for 30 s, and incubated for 60 min at 60˚C. The samples were subsequently transferred to GC vial inserts for GC-MS analysis (Parlapally *et al.*, 2016).

N-Methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA)

The freeze-dried samples containing the aliquoted stock solution and the biological samples were removed from the freezer 10 min prior to silylation to reach room temperature. A volume of 100 µL of pyridine was added to each sample. This was followed by the addition of 100 µL MSTFA to each sample, whereafter the samples were capped, vortexed for 30 s, and incubated for 30 min at 37˚C (Engel *et al.*, 2020). The samples were subsequently transferred to GC vial inserts for GC-MS analysis.

3.2.3.2 Combination derivatisation with methoximation and silylation

Methoximation and N, O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA-MeOX)

The freeze-dried samples containing the aliquoted stock solution and the biological samples were removed from the freezer 10 min prior to methoximation to reach room temperature. The derivatisation method was carried out using freshly prepared methoxyamine hydrochloride dissolved in pyridine (20 mg/mL) (section 3.2.3). Each sample received 100 µL of the methoximation reagent and was vortexed for 30 s. Subsequently, the samples were incubated for 60 min at 60˚C. Once methoximation was complete, the samples were cooled for 10 min prior to adding 100 µL BSTFA. The samples were vortexed for 30 s, and incubated for 60 min at 60˚C (Venter *et al.*, 2016), whereafter the contents were transferred to GC vial inserts for GC-MS analysis.

Methoximation and N-Methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA-MeOX)

The freeze-dried samples containing the aliquoted stock solution and the biological samples were removed from the freezer 10 min prior to methoximation to reach room temperature. The derivatisation method was carried out with freshly prepared methoxyamine hydrochloride dissolved in pyridine (20 mg/mL) (section 3.2.3). Each sample received 100 µL of the methoximation reagent and vortexed for 30 s. Subsequently, the samples were incubated for 60 min at 60˚C. Once methoximation was complete, the samples were cooled for 10 min prior to adding 100 µL MSTFA. The samples were vortexed for 30 s, and incubated for 30 min at 37˚C (Engel *et al.*, 2020), whereafter the contents were transferred to GC vial inserts for GC-MS analysis.

3.2.3.3 Alkylation

The freeze-dried samples containing the aliquoted stock solution and the biological samples were removed from the freezer 10 min prior to alkylation to reach room temperature. The freeze-dried samples were re-suspended in 200 µL of NaOH (1 mM, and vortexed for 30 s). The samples were quantitatively transferred to silanised borosilicate glass tubes, followed by the addition of 167 µL methanol and 34 µL pyridine. Next, the samples were derivatised by carrying out the following steps while on the vortex: first, 20 µL of MCF was added and vortexed for 30 s, after which another 20 µL of MCF was added to the glass tube timing the reaction for another 30 s. A volume of 400 µL chloroform (containing methylated C19:0) was added and vortexed for 10 s. Lastly, 400 µL $NaHCO_3$ was added and vortexed for the final 10 s. Once all the samples were derivatised the batch was centrifuged at 945 g for 5 min at 6˚C to allow separation of the aqueous phase and the organic phase. Using a glass Pasteur pipette, the upper aqueous phase was removed and discarded. This was repeated for all the samples. The

residual water in the samples were removed by adding ± 30 mg of sodium sulphate to the chloroform phase. Subsequently, a glass Pasteur pipette was used to transfer the chloroform phase containing the MCF derivatives to an insert in a GC vial for GC-MS analyses (Villas-Bôas *et al.*, 2003).

3.2.4 Biological samples

Ten extracted samples (five random level 1 and five random level 2) were received from the Laboratory for Inborn Errors of Metabolism situated in the Centre for Human Metabolomics, North-West University. The samples are part of the European Research Network for evaluation and improvement of screening, Diagnosis, and treatment of Inherited disorders of Metabolism (ERNDIM) quality scheme. ERNDIM is an independent not-for-profit foundation which has been providing External Quality Assurance schemes in the field of inborn errors of metabolism since 1994 (https://www.erndim.org/).

3.2.5 GC-MS analysis

A GC-MS system consisting of an Agilent 7890A GC with an Agilent 7683B autosampler coupled to an Agilent mass spectrometer 5975B was used to analyse the samples. Helium served as the carrier gas. To separate the metabolites, an Agilent J&W DB-5MS GC column (30 m X 250 µm X 0.25 µm) (Agilent Technologies) was used. A sample volume of 1 µL was injected in split mode (1:5). A constant gas flow was maintained at a flow rate of 1 mL/min. The front inlet temperature was kept at a temperature of 230°C. The oven temperature was ramped from 70°C to 300°C over a period of 49 min, with the goal to improve metabolite separation. The oven temperature initiated at 70°C and was increased to 179°C at a rate of 10°C/min, increased to 180˚C at a rate of 0.5˚C/min and held for 2 min, increased to 220˚C at a rate of 10˚C/min and held for 1min, increased to 265˚C at a rate of 2.5˚C/min and held for 1 min, increased to 280˚C at a rate of 10˚C/min and held for 1 min, and lasty increased to 300˚C at a rate of 10˚C/min and held for 1 min. A solvent delay of 8.3 min (TMS) and 7 min (MCF) was used for the two types of methods. A single ion monitoring (SIM) method with simultaneous scanning was used during analysis to improve the sensitivity of the run. The Agilent 5975B series mass selective detector (MSD) provides the ability to create MS methods that continually alternate between scanning and selective ion monitoring (SIM) acquisitions, known as SIM-Scan. This method can therefore obtain scan information almost instantaneously with the SIM information throughout the chromatographic run and make the monitoring of a list of target compounds at low concentrations possible. A characteristic ion for each derivative (TMS and MCF) was selected (Tables 3.1 and 3.2). The SIM method included 15 (TMS) and 14 (MCF) separate time segments corresponding to the retention times of 23 derivatives (compounds and internal standards) from different chemical classes.

Table 3.1: Retention times and targeted ions for TMS derivatives used in selective ion monitoring mode.

Derivative	Retention time (min)	TMS targeted ion (SIM m/z)
Lactic acid, 2TMS	9.197	117
Alanine, 2TMS	10.046	116
Glycine, 2TMS	10.425	102
2-Hydroxybutyric acid, 2TMS	10.486	131
Valine, 2TMS	12.033	144
Glycine, 3TMS	13.452	174
Succinic acid, 2TMS	13.574	147
Norleucine, 2TMS	13.639	158
Fumaric acid, 2TMS	14.102	245
3-Phenylbutyric acid, TMS	15.427	118
Aspartic acid, 3TMS	16.362	232
Methionine, 2TMS	16.377	176
Phenylalanine, TMS	16.924	120
α-Ketoglutaric acid, 2TMS	17.705	147
Glutamic acid, 3TMS	17.892	246
Phenylalanine, 2TMS	18.053	218
Glutamine, 3TMS	21.093	156
ABMBA (non-derivative)	21.910	211
Citric acid, 4TMS	22.675	273
Coumaric acid, 2TMS	22.921	219
Lysine, 4TMS	23.623	174
Tryptophan, 3TMS	28.562	202
Nonadecanoic acid methyl ester (non-derivative)	29.039	74
Cystine, 4TMS	35.300	73
Isocitric acid, 4TMS	21.974	147

Table 3.2: Retention times and targeted ions for MCF derivatives used in selective ion monitoring mode.

Derivative	Retention time (min)	MCF targeted ion (SIM m/z)
Fumaric acid	8.449	113
Lactic acid	8.676	103
Succinic acid	8.699	115
2-Hydroxybutyric acid	10.255	73
Alanine	10.592	102
Glycine	10.832	88
Valine	12.755	130
α-Ketoglutaric acid	12.755	115
3-Phenylbutyric acid	13.626	105
Norleucine	14.612	88
Aspartic acid	15.614	160
Citric acid	15.682	143
Isocitric acid (main peak 1)	16.182	115
Glutamic acid	17.388	114
Methionine	17.631	61
ABMBA (non-derivative)	18.853	213
Isocitric acid (main peak 2)	19.84	129
Phenylalanine	19.901	162
Coumaric acid	22.66	161
Lysine	25.328	142
Nonadecanoic acid methyl ester (non-derivative)	29.003	74
Tryptophan	31.954	130
Cystine	Not detected	
Glutamine	Not detected	

3.2.6 GC-MS data processing

3.2.6.1 Standard stock solution

The raw data generated by the GC-MS system was translated by MassHunter GC-MS translator B.07.01 and processed further with the use of Agilent MassHunter quantitative analysis. Peak identification was done by comparisons to NIST and the library of the Potchefstroom Laboratory for Inborn Errors of Metabolism.

3.2.6.2 Biological samples

Samples were injected and run using the same method as the previous samples (section 3.2.5). The Agilent scan data files were processed with AMDIS software (version 2.66) for GC-MS data interpretation from NIST. AMDIS automatically extracts pure (background free) component mass spectra. The automated steps included, noise analysis, component perception, spectrum deconvolution, and compound identification (Stein & Scott, 1994). The AMDIS processed extracted (identified and unidentified) data were imported into Agilent Mass Profiler Professional and retention times were aligned with the known internal standards (2-amino-3-bromo-5-methylbenzoic acid, and nonadecanoic acid methyl ester). The features were exported to a comma separated values data file.

3.2.7 Statistical analyses

3.2.7.1 Standard stock solution

Raw data obtained from the stock solutions (n=5) of analysed samples over five days, were normalised with the internal standard, 3-phenylbutyric acid (Reinecke *et al.*, 2012) and subjected to PCA analyses using the webserver MetaboAnalyst (version 5.0). Herein no filtering was applied, but the data were log transformed. The relative standard deviation (RSD) was determined for each individual compound (n=5) over 5 days (n=25), by using the standard deviation, divided by the mean, expressed as a percentage (Parsons *et al.*, 2009). The average RSD for all compounds over time was determined and plotted using GraphPad Prism (version 9.4.1).

3.2.7.2 Biological samples

Pre-processing of untargeted (scan) data were performed in Microsoft Excel, per derivatisation method. Firstly, zero filtering was applied to ensure a focused dataset (Smuts *et al.*, 2013; Lamichhane *et al.*, 2018). Secondly, the data were normalised relative to the internal standard 3-phenylbutyric acid (Reinecke *et al.*, 2012). Next, the data were uploaded to the webserver MetaboAnalyst (version 5.0) where missing value estimation was performed, followed by log transformation. Multivariate analyses by means of PCA was performed using a 95% confidence region, to visualise the natural grouping within the data (Alonso *et al.*, 2015). Univariate analysis in the form of t-test comparisons were performed (false discovery rate corrected $p < 0.05$) to find significant differences between experimental groups (Saccenti *et al.*, 2014; Alonso *et al.*, 2015; Chong *et al.*, 2019). Additionally, the mean normalised peak area of the top ten significant features identified with each derivatisation method was determined and reported.

CHAPTER 4: RESULTS AND DISCUSSION

Derivatisation is often a necessary step of the sample preparation process before GC-MS analyses can be executed (He *et al.*, 2018). It is important to familiarise oneself with the different derivatisation methods, and to determine how they influence the metabolites of interest, in order to choose the most suitable method for the experiment at hand (Orata, 2012). Derivatisation conditions are often assessed via the repeatability of replicates' relative normalised response (Christou *et al.*, 2014). Hence, in this study, five different derivatisation methods were compared using a stock solution with a variety of pre-selected compounds (section 3.2.2.2). Additionally, the derivatisation methods were evaluated on extracted biological samples as proof of concept for metabolomics studies (section 3.2.4). The results from the various comparisons are herein reported next.

4.1 DATA OVERVIEW PCA SCORE PLOTS

An overview of the GC-MS data, of the samples (n=5), prepared over five days, derivatised using a) BSTFA; b) BSTFA-MeOX; c) MSTFA; d) MSTFA-MeOX and e) MCF methods, are shown in Figure 4.1. In each instance a multivariate assessment of the data acquired by GC-MS analyses is given as PCA score plots. Throughout Figure 4.1a-e, principal component 1 (PC1) explains the largest variation in the data, depicting variation between groups, (batches), whereas PC2 projects the variation within one group analysed as a batch. Amongst all the compared methods most of the variation was accounted for by PC1 (85.6%) when BSTFA (Figure 4.1a) was used as derivatisation method, while only 62.3% of the variation was accounted for by the first PC when BSTFA-MeOX (Figure 4.1b) was used. Considering that the same (individually prepared) samples were analysed as repeats over five days, it is expected that no clear separation is seen between the batches analysed, confirming that the metabolite profiles are similar. Albeit variation is present between the different groups, prepared on different days, contributing to the different ellipses displayed in the PCA plots.

Plots with small ellipses suggest closer grouping of the data and less variation as seen in batch 1 (●) of Figure 4.1a and batch 2 (X) of Figure 4.1b. The opposite is also true, where large ellipses indicate more variation within the specific batches as seen in batch 4 (▲) of Figure 4.1a, batch 5 (◆) of Figure 4.1b, and batch 1 (●) and 2 (X) of Figure 4.1e. Interestingly the batches derivatised with the addition of methoxyamine hydrochloride solution (MeOX) (Figure 4.1b and d) showed similar groupings of the data, with relatively constant overlapping of ellipses (apart from some samples adding to variation in the data). Then again, the use of TMS methods (BSTFA and MSTFA) without MeOX showed larger overlapping ellipses for most of the batches (Figure 4.1a and c).

When adding samples derivatised by silylation (BSTFA, BSTFA-MeOX, MSFTA, MSTFA-MeOX) and alkylation (MCF) on the same PCA plot (irrespective of batch), clear separation is seen between the TMS and MCF samples, suggesting distinct differences in the GC-MS profiles (Figure 4.1f). This plot showed larger inter-variance (PC1) (77.4%) than the intra-variance (PC2) (10.9%). Also, the TMS methods (higher metabolite abundance detected) grouping to the left of the plot and the MCF method (lower metabolite abundance) seen on the right side of the plot.

The use of PCA analysis is a useful way of obtaining an overview of the data, looking for outliers, and to visually detect sample patterns of groupings (Chong *et al.*, 2019). The current study indicates distinct TMS and MCF data arrangements on the PCA plots. Differences between PCA plots showcasing TMS and MCF derivatisation methods have been previously reported, where MCF provided better discrimination between groups (Loyo *et al.*, 2021). Also, PLS-DA score plots have been used to reveal major differences between TMS and MCF metabolites from bacterial culture methods (Azizan & Baharum, 2012). When focusing on differences in metabolite profiles between BSFTA and BSTFA-MeOX prepared samples, a study by Willers *et al.* (2016) on fatty acid methyl esters from soil microbial communities revealed via PCA score plots, that larger variation is present in the samples prepared only with BSFTA, compared to BSTFA with the inclusion of MeOX, similar to the results of this study. Then again distinct differences in metabolite profiles detected in milk showed clear PCA separation between BSFTA and MSTFA derivatisation techniques (Parvatam *et al.*, 2023). In the current study BSTFA-MeOX and MSTFA-MeOX showed separation on the combined PCA score plot, while BSTFA and MSTFA showed overlapping metabolite profiles. This outcome can likely be because of the use of analytical standards, while the use of biological samples as in the study by Parvatam *et al.* (2023) will result in different derivatisation responses.

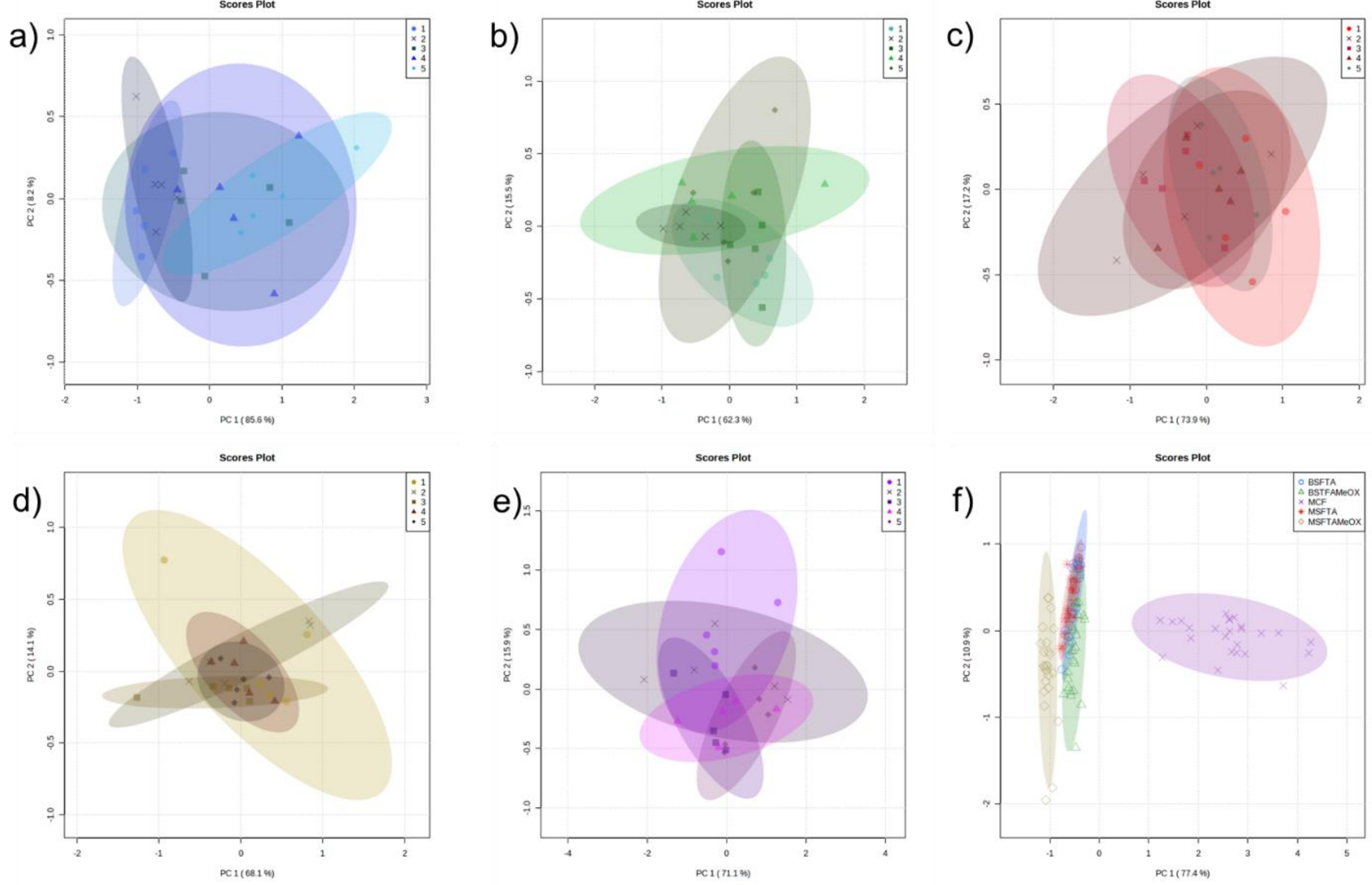


Figure 4.1: The PCA score plots obtained from 25 samples per derivatisation method analysed by GC-MS. Derivatised by: a) BSTFA; b) BSTFA-MeOX; c) MSFTA; d) MSTFA-MeOX; e) MCF; and f) PCA score plot of all tested methods (only compounds detected with all methods are shown). Batches 1 (●), 2 (X), 3 (■), 4 (▲), 5 (⬩).

4.2 REPEATABILITY DATA OVERVIEW

Repeatability can be used to describe the variation between measurements applied to the same subject and measurements that are subsequently repeated under the same conditions (He *et al.*, 2018). To evaluate this, the coefficient of variation (CV) (the ratio of the standard deviation to the mean expressed in percentage), which is equivalent to the RSD value can be used (Parsons *et al.*, 2009). Hence, this value is a measurement of the relative distribution of data points around the mean of the corresponding data set (Brown, 1998; Pélabon *et al.*, 2020). This is an important aspect since the repeatability of the method determines how reliably results can be interpreted (Hibbert, 2007; Abbiss *et al.*, 2015; Worsfold *et al.*, 2019).

The RSD was calculated from the normalised data for replicates in each method. A visual overview of the calculated average RSD (%) values for each method can be viewed in Fig. 4.2, while Table 4.1 - 4.3 give a complete layout of the results for each method. For targeted metabolomics analyses the Food and Drug Administration recommends a CV value of between 15% and 20% (t'Kindt *et al.*, 2009). However, more variation is expected when doing untargeted analysis (Schoeman & du Preez, 2012) with a CV value of > 50% indicating that the compound is not reliably measured (Venter *et al.*, 2016). In the current study, data is collected in an untargeted fashion but mined targeted as the compounds of interested have been pre-selected, thus data with RSD values higher than 50% will be seen as unreliably detected. The box-like plots in Figure 4.2 showed that the use of MSTFA-MeOX (plot 4 in yellow) as derivatisation method resulted in the largest RSD values (excluding the one outlier in the data obtained via MCF derivatisation that has an RSD > 140%). Subsequently, BSTFA and BSTFA-MeOX showed similar distribution of data points. Most compounds derivatised by BSTFA gave the highest peak areas/signals (seen as average peak areas) compared to the other derivatisation methods (Table 4.1 – 4.3; derivatives are listed in the tables in terms of detected retention times, reporting the average RSDs, with values < 20% highlighted in green and values > 50% highlighted in red). The silylation derivatisation methods using BSTFA, BSTFA-MeOX and MSTFA resulted in 13 compounds with RSDs < 20%. While the use of MSTFA-MeOX had 11 compounds and MCF derivatisation had 15 compounds with RSDs < 20%. Additionally, BSTFA and MCF derivatisation showed one compound that exceeds a 50% RSD, whereas MSTFA with methoximation had four compounds with an RSD > 50%. When using BSTFA and MSTFA as derivatisation reagents, the compound lactic acid could not be detected. With and without the inclusion of MeOX with both BSTFA and MSTFA derivatisation methods did not aid the detection of isocitric acid and α-ketoglutaric acid. The use of MCF derivatisation did not facilitate the detection of glutamine and cystine.

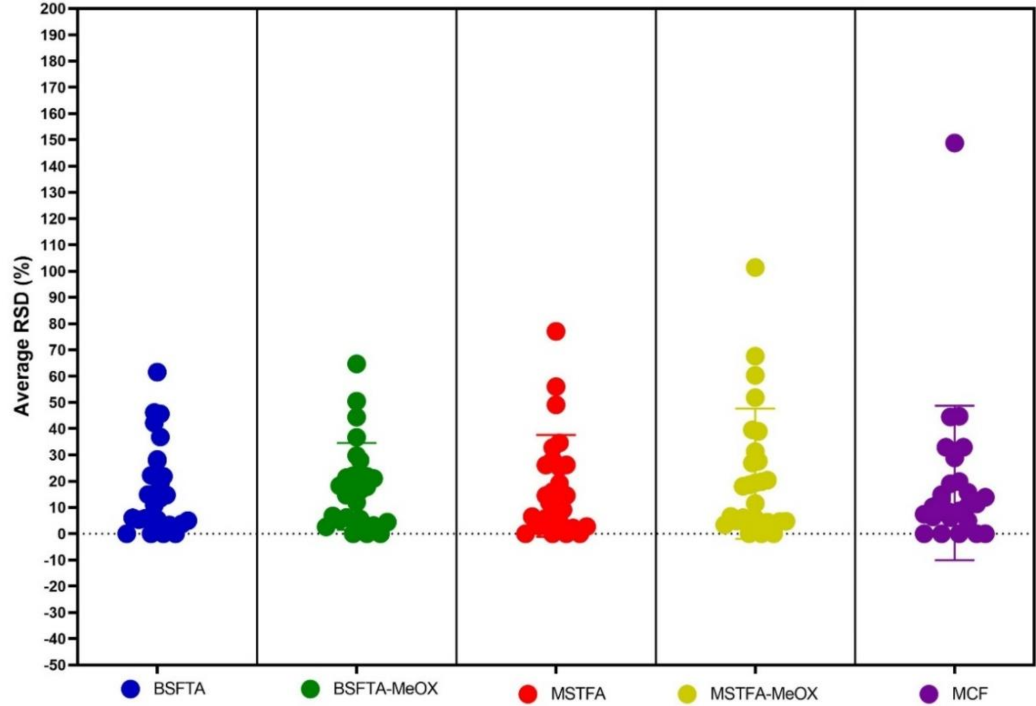


Figure 4.2: An overview of the average RSD distribution per compound of the sample replicates derivatised by five methods over five days.

Table 4.1: BSTFA derivatisation overview of compounds detected, the retention times, average peak area and average RSD (± standard deviation). Compounds with RSD < 20% are highlighted in green, compounds with RSD > 50% and highlighted in red. Compounds are listed in order of detected retention times.

Derivative	**BSTFA**				**BSTFA-MeOX**			
Parameter	**RT (min)**	**AVE AREA**	**AVE RSD (%)**	**STDEV RSD**	**RT (min)**	**AVE AREA**	**AVE RSD (%)**	**STDEV RSD**
Lactic acid, 2TMS	ND				9.197	387,130	6.2	1.29
Alanine, 2TMS	10.046	821,091	6.1	2.10	10.046	796,862	6.7	1.59
Glycine, 2TMS	10.425	409,404	5.3	0.57	10.425	306,643	17.9	3.33
2-Hydroxybutyric acid, 2TMS	10.486	707,930	5.8	2.07	10.486	676,842	5.6	1.28
Valine, 2TMS	12.033	934,518	5.0	1.69	12.033	925,337	5.2	0.98
Glycine, 3TMS	13.452	38,059	42.2	15.91	13.452	280,262	36.7	2.17
Succinic acid, 2TMS	13.574	1,023,165	3.7	1.70	13.574	991,869	3.1	0.37
Norleucine, 2TMS	13.639	1,193,957	5.5	1.85	13.639	1,173,583	4.4	1.22
Fumaric acid, 2TMS	14.102	573,955	3.3	2.19	14.102	555,969	2.6	0.51
3-Phenylbutyric acid, TMS	15.427	271,694	NA	NA	15.427	262,833	NA	NA
Aspartic acid, 3TMS	16.362	612,517	13.6	6.95	16.362	551,762	18.2	8.49
Methionine, 2TMS	16.377	465,425	10.9	6.00	16.377	458,537	12.0	5.75
Phenylalanine, TMS	16.924	108,053	21.8	7.60	16.924	143,854	22.1	8.46
Glutamic acid, 3TMS	17.892	293,470	36.8	18.12	17.892	92,301	64.6	5.50

Table 4.1: Continued.

Phenylalanine, 2TMS	18.053	407,319	19.2	9.03	18.053	315,834	29.7	10.68
Glutamine, 3TMS	21.093	274,95	61.5	26.55	21.093	3,237	14.6	5.40
ABMBA	21.910	89,271	14.7	5.33	21.910	58,781	16.5	5.44
Citric acid, 4TMS	22.675	174,641	15.1	6.29	22.675	167,894	28.0	13.24
Coumaric acid, 2TMS	22.921	174,419	14.9	6.61	22.921	177,228	16.1	5.78
Lysine, 4TMS	23.623	6,979	45.6	10.17	23.623	6,379	50.4	19.98
Tryptophan, 3TMS	28.562	973,190	46.1	22.81	28.562	362,648	44.4	14.92
C19 methyl ester	29.039	286,737	22.2	9.42	29.039	334,567	21.8	5.5
Cystine, 4TMS	35.300	422,627	27.8	12.00	35.300	438,295	21.0	9.62
Isocitric acid, 4TMS	ND				ND			
α-Ketoglutaric acid, 2TMS	ND				ND			

*ND: not detected; NA: not applicable as 3-phenylbutyric acid was used to normalise the data.

Table 4.2: MSTFA derivatisation overview of compounds detected, the retention times, average peak area and average RSD (± standard deviation). Compounds with RSD < 20% are highlighted in green, compounds with RSD > 50% and highlighted in red. Compounds are listed in order of detected retention times.

Derivative	**MSTFA**				**MSTFA-MeOX**			
Parameter	**RT (min)**	**AVE AREA**	**AVE RSD (%)**	**STDEV RSD**	**RT (min)**	**AVE AREA**	**AVE RSD (%)**	**STDEV RSD**
Lactic acid, 2TMS	ND				9.197	246,497	6.6	1.40
Alanine, 2TMS	10.046	715,201	5.9	3.19	10.046	512,720	6.1	1.68
Glycine, 2TMS	10.425	338,469	9.1	4.27	10.425	10,178	31.3	10.06
2-Hydroxybutyric acid, 2TMS	10.486	519,991	6.5	2.74	10.486	442,551	5.7	1.38
Valine, 2TMS	12.033	828,676	4.4	2.45	12.033	642,678	4.8	1.47
Glycine, 3TMS	13.452	9,998	32.8	4.54	13.452	616,703	4.7	1.65
Succinic acid, 2TMS	13.574	939,186	2.7	0.99	13.574	737,755	3.4	1.83
Norleucine, 2TMS	13.639	1,124,537	2.8	1.70	13.639	845013	4.6	1.86
Fumaric acid, 2TMS	14.102	576,192	2.2	0.95	14.102	440,519	4.4	2.30
3-Phenylbutyric acid, TMS	15.427	240,412	NA	NA	15.427	175,539	NA	NA
Aspartic acid, 3TMS	16.362	573,407	11.9	3.91	16.362	363,105	27.7	12.70
Methionine, 2TMS	16.377	455,005	7.5	2.35	16.377	298,898	19.8	9.66
Phenylalanine, TMS	16.924	69,145	26.1	4.66	16.924	111,902	26.9	12.65
Glutamic acid, 3TMS	17.892	306,993	34.6	12.10	17.892	67,933	67.6	18.55
Phenylalanine, 2TMS	18.053	408,604	16.0	5.85	18.053	204,026	39.5	15.79
Glutamine, 3TMS	21.093	25,193	77.1	22.43	21.093	930	101.3	34.27
ABMBA	21.910	41,664	19.2	7.45	21.910	39432	19.3	7.11
Citric acid, 4TMS	22.675	124,468	14.6	5.88	22.675	47,687	51.8	31.02
Coumaric acid, 2TMS	22.921	124,786	14.5	6.18	22.921	104295	11.5	4.74

Table 4.1: Continued.

Lysine, 4TMS	23.623	1,170	49.1	15.85	23.623	9,694	60.3	21.38
Tryptophan, 3TMS	28.562	339,622	56.0	23.37	28.562	105,062	39.0	13.28
C19 methyl ester	29.039	136,073	28.0	11.32	29.039	107619	20.4	5.14
Cystine, 4TMS	35.300	141,692	25.4	10.52	35.300	156,695	18.1	5.87
Isocitric acid, 4TMS	ND				ND			
α-Ketoglutaric acid, 2TMS	ND				ND			

***ND: not detected; NA: not applicable as 3-phenylbutyric acid was used to normalise the data.**

Table 4.3: MCF derivatisation overview of compounds detected, the retention times, average peak area and average RSD (± standard deviation). Compounds with RSD < 20% are highlighted in green, compounds with RSD > 50% and highlighted in red. Compounds are listed in order of detected retention times.

Derivative	**MCF**			
Parameter	**RT (min)**	**AVE AREA**	**AVE RSD (%)**	**STDEV RSD**
Fumaric acid	8.449	325,613	19.0	11.05
Lactic acid	8.676	92,317	6.1	1.41
Succinic acid	8.699	579,888	10.5	9.81
2-Hydroxybutyric acid	10.255	160,218	13.9	12.10
Alanine	10.592	324,634	44.4	31.88
Glycine	10.832	13,444	148.7	64.28
Valine	12.755	548,016	7.39	7.45
Coumaric acid	12.755	311,438	32.9	23.32
C19 methyl ester	12.755	1,882,358	19.8	6.80
3-Phenylbutyric acid	13.626	400,143	NA	NA
Aspartic acid	15.614	512,511	6.4	2.09
Citric acid	15.682	355,076	10.2	6.27
Isocitric acid (peak 1)	16.182	907,19	28.8	26.48
Glutamic acid	17.388	204,337	15.8	12.31
Methionine	17.631	124,033	14.8	8.84
α-ketoglutaric acid	18.853	174,206	8.2	8.18
Isocitric acid (peak 2)	19.840	50,023	13.2	2.90
Phenylalanine	19.901	587,351	11.3	4.56
Lysine	25.328	238,056	32.9	7.01
ABMBA	29.003	5,020	8.7	3.52
Tryptophan	31.954	2,660,031	44.7	12.12
Norleucine	31.954	295,780	4.9	2.27
Glutamine	ND			
Cystine	ND			

*ND: not detected; NA: not applicable as 3-phenylbutyric acid was used to normalise the data.

The use of box plots may be helpful in summarising complex results from multivariate analyses (Nuzzo, 2016), while it can also facilitate the identification of outlier data values (Williamson *et al.*, 1989; Krzywinski & Altman, 2014). This is verified in the current study, where box-like plots were used to compare the five different derivatisation methods, indicating the distribution of samples analysed over five days. As an overview there were differences in the data distribution when MCF derivatisation is used compared to TMS derivatisation. MCF derivatisation had more samples with an average RSD below 20% and were clustered closer together. This outcome is in agreement with the results found by Villas-Bôas *et al.* (2011), where MCF derivatisation resulted in more compounds with an RSD < 20%, when compared to compounds derivatised with MSTFA-MeOX. Furthermore, Moros *et al.* (2017) concluded that MSTFA derivatisation showed better repeatability than the use of BSTFA. In contrast, He *et al.* (2018) found that the use of BSTFA as derivatisation method gave more reproducible results than the use of MSTFA when profiling the intracellular metabolome of leukaemia cells. From the current findings, the data points from MSTFA derivatisation were more spread out (indicating larger RSD values) than BSTFA derivatisation findings. Additionally, while investigating an automation derivatisation protocol, Abbiss *et al.* (2015) compared BSTFA and MSTFA derivatisation and concluded that there was no clear indication of a superior derivatisation method in terms of compound class. Then again, when Sogin *et al.* (2019) analysed seawater samples, BSTFA derivatisation resulted in higher peak intensities compared to MSTFA derivatisation. They hypothesised that BSTFA is less susceptible to water inhibition, leaving more BSTFA available for compound derivatisation, hence resulting in the production of more derivatives and greater peak intensities. The results obtained by Sogin *et al.* (2019) were also similar to current findings, as BSTFA derivatisation resulted in the highest metabolite signals for most of the pre-selected compounds.

Given that derivatisation efficiency is reliant on temperature and reaction time, the average RSD (%) may differ between compounds with more than one derivative (Christou *et al.*, 2014; Moros *et al.*, 2017). Therefore, each derivatisation method with individual derivatives represented will be discussed separately in the following section.

4.3 REPEATABILITY PER DERIVATISATION METHOD

4.3.1 BSTFA

A visual overview of samples (n=25) derivatised with BSTFA as five samples in five batches can be seen in Figure 4.3. Only the compounds (total = 21) that were detected and identified via the BSTFA processing steps are shown in the graph. Following BSTFA derivatisation, multiple peaks were detected for glycine (2TMS and 3TMS) and phenylalanine (TMS and

2TMS), resulting in two metabolites listed. Glutamine was the only compound that had an average RSD value higher than 50% when using BSTFA. Additionally, cystine, glutamic acid, glycine (3TMS), lysine, nonadecanoic acid methyl ester, phenylalanine (TMS) and tryptophan had an average RSD value between 20% and 50%, and 2-hydroxybutyric acid, ABMBA, alanine, aspartic acid, citric acid, coumaric acid, fumaric acid, glycine (2TMS), methionine, norleucine, phenylalanine (2TMS), succinic acid and valine had an average RSD value lower than 20% (Table 4.1, Figure 4.3).

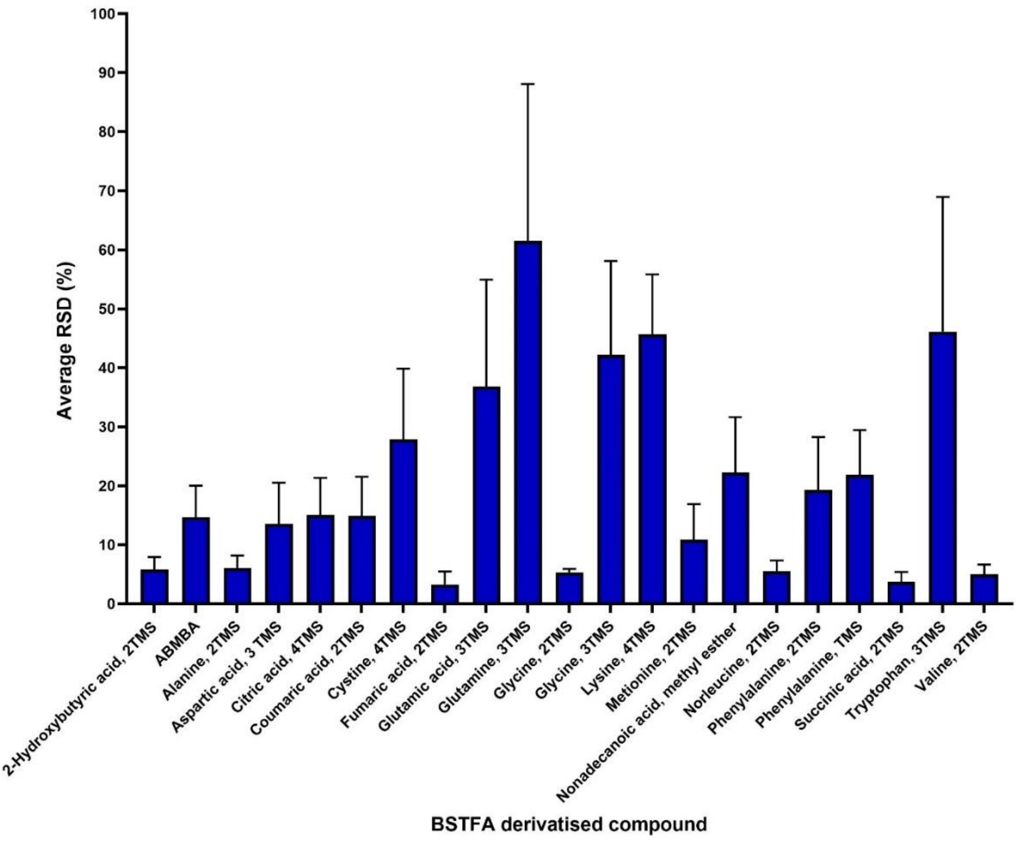


Figure 4.3: The average (±STEDV) RSD of BSTFA derivatised compounds over five analysed batches (n=25).

A well-known problem in analytical chemistry is the formation of artefacts during silylation, which leads to multiple derivatives for the same compound (Little, 1999; Rohloff, 2015; Moldoveanu & David, 2018). This was the case for glycine when Bani Rashaid *et al.* (2014) optimised a method for amino acid analysis, where both the di-trimethylsilyl- and tri-trimethylsilyl derivatives were obtained for the same compound after BSTFA derivatisation. This phenomenon of obtaining multiple peaks for the sample compound was again confirmed in the current investigation when considering glycine and phenylalanine. The occurrence of multiple peaks remains a problem when compound identification with biological application is the aim of a study (Little, 1999), as special attention will need to be added to data processing to ensure that the compound in question is reported correctly (Xu *et al.*, 2009). Multiple peaks are often present when silylation is incomplete, attributed to factors such as derivatisation time

and temperature (Fritsche-Guenther *et al.*, 2021). Resultantly it can be argued that the BSTFA silylation conditions of incubation of 60 min at 60°C is not optimal for glycine and phenylalanine. Furthermore, the reproducibility (RSD %) of derivatisation is an indication of how reliable a specific compound can be measured (Hibbert, 2007; Abbiss *et al.*, 2015). Pietrogrande *et al.* (2010) found that the use of BSTFA derivatisation for low-molecular-weight dicarboxylic acid analysis, such as succinic acid from atmospheric organic particulate matter, gave satisfactory reproducibility (RSD < 15%). The current study also showed good repeatability for dicarboxylic acids such as succinic acid or fumaric acid following BSTFA derivatisation. What is more is that all the derivatisation methods tested in this study showed good repeatability for succinic acid and fumaric acid, potentially making them easy compounds to detect irrespective of the derivatisation method used.

4.3.2 BSTFA-MeOX

A visual overview of samples (n=25) derivatised by BSTFA-MeOX in five batches can be seen in Figure 4.4. A total of 22 compounds were detected and identified using the BSTFA with methoximation as derivatisation method. As was the case when using BSTFA only, glycine (2TMS and 3TMS) and phenylalanine (TMS and 2TMS) resulted in the detection of multiple peaks. Glutamic acid and lysine showed average RSD values higher than 50%. Moreover, citric acid, cystine, glycine (3TMS), nonadecanoic acid methyl ester, phenylalanine (TMS and 2TMS) and tryptophan had an average RSD value between 20% and 50%, and 2-hydroxybutyric acid, ABMBA, alanine, aspartic acid, coumaric acid, fumaric acid, glutamine, glycine (2TMS), lactic acid, methionine, norleucine, succinic acid and valine resulted in an average RSD value lower than 20% (Table 4.1, Fig. 4.4).

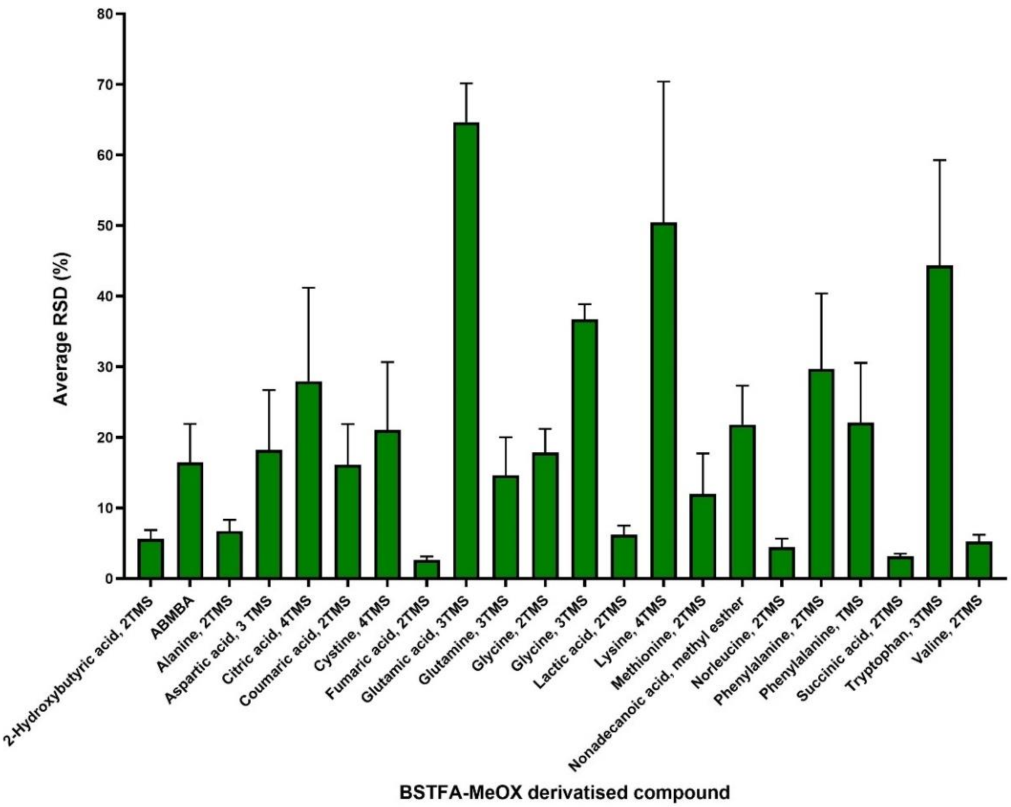


Figure 4.4: The average RSD (±STDEV) of BSTFA-MeOX derivatised compounds over five analysed batches (n=25).

Methoxyamine hydrochloride is an established methoximation reagent and is frequently used as part of a two-step derivatisation method where methoximation is carried out prior to silylation (Kouremenos *et al.*, 2010). The goal of this extra step is to protect the alpha-keto acids against decarboxylation and in turn results in simplified chromatograms (Parkinson, 2012; Glicksberg & Kerrigan, 2020; Khodadadi & Pourfarzam, 2020). A study by López-Bascón *et al.* (2019) compared the use of BSTFA with and without a methoximation step, and found that the addition of methoximation prior to silylation resulted in the detection of three additional carboxylic acids. Furthermore, Yin *et al.* (2017) optimised conditions for GC-TOF/MS analysis of host-gut microbiota metabolic interactions (involving organic acid, amino acids, amines, fatty acids and carbohydrates) using methoximation with BSTFA as derivatisation method. The repeatability from their study, performed on reference standards, resulted in RSD values < 20%, enabling quantitative detection of gut microbial metabolism. Also, when Carrasco-Pancorbo *et al.* (2009) prepared a standard mixture of compounds (containing mostly amino acids and organic acids) with BSTFA derivatisation and methoximation, suitable repeatability (< 10%) was obtained, supporting the application of this method to metabolic profiling of human cerebrospinal fluid samples. Overall, within the current investigation the inclusion of methoximation along with BSTFA derivatisation on the pre-selected list of compounds, did not result in higher repeatability, compared to the use of only BSTFA. Even with the addition of methoximation to BSTFA, multiple peaks for amino acids were detected as seen above when only using BSTFA

as derivatisation method. The occurrence of multiple peaks for the same compound has been previously reported by Carrasco-Pancorbo *et al.* (2009), with compounds, such as glycine, valine and isoleucine with more than one TMS derivative, similar to the results of this study.

4.3.3 MSTFA

A visual overview of samples (n=25) derivatised by MSTFA in five batches can be seen in Figure 4.5. A total of 21 compounds were detected and identified when using MSTFA as derivatisation method. Again, glycine and phenylalanine showed multiple peaks. Glutamine and tryptophan had an average RSD value higher than 50%. Cystine, glutamic acid, glycine (3TMS), lysine, nonadecanoic acid methyl ester and phenylalanine (TMS) had an average RSD value between 20% and 50%, and 2-hydroxybutyric acid, ABMBA, alanine, aspartic acid, citric acid, coumaric acid, fumaric acid, glycine (2TMS), methionine, norleucine, phenylalanine (2TMS), succinic acid and valine had an average RSD value lower than 20% (Table 4.2, Fig. 4.5).

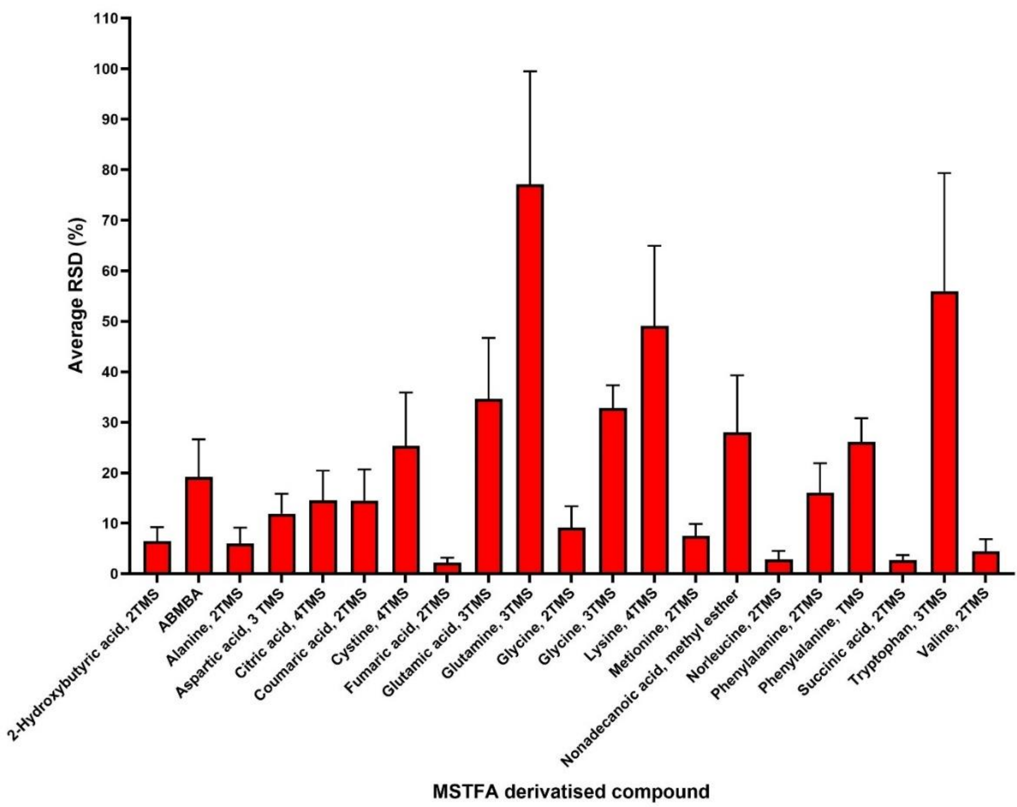


Figure 4.5: The average RSD (±STDEV) of MSTFA derivatised compounds over five analysed batches (n=25).

MSTFA is a frequently used reagent for trimethylsilylation (Fiehn, 2000; Roessner *et al.*, 2000; Huang *et al.*, 2010; Moros *et al.*, 2017; Beale *et al.*, 2018; Engel *et al.*, 2020; Bhumireddy *et al.*, 2021) and is considered more reactive and volatile than BSTFA (Parkinson, 2012). Verma *et al.* (2021) confirmed that MSTFA produced more volatile by-products during the

development of a method for the detection of medication. From the current study, MSTFA and BSTFA derivatisation resulted in the detection of the same number of compounds, with similar RSDs. Moreover, Engel *et al.* (2020) revealed that the amino acids lysine and glycine were converted into two derivatives using MSTFA as derivatising agent, which remains an eminent problem with silylation (Rohloff, 2015; Moldoveanu & David, 2018). This occurrence of two derivatives can be explained by the consecutive and slow-moving reactions of the H-atom of the alpha-amino group of glycine during derivatisation, meaning that longer derivatisation time may result in only one glycine TMS derivative (Moros *et al.*, 2017; Engel *et al.*, 2020; Mojsak *et al.*, 2020). These findings also correspond with the results of the current study since the amino acids glycine and phenylalanine produced two derivatives after derivatisation with MSTFA, as was the case following the use of BSTFA. To aid repeatability it is suggested to optimise derivatisation time, temperature and reagent volume (Barceló-Barrachina *et al.*, 2005). These are all factors that were not assessed in the current study yet it may affect the outcome of the detected peaks as seen in the current set of results (Moros *et al.*, 2017).

4.3.4 MSTFA-MeOX

A visual overview of samples (n=25) derivatised by MSTFA with methoximation in five batches can be seen in Figure 4.6. A total of 22 compounds were detected and identified using MSTFA-MeOX as derivatisation method. Multiple peaks were detected for glycine (2TMS and 3TMS) and phenylalanine (TMS and 2TMS). Four compounds (citric acid, glutamic acid, glutamine, and lysine) had average RSD values higher than 50%. Several compounds (aspartic acid, glycine (2TMS), nonadecanoic acid methyl ester, phenylalanine (TMS and 2TMS) and tryptophan) had average RSD values between 20% and 50%, and eleven compounds (2-hydroxybutyric acid, ABMBA, alanine, coumaric acid, fumaric acid, glycine (3TMS), lactic acid, methionine, norleucine, succinic acid and valine) had average RSD values lower than 20% (Table 4.2. Fig. 4.6).

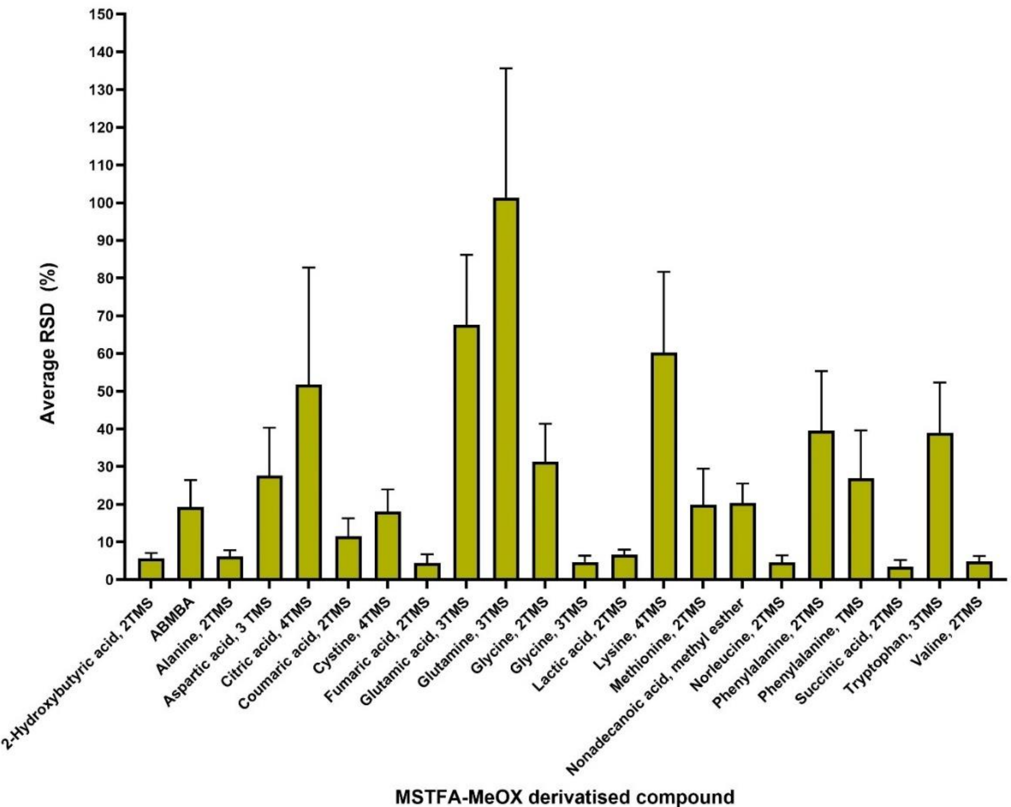


Figure 4.6: The average (±STDEV) RSD of MSTFA-MeOX derivatised compounds over five analysed batches (n=25).

The inclusion of methoximation prior to silylation produces oximes from aldehyde and keto groups and consequently inhibits the decarboxylation of ketoacids by disrupting charge transfer (Fiehn, 2000; Hušek *et al.*, 2016; Moros *et al.*, 2017; Beale *et al.*, 2018; Engel *et al.*, 2020). According to Pechlivanis *et al.* (2014) the use of methoximation with MSTFA derivatisation could be uniquely beneficial, given that MSTFA targets a wide range of polar compounds. By implementing this two-step derivatisation protocol 86 metabolites from different functional groups were identified in rat plasma samples (Pechlivanis *et al.*, 2014). During an investigation of derivatisation conditions by Moros *et al.* (2017), MSTFA with methoximation derivatisation, was found to be the superior derivatisation method in terms of compound detection, peak intensity and repeatability. Yet, in the current study, MSTFA with methoximation derivatisation had the lowest peak intensities for the majority of the compounds, with little to no distinction between compound detection and repeatability compared to the other derivatisation methods. Furthermore, when focussing on method repeatability (RSD), more than half of the compounds detected by MSTFA-MeOX in the present study, had an average RSD < 20% (noting that four specific compounds had an RSD > 50%). In a study of microbial metabolomes, the use of MSTFA-MeOX resulted in RSD < 10% for most of the tested standard compounds (including amino acids, organic acids, sugars, and sugar phosphates) (Koek *et al.*, 2006). Moreover, as was the case when using only MSTFA derivatisation, two derivatives for

glycine and phenylalanine were also present when methoximation was included in the present study, which are similar to the results obtained by Engel *et al.* (2020) when they analysed a standard metabolite mixture during a derivatisation condition investigation. Likewise, in an untargeted metabolomics study, multiple peaks were detected for some metabolites (e.g., glutamine and tryptophan) found in milk samples, reasoning that incomplete or partial derivatisation was the cause (Bhumireddy *et al.*, 2021).

4.3.5 MCF

A visual overview of samples (n=25) derivatised by MCF in five batches can be seen in Figure 4.7. A total of 21 compounds were detected and identified using MCF as derivatisation method. Detection of isocitric acid resulted in two peaks. Cystine and glutamine were not detected with the use of this method, while α-ketoglutaric acid was only detected with this method. Glycine had an average RSD value higher than 50%. Moreover, alanine, coumaric acid, isocitric acid (peak 1), lysine, nonadecanoic acid methyl ester and tryptophan have had average RSD value between 20% and 50%, while 2-hydroxybutyric acid, ABMBA, α-ketoglutaric acid, aspartic acid, citric acid, fumaric acid, glutamic acid, isocitric acid (peak 2), lactic acid, methionine, norleucine, phenylalanine, succinic acid and valine had an average RSD value lower than 20%. Consequently, the use of MCF resulted in the most compounds with an average RSD < 20% (Table 4.3, Fig. 4.7).

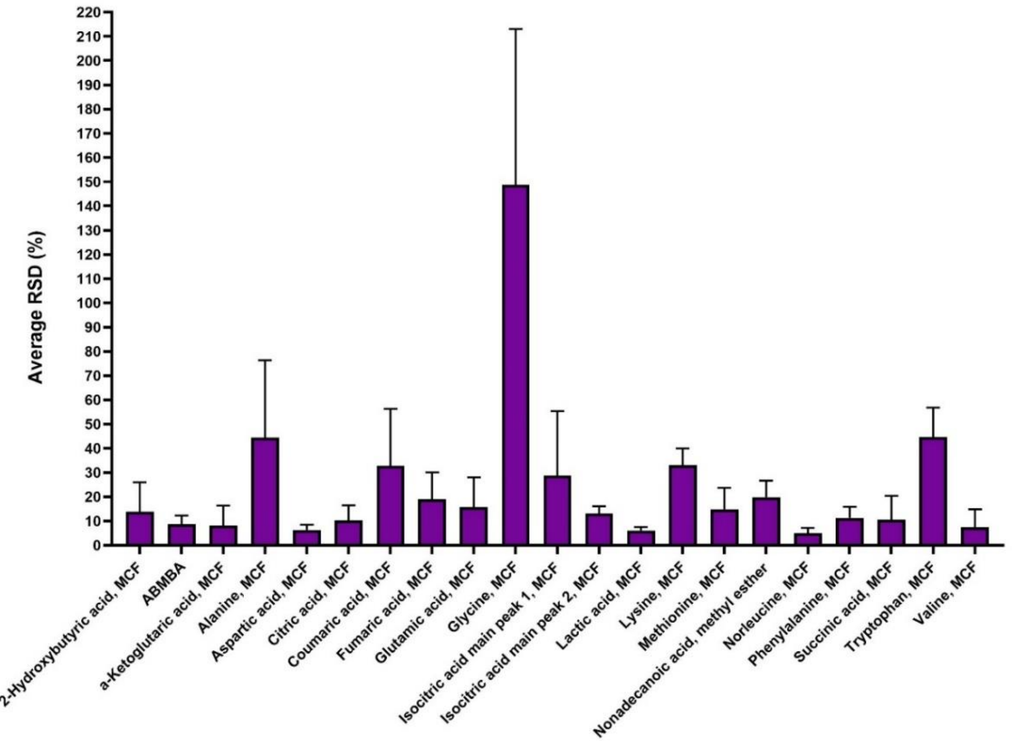


Figure 4.7: The average (±STDEV) RSD of MCF derivatised compounds over five analysed batches (n=25).

Alkyl chloroformates are versatile derivatisation reagents and are useful in the detection of amino- and organic acids (Kvitvang *et al.*, 2011; Hušek *et al.*, 2016), as showcased in a metabolomics study on abalone by Nguyen *et al.* (2023). Considering an array of amino acids, organic acids, fatty acids and nucleotides, Villas-Bôas *et al.* (2011) reported repeatability lower than 10% for most compounds analysed with the use of MCF derivatisation. Likewise, in the present study, the majority of the derivatives produced through MCF derivatisation, had an average RSD < 20%. Also, single derivatives were largely produced in the current investigation for the selected compounds (excluding isocitric acid), similar to a previous study conducted on milk samples for the analysis of amino- and organic acids (Bhumireddy *et al.*, 2021).

4.4 DERIVATISATION METHOD COMPARISON PER COMPOUND

A visual overview of the average RSD per compound derivatised by BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF are shown in Figure 4.8. In each subframe, a single compound is compared across the five different derivatisation methods. For each specific compound the lowest RSD was obtained using the following respective derivatisation methods: 2-hydroxybutyric acid when using BSTFA-MeOX; α-Ketoglutaric acid with MCF; ABMBA with MCF; alanine by utilising MSTFA; aspartic acid with MCF; citric acid with MCF; coumaric acid when using MSTFA-MeOX; cystine when using MSTFA-MeOX; fumaric acid by utilising MSTFA; glutamic acid with MCF; glutamine when using BSTFA-MeOX; glycine 1 (2TMS) by utilising BSTFA; glycine 2 (3TMS) when using BSTFA-MeOX; isocitric acid 1 and isocitric acid 2 with MCF; lactic acid and lysine with MCF; methionine by utilising MSTFA; C19 with MCF; norleucine by utilising MSTFA; phenylalanine 1 (TMS) and phenylalanine 2 (2TMS) by MSTFA and MCF respectively; succinic acid by utilising MSTFA; tryptophan when using MSTFA-MeOX and valine by utilising MSTFA. A summary of the results is showcased in Table 4.4, in terms of compound performance where the method with the highest (▲) and lowest (▼) compound RSDs are highlighted.

Figure 4.8: The comparison of the average percentage RSD for each compound derivatised by BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF. A total of 25 samples were derivatised per method.

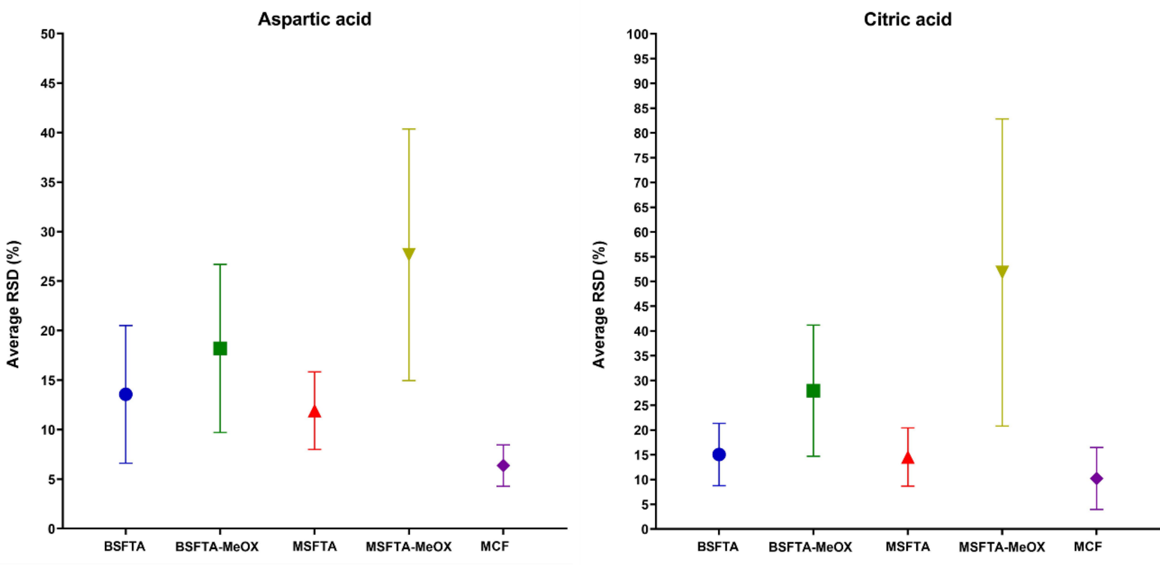
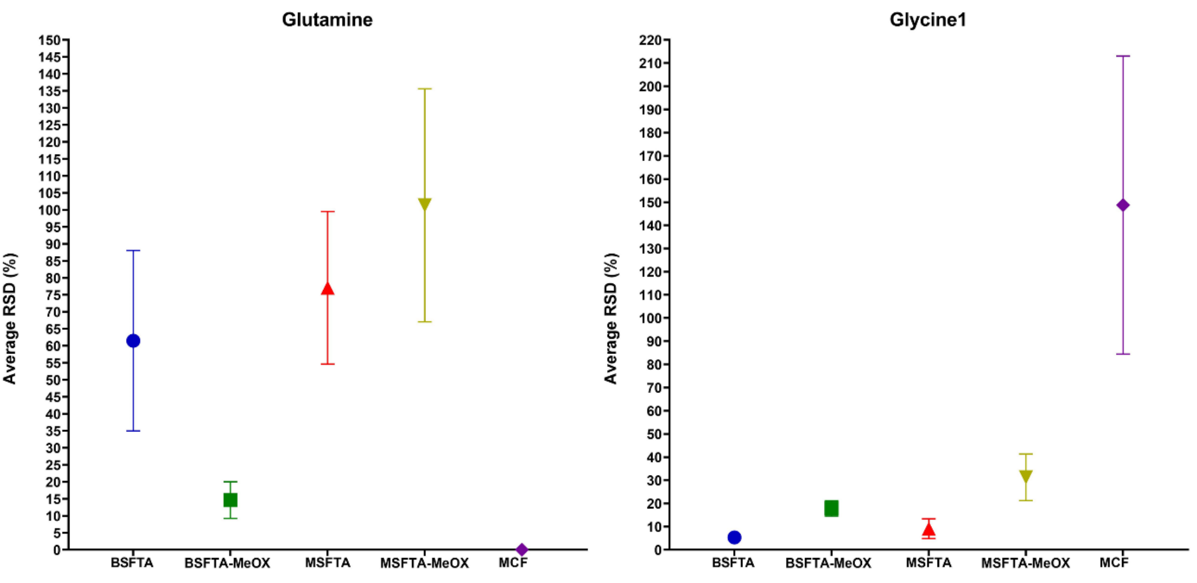


Figure 4.8: Continued.



Glycine2

Average RSD (%)

70 65 60 55 50 45 40 35 30 25 20 15 10 5 0

BSFTA BSFTA-MeOX MSFTA MSFTA-MeOX MCF

Isocitric acid1

Average RSD (%)

60 55 50 45 40 35 30 25 20 15 10 5 0

BSFTA BSFTA-MeOX MSFTA MSFTA-MeOX MCF

Isocitric acid2

Average RSD (%)

20 15 10 5 0

BSFTA BSFTA-MeOX MSFTA MSFTA-MeOX MCF

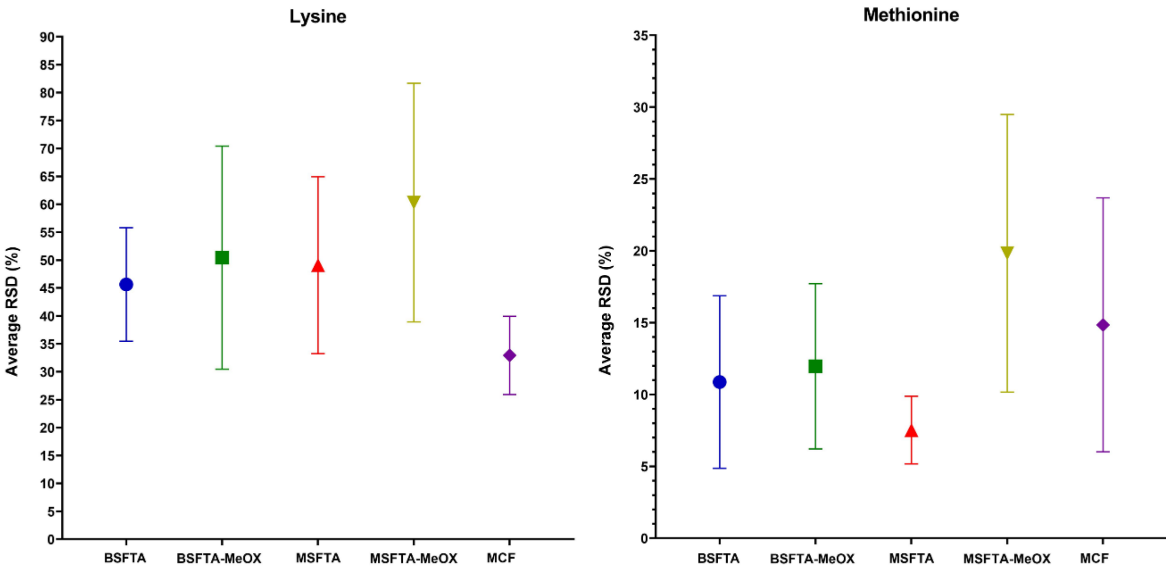
Lactic acid

Average RSD (%)

10 5 0

BSFTA BSFTA-MeOX MSFTA MSFTA-MeOX MCF

Figure 4.8: Continued.



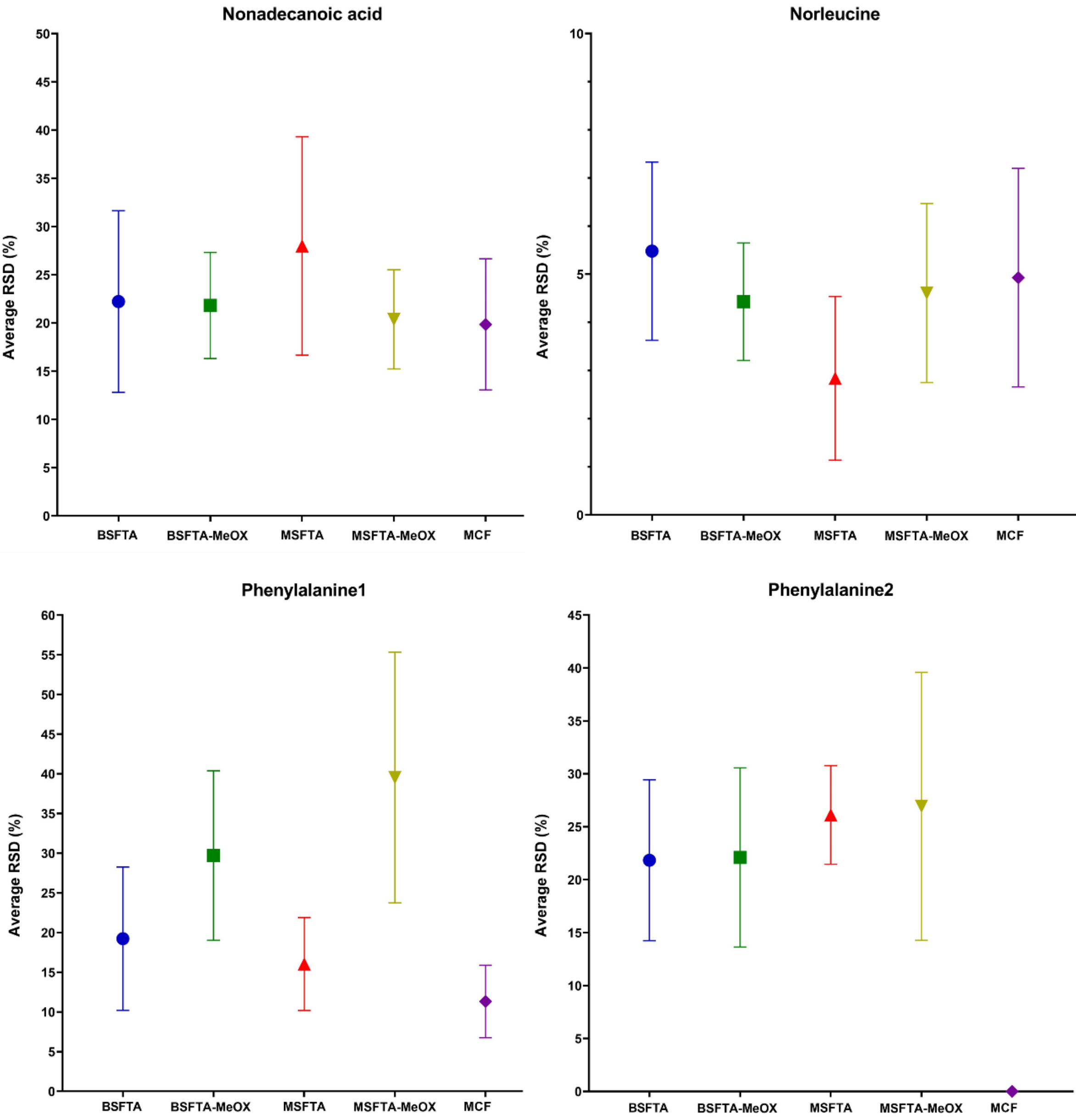
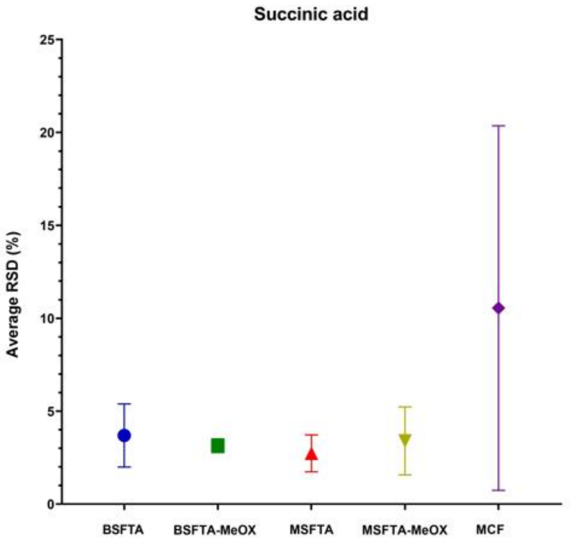
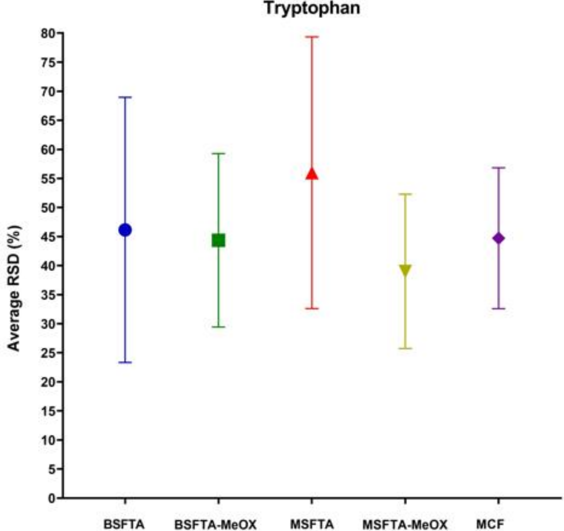


Figure 4.8: Continued.





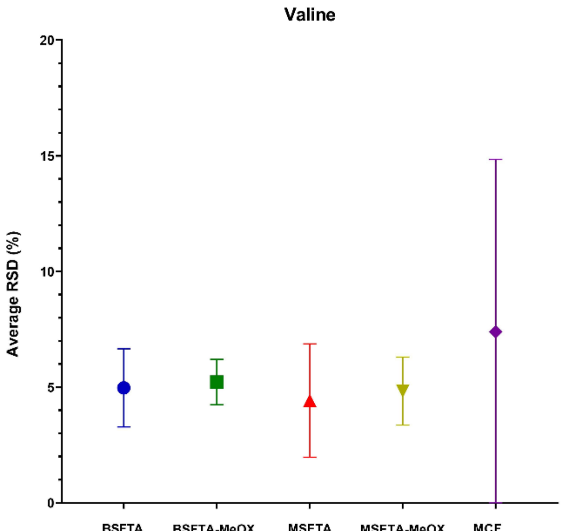


Figure 4.8: Continued.

Table 4.4: The comparison of the average percentage RSD for each identified compound derivatised by BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF. ▲ Highest RSD; ▼ Lowest RSD; ♦ Between lowest and highest values; ▪ Not detected. Compounds are listed in alphabetical order.

Compound	BSFTA	BSFTA-MeOX	MSTFA	MSTFA-MeOX	MCF
2-Hydroxybutyric acid	♦	▼	♦	♦	▲
α-Ketoglutaric acid	▪	▪	▪	▪	▼
ABMBA	♦	♦	♦	▲	▼
Alanine	♦	♦	▼	♦	▲
Aspartic acid	♦	♦	♦	▲	▼
Citric acid	♦	♦	♦	▲	▼
Coumaric acid	♦	♦	♦	▼	▲
Cystine	▲	♦	♦	▼	▪
Fumaric acid	♦	♦	▼	♦	▲
Glutamic acid	♦	♦	♦	▲	▼
Glutamine	♦	▼	♦	▲	▪
Glycine 1 (2TMS)	▼	♦	♦	♦	▲
Glycine 2 (3TMS)	▲	▼	♦	♦	▪
Isocitric acid 1	▪	▪	▪	▪	▼
Isocitric acid 2	▪	▪	▪	▪	▼
Lactic acid	▪	♦	▪	▲	▼
Lysine	♦	♦	▪	▲	▼
Methionine	♦	♦	▼	▲	♦
C19	♦	♦	▲	♦	▼
Norleucine	▲	♦	▼	♦	♦
Phenylalanine 1 (TMS)	▼	♦	♦	▲	▪
Phenylalanine 2 (2TMS)	♦	♦	♦	▲	▼
Succinic acid	♦	♦	▼	♦	▲
Tryptophan	♦	♦	▲	▼	♦
Valine	♦	♦	▼	♦	▲

The choice of derivatisation method is highly dependent on the nature of compounds analysed (Parkinson, 2012), given that derivatisation methods have been accused of being biased towards certain functional groups (Engel *et al.*, 2020). This is evident from the results depicted in Figure 4.8 and Table 4.4, where every compound analysed reacted differently depending on the derivatisation method used. Therefore, it is important to familiarise oneself with the different derivatisation methods in order to improve or develop analytical methods suitable for the ultimate goal of a study (Orata, 2012). When focussing on the repeatability of derivatisation reactions, Villas-Bôas *et al.* (2011) demonstrated that the variability (and hence the RSD values) of amino acids and nucleotides were noticeably higher with the use of MSTFA with methoximation derivatisation compared to MCF derivatisation. However, in the present study, MCF derivatisation did not produce significantly more amino acids with lower RSD values (compared to TMS derivatisation). Additionally, when considering the repeatability of dicarboxylic acids in the current investigation, TMS derivatisation resulted in lower RSD values for succinic- and fumaric acid compared to MCF derivatisation, which supports the results obtained by Pietrogrande *et al.* (2010). They found that with the use of BSTFA derivatisation

the RSD values of the target compounds were sufficient (RSD < 15%) to quantitatively analyse dicarboxylic acids in atmospheric organic particulate matter (Pietrogrande *et al.*, 2010).

Furthermore, during the analysis of haemolymph, by Killiny *et al.* (2017), it was found that TMS derivatisation (MSTFA with methoximation) resulted in the detection of more metabolites than with MCF derivatisation. Likewise, Gao *et al.* (2009) and Gao *et al.* (2010) illustrated in two different investigations that TMS derivatisation (BSTFA with methoximation) identified 133 compounds (such as carbohydrates, polyols and bile acids), whereas alkyl chloroformate derivatisation identified 73 compounds (such as amino acids, phenolics, and carboxylic acids) in human faecal water samples. Yet, there was no substantial difference in the total compounds detected between the different derivatisation methods in the current study, although MCF derivatisation was the only method to make the detection of α-ketoglutaric acid and isocitric acid easy.

Moreover, with the inclusion of methoximation it is thought to protect α-keto acids and aldehydes against decarboxylation to make their detection possible (Gullberg *et al.*, 2004). This was previously demonstrated in an untargeted analysis by López-Bascón *et al.* (2019) when three α-keto acids were only detected when methoximation was included. Then again, the overall number of compounds detected was more when methoximation was not carried out (López-Bascón *et al.*, 2019). Methoximation also inhibits the cyclisation of open chain and cyclic structures of sugars, thereby preventing the appearance of multiple peaks belonging to an individual sugar compound (Gullberg *et al.*, 2004; Pasikanti *et al.*, 2008; Miyagawa & Bamba, 2019), as proved in a study conducted by Gao *et al.* (2010) on human faecal water. Their results indicated that the formation of several sugar peaks, such as glucose, decreased when methoximation was included with TMS derivatisation (Gao *et al.*, 2010). Similarly, Engel *et al.* (2020) observed a reduction in derivatives per monosaccharide when methoximation was executed prior to TMS derivatisation. However, when considering this in the present study, the inclusion of methoximation with TMS derivatisation seemed redundant since no sugars were included in the stock solution, and the only keto acid (α-ketoglutaric acid) included in the stock solution was not detected by TMS derivatisation. Interestingly lactic acid was only detected by TMS derivatisation when a methoximation step was implemented. Lactic acid has a high polarity (Zhang *et al.*, 2019) requiring the highest temperature to enable volatility (Kiseleva *et al.*, 2021). It can potentially be reasoned that the conditions applied for BSTFA and MSTFA derivatisation in this study were not optimal to ensure the detection of lactic acid. However, in the presence of methoxyamine hydrochloride solution, lactic acid was detected by TMS derivatisation as reported in literature (Roessner *et al.*, 2000; Zarate *et al.*, 2016; Parvatam *et al.*, 2023).

4.5 DERIVATISATION RESPONSE OVER TIME

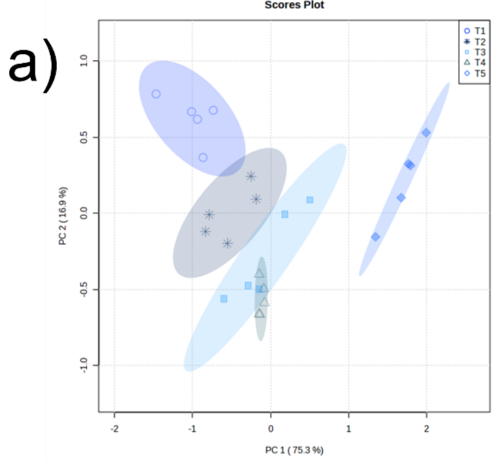
An overview of the samples (n=5) derivatised using BSTFA; BSTFA-MeOX; MSFTA; MSTFA-MeOX and MCF methods, analysed at five time periods (by injecting the same batch at each time period) are shown in Figure 4.9. The re-injection of the last batch per sample set was done to determine the stability of the derivatives produced by each derivatisation method, where T1 (0 h) was analysed upon completion of the batch derivatisation, T2 = 12 h, T3 = 36 h, T4 = 60 h and T5 = 84 h following completion of the derivatisation reaction. In each case, a multivariate assessment of the data acquired by GC-MS analyses is given in the PCA score plots with their corresponding box-like plots. Considering that five individually prepared samples were analysed repeatedly at five timepoints, it is anticipated that some separation will be seen within samples between the different time intervals.

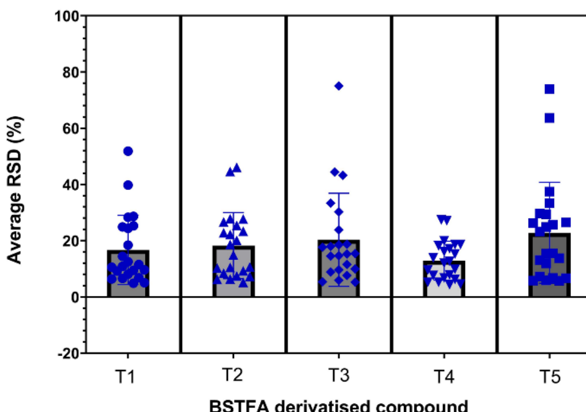
The use of BSTFA as derivatisation method showed steady (similar) repeatability within the first three timepoints (up to 36 hours post derivatisation), with the average RSD values below 20%, then at the 60 hours post preparation a decrease is seen in the calculated RSDs (T4 = average RSD, across compounds, 12.9%) followed by an increase (T5 = average RSD 22.8%) at 84 hours (Figure 4.9a). The overall change in metabolite profile is further reflected in the PCA score plot, where timepoints 1 (○) and 5 (◆) are grouped on the opposite sides of the plot, suggestive of the most different profiles. The PCA plot does show progression of stability over time with the groups gradually shifting from timepoint1 to timepoint 5. A highly similar pattern is seen in the PCA score plot evaluating the use of BSTFA with methoximation as derivatisation method, where timepoints 1 and 5 are grouped on the opposite ends of the plot (albeit in a different direction than the BSTFA data), along with a gradual shift of groups as time progresses (Figure 4.9b). In addition, the corresponding average RSD values for BSTFA-MeOX showed similar beginning and end points, starting with an average RSD < 20% (T1 = RSD 15.4%), followed by an increase (T2, 29.0%), decrease (T3, 20.7%) and increase (T4, 25.2%), and at the last timepoint a very similar average RSD to the first timepoint (T5 = RSD 14.5%).

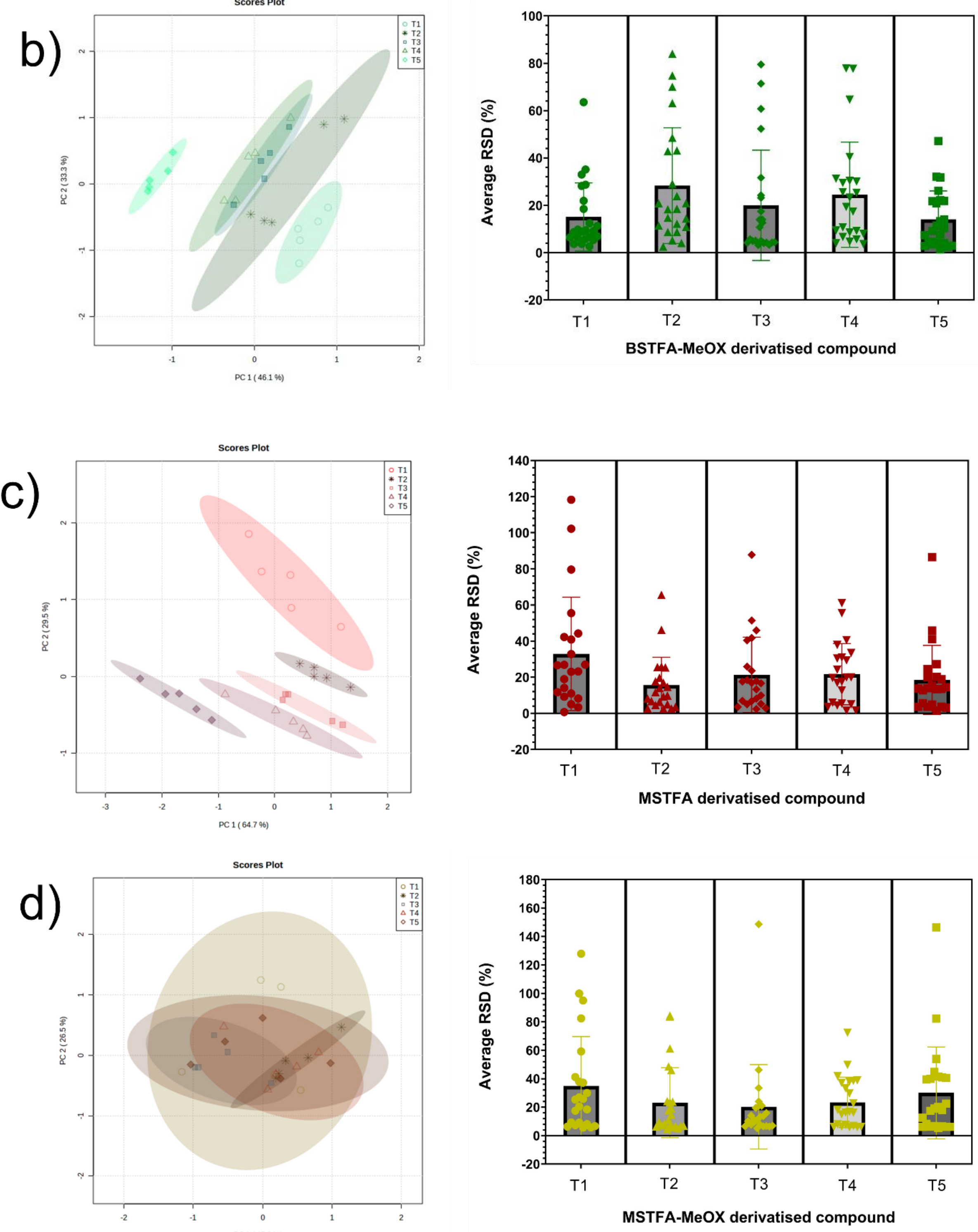
When MSTFA was applied as derivatisation method, the first timepoint showed the worst repeatability (T1 = average RSD 32.9%), given that most compounds have an average RSD value higher than 20%. At 12 hours post preparation a decrease was seen (T2 = average RSD 15.7%), followed by an increase in average RSD at timepoint 4 (T4 = 21.3%), and another decrease at 84 hours (T5 = 18.5%) (Figure 4.9c). The first timepoint (○) is further portrayed by the largest ellipses in the PCA score plot, again suggesting the worst repeatability. As with BSTFA derivatisation, the groups in the PCA score plot gradually shift from timepoint 1 to

timepoint 5 (◆), indicating a change in stability over time. This shift is not displayed in the PCA plot when MSTFA with methoximation derivatisation (Figure 4.9d) was used, rather an overlay of ellipses is seen with the inclusion of methoximation with MSTFA derivatisation, suggesting similar metabolic profiles over the five timepoints. Furthermore, the box-like plots from timepoints 2 (T2 = average RSD 23.1%) to 4 (T4 = average RSD 23.3%) depicted steady repeatability when MSTFA with methoximation derivatisation was applied, with RSD values below 25%, while an increase of calculated RSD values is observed in timepoints 1 (T1 = 34.9%) and 5 (T5 = 30.0%).

The highest average RSD (39.4%) was seen at 0 h when MCF derivatisation was applied suggesting the most variation (Figure 4.9e). As time progressed to 12 hours, the average RSD decreased to 18.8% (T2) and remained relatively steady until 84 hours had passed (T5 = 17.8%). From the PCA score plot, an overlap of metabolite groupings is seen over the five timepoints analysed, suggesting similarity between profiles over time along with reasonable stability over time. Timepoint 1 showed the largest grouping, while timepoint 5 still overlapped with the other batches, in a different direction potentially indicating a change in the metabolite response.






b)
Scores Plot
PC 1 (46.1 %)
PC 2 (33.3 %)
T1
T2
T3
T4
T5
Average RSD (%)
BSTFA-MeOX derivatised compound
c)
Scores Plot
PC 1 (64.7 %)
PC 2 (29.5 %)
Average RSD (%)
MSTFA derivatised compound
d)
Scores Plot
PC 1 (47.6 %)
PC 2 (26.5 %)
Average RSD (%)
MSTFA-MeOX derivatised compound

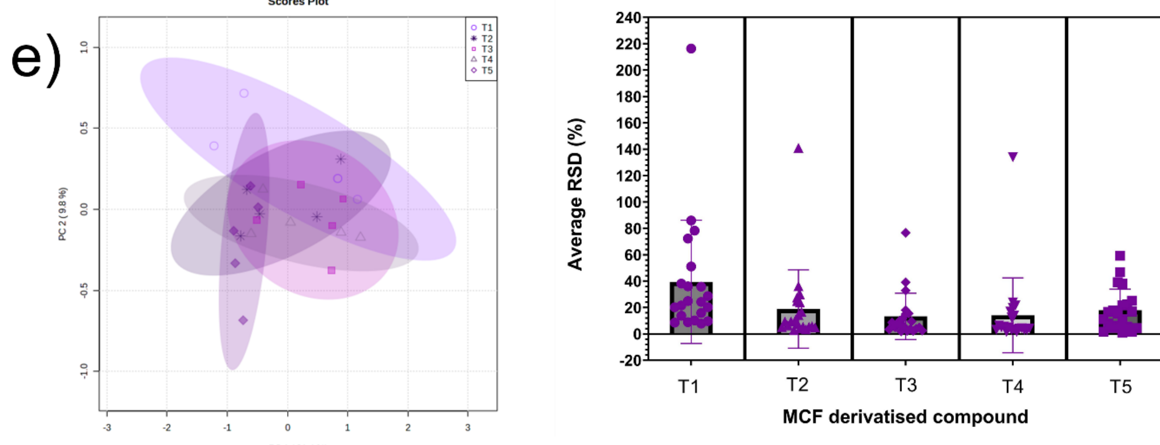


Figure 4.9: An overview of the distribution of the average RSD values for all derivatives of the sample replicates of each derivatisation method a) BSTFA; b) BSTFA-MeOX; c) MSFTA; d) MSTFA-MeOX; e) MCF. This response was measured at five time periods (T1: 0h; T2: 12h; T3: 36h; T4: 60h; T5: 84h) as summarised by the PCA score plots and boxplots.

Derivative stability is an important aspect to acknowledge in metabolomics studies, especially when large-scale analyses are involved and samples reside in the autosampler for long time periods (Carrasco-Pancorbo *et al.*, 2009; Villas-Bôas *et al.*, 2011; Pechlivanis *et al.*, 2014). Ultimately the aim of silylation is to increase the stability of a compound and improve GC behaviour (Gumbi *et al.*, 2019). The same is true for alkylation, where increased compound stability is one of the advantages of using alkylation, or more precisely, MCF derivatisation (Casas & Matamoros, 2021). The use of BSTFA derivatisation (Figure 4.9a) showed stable results within the first 3 timepoints analysed, but after 36 hours a decrease in average RSD was experienced. The initial higher average RSD values can be ascribed to the derivatisation nature of some derivatives suggesting that longer silylation times or different incubation temperatures might be needed to ensure a complete reaction for some of the compounds. For example, a derivatisation temperature of 50°C and reaction time of 30 min was found to be best for the analyses of organic acids when using BSTFA with TMCS as reagents (Christou *et al.*, 2014)

In the current study when MSTFA and MSTFA-MeOX were used as derivatisation methods a decrease in average RSD values was seen, indicating an increase in stability, after 12 hours. This suggests that longer derivatisation times are likely needed for complete derivatisation when using MSTFA. However, literature supports the notion that MSTFA is more reactive than BSTFA, achieving complete derivatisation at a faster rate (Parkinson, 2012; Zhu *et al.*, 2017). A study conducted by Villas-Bôas *et al.* (2011) suggested variable degrees of instability over 72 hours for amino and non-amino organic acids when derivatised by MSTFA with methoximation. Similarly, the response seen in the current study shows a variety of RSD

values for the pre-selected compounds analysed when using MSTFA over 84 hours. Moreover, in the present study when MSTFA with methoximation derivatisation was used, an overlap in the ellipses of the PCA score plot were seen between batches suggesting similar metabolite profiles, unlike the plot seen when just using MSTFA, where all five timepoints group away from one another. Likewise, Engel *et al.* (2020) illustrated on a mixture of standards that more stable signals were achieved for the metabolites glycine and lysine when MSTFA with methoximation was applied as derivatisation method. However, the increase in derivatisation time was not beneficial for the other compounds detected in the mixture of standards (including amino acids, organic acids and sugars) (Engel *et al.*, 2020). In the current study, outliers were also present in the box-like plots suggesting that some metabolites perform better at different derivatisation times.

The use of MCF derivatisation in the present study showed a decrease in the average RSD values after 12 hours and remained relatively steady (RSD < 20%) until 84 hours had passed. It was anticipated that the MCF derivatisation reaction would reach completeness faster due to its rapid derivatisation rate (Bhumireddy *et al.*, 2021), yet a decrease in average RSD values were only seen after 12 hours of derivatisation, suggesting that some compounds have not been completely derivatised within timeframes used in the applied method. Additionally, when using MCF, single derivatives were produced in this study, for compounds such as glycine and phenylalanine making the MCF derivatives more stable than some of the TMS derivative counterparts, which produced multiple peaks for those compounds. In a study conducted by Villas-Bôas *et al.* (2011), they compared the stability of amino acids, organic acids, and sugars (represented by RSD values), over 72 hours, following alkylation (MCF) and silylation (TMS). From their study it was concluded that MCF derivatives (excluding. alanine) have greater stability over time (RSD < 10%) compared to the TMS counterparts (RSD > 10%) (Villas-Bôas *et al.*, 2011). Based on the average RSD outcomes of the current study, the same is true where MCF derivatisation showed lower RSD values than TMS derivatisation at timepoint 5 (reaching over 72 hours as in the study by Villas-Bôas *et al.* (2011). This is a useful observation in scenarios where samples are kept in the autosampler for extended times, suggesting that good repeatability will still be achieved within the results if MCF derivatisation was applied.

4.6 SUMMARISED COMPOUND RESPONSE OVER TIME

In Figure 4.10 the average normalised abundance of each derivative (as measured in five samples) per five timepoints, per derivatisation method, is summarised. Herein it is clear that the abundances of TMS derivatised compounds are higher than their MCF counterparts. Next, the MCF derivatisation method showed better stability, with most of the derivatives exhibiting highly similar values as seen by the display of data points in a relatively straight line. The

results from the TMS derivatisation methods showed more patterns, suggesting varying responses over time.

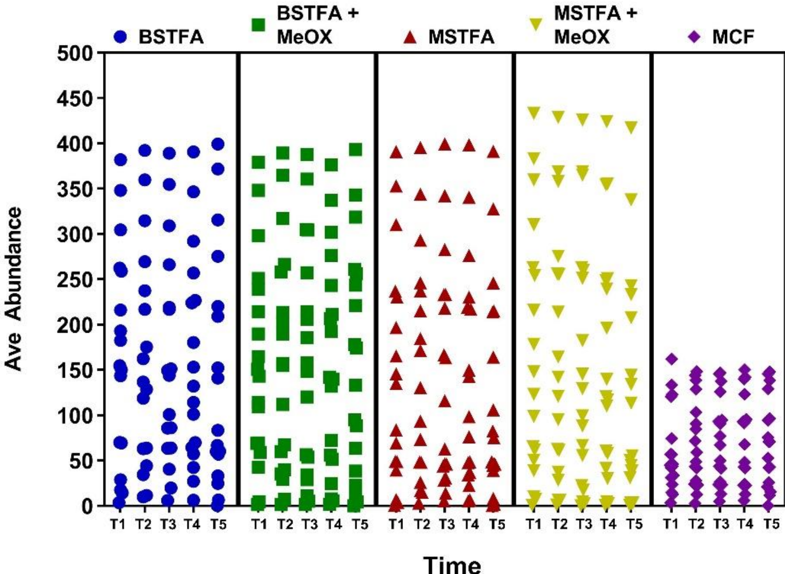


Figure 4.10: Average normalised abundance per derivative (as determined from the analyses of five samples) measured over time derivatised by BSTFA (●), BSTFA-MeOX (■), MSTFA (▲), MSTFA-MeOX (▼) and MCF (♦) methods to infer stability.

The similar derivative abundances produced by MCF derivatisation over time are suggestive of better stability of compound derivatives when using this derivatisation method (Figure 4.10). It has been reported that MCF derivatisation showed improved compound stability compared to TMS derivatisation methods when focusing on amino and non-amino organic acids (Rohloff, 2015). Interestingly a study focusing on yeast samples deemed the use of BSTFA as the preferred derivatisation method due to the stability of the resulting TMS derivatives. Herein they reported that GC-MS analyses should happen immediately after derivatisation to obtain the most reproducible results (Williams *et al.*, 2021). When considering only the results obtained by BSTFA derivatisation in the current study, an increase or decrease in the average abundance of the derivatives is experienced following the first timepoint. Changes in the samples can occur following derivatisation due to compound degradation, as seen with the analyses of psychoactive substances where the storage of samples in a cooled autosampler was suggested to minimise analyte losses (Woźniak *et al.*, 2020). Arguably, the results of this study show that changes could have occurred between timepoints 2 and 5 due to sample degradation, not only when using BSTFA derivatisation, but also when utilising the other derivatisation methods. This is indeed an area of method performance that will benefit from future studies. Also, within Figure 4.10 high abundances were obtained following TMS derivatisation methods, as previously reported in serum (Tang *et al.*, 2023) and plasma (Zarate *et al.*, 2016) samples. Comparisons of peak intensities between MCF and TMS derivatisation

methods are scarce in literature deeming the results from the current study as an interesting aspect for future studies aimed at comparing derivative signals.

4.7 INDIVIDUAL COMPOUND RESPONSE OVER TIME

The average normalised abundance of each derivative (as measured in five samples) was plotted next (Figure 4.11) to visualise their stability over 84 hours (5 batches) produced by the various derivatisation methods (BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF). Compounds such as 2-hydroxybutyric acid, succinic acid and valine showed similar values over time, irrespective of the derivatisation method used. Then again, the abundances of compounds like aspartic acid, lactic acid and methionine were closely related when derivatised with MCF, while the stability of compounds like coumaric acid, fumaric acid and lysine were comparable when TMS derivatisation methods were used. Interestingly, the abundance of some derivatives decreased as time progressed, for example, the first timepoint of the compound glycine showed the highest abundance when derivatised with BSTFA, BSTFA-MeOX and MSTFA, but decreased with each successive timepoint. The opposite is also true where the abundance of some derivatives increased with increasing time, as illustrated by tryptophan when derivatised with MCF. The compound glutamine showed better stability when methoxyamine was included in the derivatisation reaction, while citric acid reflected more stable results when using BSTFA and MSTFA without methoxyamine. Other derivatives such as cystine, glutamic acid and phenylalanine showed varying responses amongst the tested times and derivatisation methods.

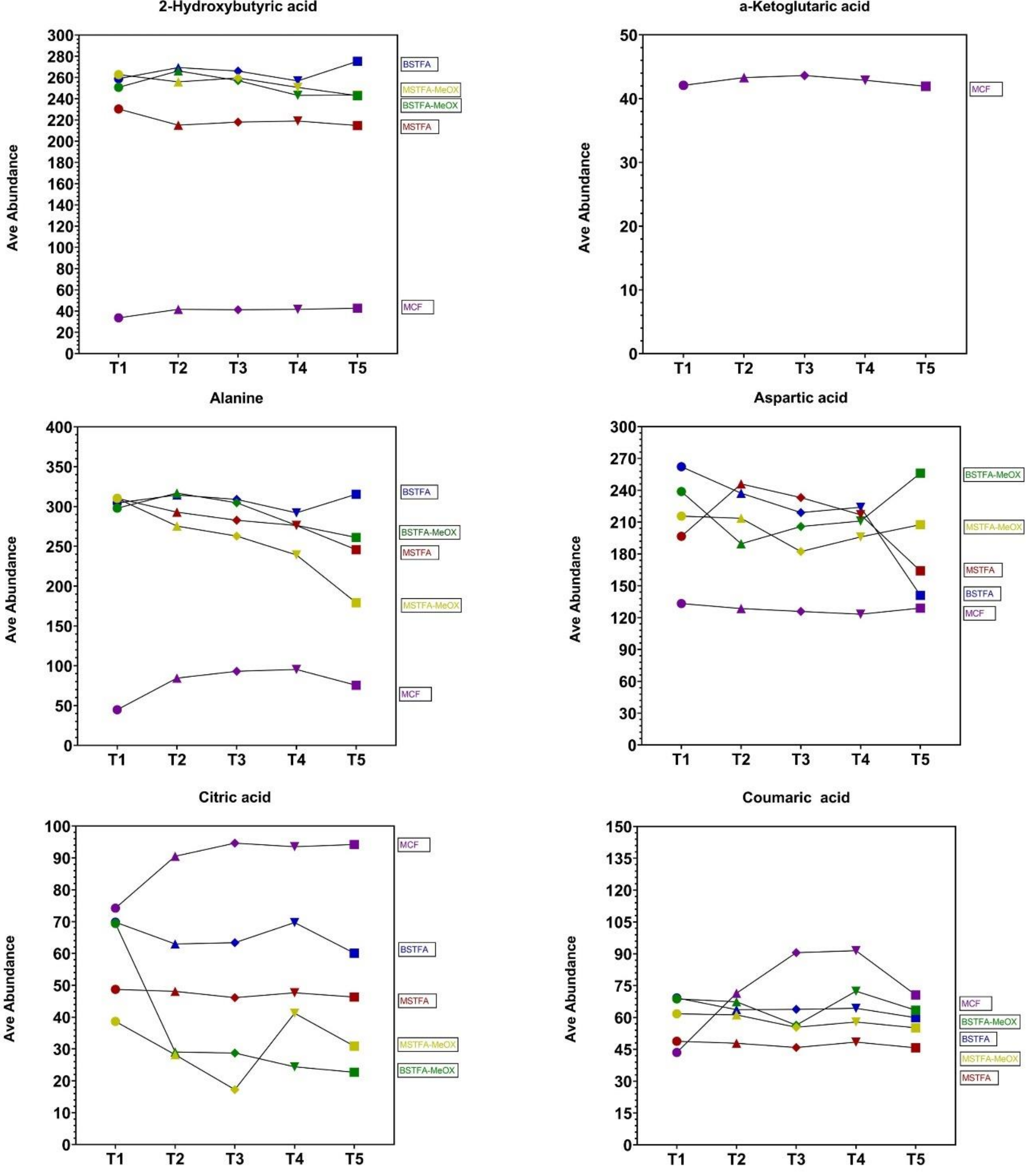


Figure 4.11: Compound stability (average normalised abundance) measured over time derivatised by BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF. 1(●): 0 h; 2(▲): 12 h; 3(♦): 36 h; 4(▼): 60 h; 5(■): 84 h.

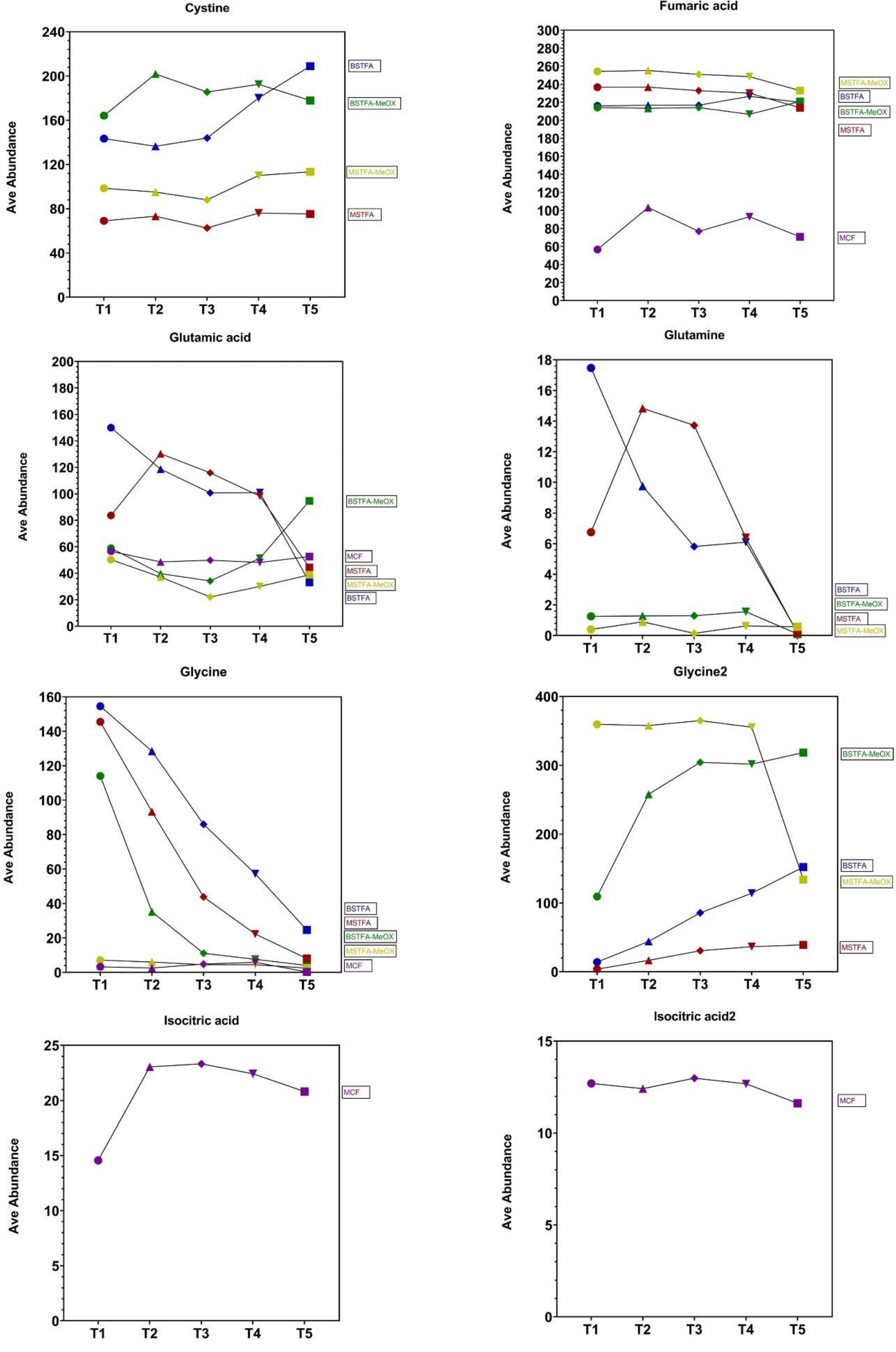


Figure 4.11: Continued.

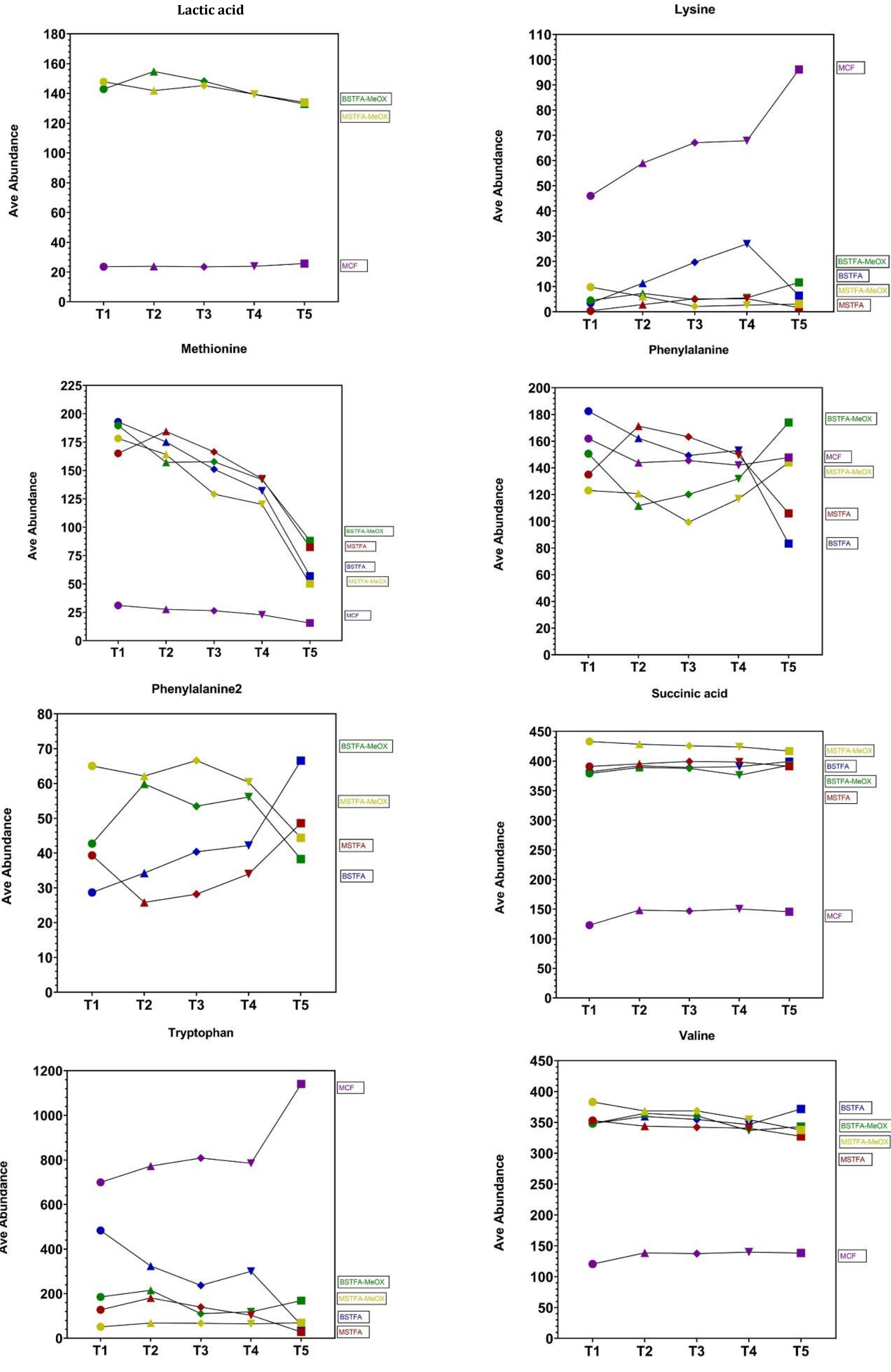


Figure 4.11: Continued.

The efficiency of a derivatisation method is determined by, among other things, the stability of the produced derivatives (Villas-Bôas *et al.*, 2005). This is evident in the present study where the average abundance obtained from normalised data, per compound, per derivatisation method, describes the stability over time (Fig. 4.11). From these results it is true that most derivatives showed a unique response over time when derivatised with the various reagents. Yet, specific compounds showed better stability over time with specific derivatisation methods.

Good stability over time, irrespective of the derivatisation method implemented, was seen in the current study with compounds such as 2-hydroxybutyric acid, succinic acid, and valine, suggesting that these compounds are not sensitive to changes following derivatisation with any reagent or method. In a study on plant metabolites, the use of MSTFA-MeOX resulted in stable quantitative values for fumaric acid and a range of sugars (Quéro *et al.*, 2014). The TMS derivatives of fumaric acid showed relatively good stability in the present study, with little changes observed in the abundance over the 84-hour timeframe. Next, MCF derivatisation resulted in good stability for amino and non-amino organic acids in a different study (Villas-Bôas *et al.*, 2011). Ultimately, metabolite classes show differences in stability of derivatised metabolites necessitating optimisation for the metabolites of interest. Literature reports that class 3 metabolites (with amide, thiol, or sulfonic functional groups) are more likely to decompose in the analytical system (Koek *et al.*, 2006). This can be seen as a reason why methionine in the current study declined over time when derivatised with any reagent.

A second pattern observed in the stability results in the current study, is a decline in average abundance amongst the analysed batches considering various reagents. While looking at the peak area of the metabolite glutamine and glutamic acid following MSTFA-MeOX derivatisation, Miyagawa and Bamba (2019) concluded that variation can be attributed to the methoximation process. Yet, in the current study the addition of the methoximation process produced more stable results compared to the non-methoxymated samples for these two compounds. The different response can likely be attributed to incubation time and temperature of the methoximation step which was performed for 60 min at 60°C in this study, compared to the 90 min at 37°C from their study. In a different study the metabolites glutamine, glutamic acid and methionine were reported as instable, following quantitative MCF derivatisation (Tumanov *et al.*, 2016). The result from the current study contradicts these findings, with glutamic acid and methionine showing good stability while glutamine was not detected after MCF derivatisation. Furthermore, a 50% loss in abundance after 12 hours in the compounds lysine and phenylalanine were detected when derivatised with MSTFA-MeOX (Quéro *et al.*, 2014). In the current study lysine also showed around a 50% decline in abundance after 12 hours, with stable results further detected. Then again, phenylalanine declined until the third time point whereafter an increase was observed when using MSTFA-MeOX. Noctor *et al.*

(2007) reported that batches of TMS derivatives had to be analysed within a period of 2 hours following derivatisation to compensate for instability (Noctor *et al.*, 2007). Moreover, in a study by Roessner *et al.* (2000) the stability of compounds including amino acids, organic acids and sugars, were measured every 2 hours within a 30-hour timeframe. The results indicated that, among all the compounds, the abundances of tryptophan and glutamine decreased with time when MSTFA was used (Roessner *et al.*, 2000). Likewise, in the current investigation after an initial increase is seen in the abundances of tryptophan and glutamine derivatised with MSTFA, decreases follow from the second timepoint onwards. Considering varying responses over time, per derivatisation method, it is advised to perform a stability test and determine where the limitations of the study at hand are. For example, if derivatives are unstable, derivatisation of smaller batches can help to prevent derivative changes. However, this can induce more variability among the injected batches (Zarate *et al.*, 2016).

4.8 BIOLOGICAL SAMPLES

The use of blood/serum as sample medium in metabolomics studies has facilitated the discovery of novel biomarkers, which contributed to improved disease diagnosis, prognosis and therapy (van Ravenzwaay *et al.*, 2007; Zhang, 2012b; Yin *et al.*, 2015). Serum is the liquid portion of blood and is prepared by a coagulation process (Denery *et al.*, 2011; Liebenberg *et al.*, 2021; Sotelo-Orozco *et al.*, 2021). Blood/serum contains a vast diversity of metabolites, making it rich in biological information (Yin *et al.*, 2015; Bi *et al.*, 2020; Doğan *et al.*, 2021; Zhu *et al.*, 2023). However, blood/serum needs to be deproteinised before GC-MS analysis (Álvarez-Sánchez *et al.*, 2010). Variations in metabolite concentrations may be indicative of many diseases, hence serum metabolomics has played an integral part in the fields of physiology, diagnostics, functional genomics, pharmacology, toxicology and nutrition (Winder *et al.*, 2011; Zhang *et al.*, 2012; Doğan *et al.*, 2021; Vignoli *et al.*, 2022).

In this section a subset of ERNDIM serum samples (n=10), grouped as A and B (section 3.2.4) were derivatised using a) BSTFA; b) BSTFA-MeOX; c) MSFTA; d) MSTFA-MeOX and e) MCF methods. The data were processed as per in-house metabolomics workflow methods (section 3.2.7) resulting in the PCA score plots as seen in Figure 4.12. As previously described, PC1 represents the intergroup variation while PC2 accounts for intragroup variation. From the data groupings the samples classified as group B (●) showed larger variation than the samples of group A (■), explained by the larger ellipses seen. All the derivatisation methods used resulted in clear separation between groups A and B, depicting distinct differences within the metabolite profiles of the biological samples analysed. Yet, the samples derivatised with BSTFA-MeOX showed the highest score within PC1 (68%), suggesting the largest differences between groups A and B, while the use of MSTFA showed the lowest score (55.8%). Also, the use of

BSTFA-MeOX resulted in the lowest PC2 score of 9.1%, suggesting the smallest within-group variation, while MSTFA showed the highest value of 16% on PC2.

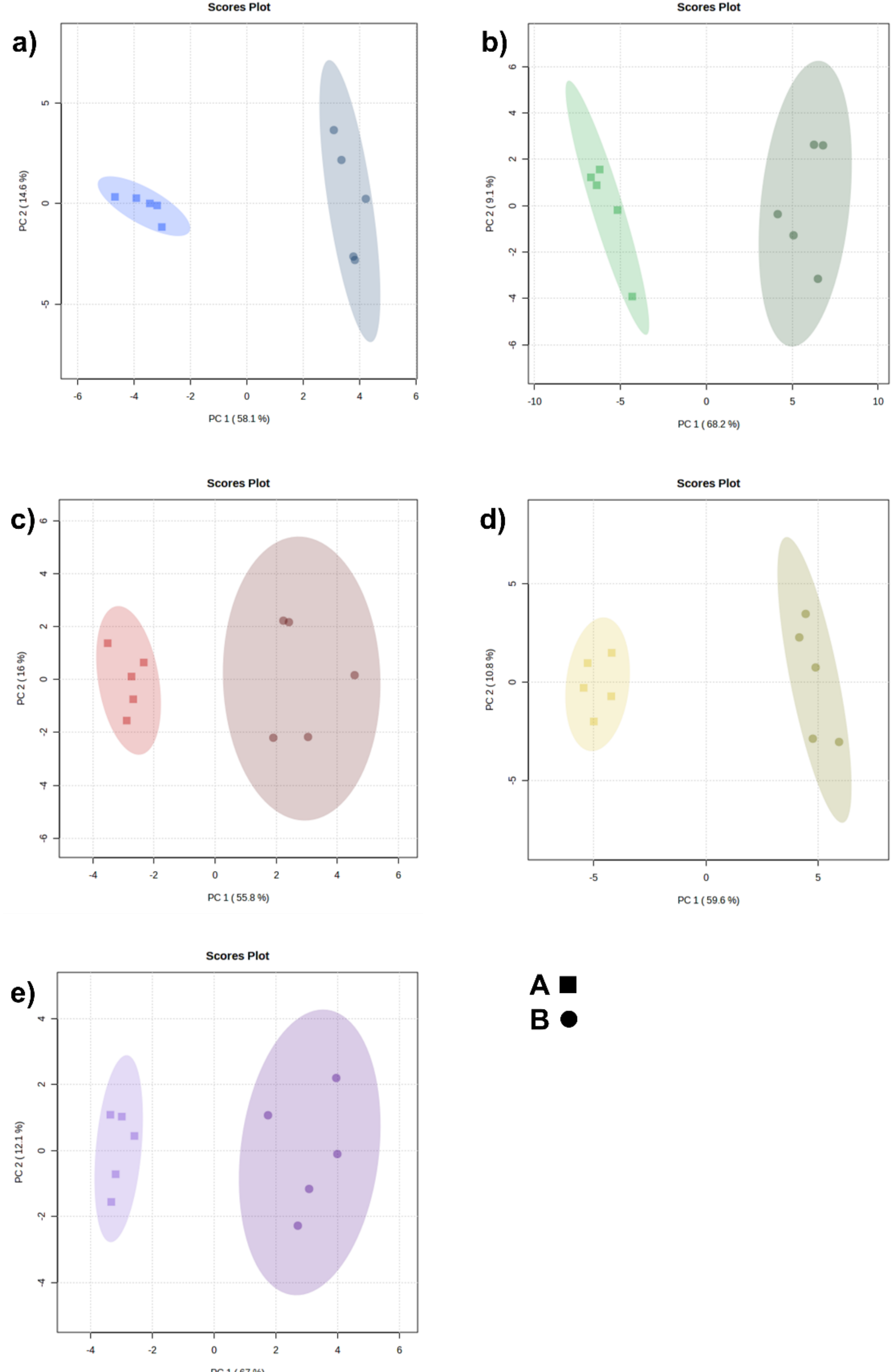


Figure 4.12: The PCA score plots obtained from analyses of biological samples derivatised with: a) BSTFA; b) BSTFA-MeOX; c) MSFTA; d) MSTFA-MeOX; e) MCF; Group A (■), Group B (●), n=5 per group.

The use of PCA analyses to indicate differentiating groups are well reported in literature (Zeki Ö *et al.*, 2020). It is important to consider the influence of biological variability on these groups during analysis, given that genetics, diet, lifestyle, the environment and more can have an impact on metabolic profiles. Post-collection procedures of samples can also play a role, such as storage, extraction and derivatisation (Crews *et al.*, 2009). Considering this, Roessner *et al.* (2000) demonstrated that the variability found in their results obtained for MSTFA with methoximation derivatised potato tuber samples, were mainly caused by the variability within the biological samples. The variation detected in the biological data within the current study, can largely be attributed to biological variation as the same pattern repeats for all derivatisation types tested. Herein group A falls towards the left side of the PCA score plots, and group B on the right side. Also, in all cases, group A has a smaller grouping (confidence ellipse) compared to group B, which is more stretched out. Seeing that a PCA summarises the main patterns of variation in the data and allows one to determine how one sample is different from another (Chong *et al.*, 2019; Pang *et al.*, 2022), it enables an overview of the current data obtained via different derivatisation techniques. Yet, as the purpose of this study is not to infer biological meaning to the results and to determine which variable contributes most to the differences, the PCA scores reported in Figure 4.12 are mostly to provide an overview of the data. Following derivatisation of the same set of samples with BSTFA; BSTFA-MeOX; MSFTA; MSTFA-MeOX; and MCF, highly similar metabolic results were obtained. When considering the within group variation as seen by PC2 it becomes apparent that the use of BSTFA with methoximation resulted in the lowest variation (PC2 9.1%) within a group, suggesting the least number of influences from external (non-biological) factors. Similarly, a study on soil microbial communities showed lower variability when using BSTFA with the inclusion of methoximation, compared to only BSTFA as derivatisation method (Willers *et al.*, 2016).

To further evaluate the performance of the different derivatisation methods with regards to the analyses of biological samples, Table 4.5 was composed. Herein, the number of features detected following in-house data processing methods (section 3.2.6), the number of statistically significant features following t-test analyses (section 3.2.7) and the number of features classified as unknowns (section 3.2.6) are given. From the derivatisation methods utilised the use of BSTFA with methoximation resulted in the largest number of features detected, the most statistically significant features and the most unknown features. Then again, the use of MSTFA resulted in the lowest numbers across all three accessed factors.

Table 4.5: The comparison of the total, unknown and statistically significant features detected, derivatised by BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF.

Method	Number of features detected	Number of statistically significant features	Number of unknown features
BSTFA	57	31	36
BSTFA-MeOX	119	76	86
MSTFA	32	14	19
MSTFA-MeOX	82	44	54
MCF	57	26	31

The goal of an untargeted metabolomics analysis approach is to measure as many metabolites as possible in an unbiased fashion (Johnson *et al.*, 2016; Wang *et al.*, 2019; Villaret-Cazadamont *et al.*, 2020). In the current investigation, when BSTFA with methoximation derivatisation was used, the highest number of features (119) were detected. Yet, in a study conducted by Moros *et al.* (2017), the most features (1419) were detected when MSTFA with methoximation derivatisation was applied on blood plasma samples (compared to BSTFA with methoximation derivatisation). Likewise, He *et al.* (2018) also concluded that more compounds can be detected when using MSTFA-MeOX compared to using BSTFA-MeOX. Then again, no significant difference was seen in the number of peaks detected by BSTFA with methoximation derivatisation or MSTFA with methoximation derivatisation, after an investigation by Abbiss *et al.* (2015) on urine samples. Furthermore, when MSTFA with methoximation derivatisation was compared to MCF derivatisation by Villas-Bôas *et al.* (2011), more peaks were detected after MCF derivatisation of culture medium samples. The opposite is true in the present study, given that more features were detected by MSTFA with methoximation derivatisation (82 detected features), than with MCF derivatisation (57 detected features). It should be highlighted that the number of features detected can vary from case to case depending on the data processing parameters implemented. For example, the features detected in a specific data matrix are influenced by steps such as noise filtering, peak detection and matching, retention times, normalisation etc. (Karaman, 2017). Hence the above comparisons are made based on final results, not considering the various pre-processing steps utilised which can also affect the outcome.

Once a peak list has been obtained, the metabolite (feature) concentrations that differ statistically can then be used to distinguish between the investigated groups (Peisl *et al.*, 2018; Yang *et al.*, 2021). The use of univariate or multivariate methods can help to identify variables that are significantly different amongst conditions. A list of features that have changed significantly under different conditions is utilised for data interpretation or to generate a hypothesis (Chong *et al.*, 2019). Following t-test analyses the use of BSTFA-MeOX resulted in the highest number of statistically significant features between groups A and B, while the use of MSTFA resulted in the lowest number of features. After GC-MS profiling of leukaemia

cells it was reported that BSTFA-MeOX resulted in higher metabolome coverage than MSTFA-MeOX, and was chosen as the derivatisation method of choice for further leukaemia metabolomics studies (He *et al.*, 2018). Then again, the use of BSTFA and MSTFA showed similar significant peaks detected when performing profiling of urinary metabolites (Pasikanti *et al.*, 2008). More often than not different derivatisation types will be applied to the same study to expand the coverage of the metabolome. Such examples of using TMS and MCF derivatisation have been reported when profiling fermented avocado seeds (Zhao *et al.*, 2023), and wine yeast (Pinu *et al.*, 2019). A similar approach can be considered in the current investigation of serum samples to enhance the metabolome coverage, if biological interpretation was the desired outcome, in future studies.

Lastly, the number of unknown features detected following derivatisation with all five methods is listed in Table 4.5. An untargeted metabolomics approach can be challenging due to the resulting complex data set and identification of unknowns (Johnson *et al.*, 2016; Pereira Braga & Adamec, 2019). A feature in a metabolomics data set that is unknown could represent a novel compound, though it can also be a redundant signal, artefact, or contaminant that is not included in metabolomics libraries and databases (Peisl *et al.*, 2018; Sindelar & Patti, 2020). Herein the largest number of unknown features were detected with the use of BSTFA-MeOX while the use of MSTFA resulted in the lowest number. This coincides with the other factors also evaluated (i.e., total number of features and the number of statistically significant features). It has been reported that the profiling of TMS derivatives include extensive unknowns (Rohloff, 2015), as seen in the current study. Yet, Parvatam *et al.* (2023) concluded that MSTFA-MeOX was superior to BSTFA-MeOX due to the wide range of identified metabolites. Yet, considering the results obtained in the current study the use of BSTFA with methoximation as derivatisation method was superior in terms of smallest variation within a group (PC2 9.1%) and number of features detected (119). However, since derivatisation methods have been accused of being biased towards certain functional groups (Engel *et al.*, 2020), the metabolome coverage is expanded by applying different types of derivatisation methods. Also, as mentioned, the data processing parameters implemented can have an influence on the final results acquired. Therefore, the choice of derivatisation method should be carefully considered and standardised for the study in question.

CHAPTER 5: CONCLUSIONS AND FUTURE PROSPECTS

5.1 INTRODUCTION

In untargeted metabolomics studies, the aim is generally to detect the maximum number of metabolites present in a specific sample type without prejudice. Gas chromatography-mass spectrometry is often seen as a gold standard platform in most laboratories for metabolomics studies. However, only volatile compounds can be analysed by GC-MS, therefore, derivatisation is mostly a necessary step in sample preparation to volatilise semi-volatile and non-volatile metabolites. Given the biased nature of derivatisation methods toward certain functional groups, it is important to familiarise oneself with the different methods, and to determine how they influence the metabolites of interest, to choose the most suitable method for the experiment at hand (Chapter 1). To obtain insight into the metabolite response of amino acids and small organic acids, when subjected to different derivatising reagents, five different derivatisation methods (BSTFA; BSTFA-MeOX; MSFTA; MSTFA-MeOX and MCF) chosen from literature and in-house experience were implemented within this study (Chapter 2). A stock solution of 19 pre-selected compounds (and four internal standards 2-amino-3-bromo-5-methylbenzoic acid, 3-phenylbutyric acid, norleucine and nonadecanoic acid methyl ester) as well as two sets of pre-extracted biological samples were analysed via GC-MS (Chapter 3). For the evaluation of the standard solution, comparisons focused on metabolite response, repeatability, and stability. While evaluations of the biological samples were assessed in terms of natural grouping and number of detected features (Chapter 4). Lastly, the summarised findings and future recommendations are given (Chapter 5), linking back to the aim of this study which was to: **Compare pre-selected derivatisation techniques on a set of key metabolites including a range of amino acids and small organic acids as a prerequisite of gas chromatography-mass spectrometry metabolomics analyses.**

5.2 CONCLUSIONS

Here I summarise the major findings reported in Chapter 4 in the context of the objectives listed in Chapter 1.

The initial focus of the present study was to investigate different derivatisation reagents used prior to performing GC-MS analyses, and to choose the most used methods based on the literature, as stated in the **first objective.** To simplify the selection, a table was used to showcase the various literature-based derivatisation reagents when considering a list of specific metabolites. From this table, the use of BSTFA, MSTFA (with and without the use of MeOX prior to silylation) and MCF were identified as reagents to use in this study, when comparing the performance of a derivatisation method.

The **second objective** was achieved by preparing the selected compounds as individual standards, creating a stock solution, and analysing the stock solution using GC-MS, after TMS and MCF derivatisation. The TMS and MCF detected derivatives were tabulated per individual compound, stating the retention time and target ion. The results obtained facilitated the setup of SIM-Scan methods, ensuring detection of the corresponding TMS and MCF derivatives, which were implemented as the methods used for the subsequent GC-MS analyses.

The derivatisation methods (BSTFA, BSTFA-MeOX, MSTFA, MSTFA-MeOX and MCF) were compared by derivatising and analysing pre-aliquoted samples from a standard stock solution. The results obtained helped to reach the **third objective**. An overview of the acquired results per derivatisation method was visually displayed via PCA score plots. The most striking result was the clear separation between silylated and alkylated samples on the same PCA plot, suggesting different GC-MS profiles of the same set of compounds. Additionally, the results indicated similar groupings of BSTFA and MSTFA derivatised batches with the inclusion of MeOX. While the exclusion of MeOX to both BSTFA and MSTFA derivatisation resulted in higher variation amongst the analysed batches. Further comparisons were made based on compound response and repeatability (average RSD%) per derivatisation method. Overall MCF derivatisation resulted in the lowest RSD values for the largest number of compounds. Additionally, each derivatisation method was reviewed separately focusing on the repeatability of each individual derivative. The use of silylation (BSTFA and MSTFA) without the methoximation step resulted in the same number of compounds detected, the formation of multiple derivatives for the same compounds and displayed very similar repeatability values for the same compounds. Interestingly without the methoximation step, neither of these methods were able to detect lactic acid, highlighting the importance of methoximation to stabilise the keto-groups of compounds. The silylation methods with the inclusion of the methoximation step also resulted in a similar number of compounds detected, the formation of multiple derivatives for the same compounds and similar repeatability. However, BSTFA-MeOX performed slightly better than MSTFA-MeOX due to the fact that more compounds had RSD values < 50%. Evaluation of alkylation with MCF resulted predominantly in the production of only single derivatives per compound, (except for isocitric acid). This finding is extremely important since the formation of multiple derivates per compound is a well-known problem when using silylation as a derivatisation method in metabolomics studies. The average RSD value per method per compound was also compared in a table demonstrating that the compound of interest should guide the derivatisation choice, as different compounds showed higher repeatability (lower RSDs) with different derivatisation methods.

One of the goals of silylation and alkylation is to improve the stability of compounds. Therefore, to achieve the **fourth objective**, a batch of analysed samples from each derivatisation method

was re-injected an additional four times (over a period of 84 hours) to compare the stability of the derivatives produced by each derivatisation method over time. Principal component analysis score plots and boxplots were used to visually summarise the results. Varying average RSD values were seen over the 84 hours for most of the derivatisation methods used. However, the use of MSTFA with methoximation and MCF derivatisation showed overlapping ellipses within their specific PCA score plots suggesting similar metabolic profiles and compound stability over the five timepoints. Also, the use of MSTFA, MSTFA-MeOX and MCF showed the highest variation within the first batch of samples analysed (timepoint 1), which might indicate that derivatisation was not completed by the time of analyses. As a next evaluation, the average normalised abundance of each derivative measured over time, per derivatisation method was summarised. Overall, the results indicated that for this set of compounds, MCF derivatisation produced the most stable derivatives as the data points form a relatively straight line. However, TMS derivatives displayed higher abundances. In addition, when focussing on the individual compounds, it yet again highlighted the influence of derivatisation method on a compound of interest. For example, aspartic acid and glutamic acid showed consistent responses over time when derivatised with MCF, while fumaric acid and coumaric acid showed similar responses when TMS derivatisation methods were used. The present results provide much-needed information for large-scale metabolomics analyses, given that samples reside in the autosampler for long time periods and derivatives need to be stable to acquire reliable results.

As stipulated by the **fifth objective**, two sets of biological samples were used to evaluate and compare the derivatisation methods as proof of concept for metabolomics studies. The sample sets were compared based on their group variation and patterns (PCA score plots), as well as number of features detected. Given that the grouping patterns (groups A and B), seen in the PCA score plots, were similar for all derivatisation methods tested, it can be concluded that the variation detected can be largely attributed to biological variation. In addition, the application of BSTFA-MeOX, detected the highest number of features that were statistically significant between groups A and B, enabling the biggest dataset for use in the next step of the metabolomics workflow.

5.3 FINAL REMARKS

To conclude, this study confirmed that the derivatisation reagent is a key factor affecting the repeatability and intensity of individual compounds. For example, the most compounds had the lowest RSD values, meaning high repeatability, when derivatised with MCF, yet the peak intensity was the highest when derivatised with BSTFA. However, when derivatising biological samples the use of BSTFA-MeOX resulted in the highest number of detected features. To

expand the coverage of the metabolome, the application of different derivatisation types may be required in the same study, since there is no “universal” derivatising agent available that can produce satisfactory results for every compound, independent of its chemical class. This is proved in the current investigation when better repeatability was obtained for organic acids, succinic- and fumaric acid when TMS derivatisation was performed and better repeatability for citric- and lactic acid with MCF derivatisation.

Although no superior derivatisation method was identified, it can be concluded that the aim of this investigation was reached. As this study successfully compared pre-selected derivatisation techniques on a set of key metabolites as a precursor of gas chromatography-mass spectrometry metabolomics analyses.

5.4 FUTURE RECOMMENDATIONS

The following recommendations should be viewed as research initiatives directed towards future research aimed at derivatisation for GC-MS metabolomics studies.

- Although α-ketoglutaric acid was added to the standard stock solution as one of the ways to evaluate the methoximation step prior to silylation, it could be beneficial to add more compounds (such as, other ketoacids and carbohydrates) to test the usefulness of MeOX to prevent the formation of multiple peaks.
- To eliminate unwanted variation and decrease the relative standard deviation, the instrument repeatability should be thoroughly evaluated. To do so, a sample from each derivatisation method can be injected multiple times. The relative standard deviation can then be calculated from the data obtained to determine the instrument repeatability.
- Investigation into other factors, such as derivatisation reagent volume, and incubation temperature and time, may improve derivatisation performance and provide conditions where completed derivatisation for most compounds are achieved.

To determine how different derivatisation methods would influence the biological results of a study a higher number of biological samples can be assessed in a next study. By adding more samples, one increases the statistical power of the study, which also aids compound identification, making it possible to infer biological meaning to the results.

REFERENCES

Abbiss, H., Rawlinson, C., Maker, G.L. & Trengove, R. 2015. Assessment of automated trimethylsilyl derivatization protocols for GC–MS-based untargeted metabolomic analysis of urine. *Metabolomics*, 11(6):1908-1921.

Adamski, J. 2020. Introduction to metabolomics. In. *Metabolomics for Biomedical Research*: Elsevier. pp. 1-15.

Alonso, A., Marsal, S. & Julià, A. 2015. Analytical methods in untargeted metabolomics: state of the art in 2015. *Frontiers in bioengineering and biotechnology*, 3:23.

Álvarez-Sánchez, B., Priego-Capote, F. & de Castro, M.L. 2010. Metabolomics analysis II. Preparation of biological samples prior to detection. *TrAC Trends in Analytical Chemistry*, 29(2):120-127.

Antonelli, J., Claggett, B.L., Henglin, M., Kim, A., Ovsak, G., Kim, N., ... Watrous, J.D. 2019. Statistical workflow for feature selection in human metabolomics data. *Metabolites*, 9(7):143.

Armitage, E.G., Godzien, J., Alonso-Herranz, V., López-Gonzálvez, Á. & Barbas, C. 2015. Missing value imputation strategies for metabolomics data. *Electrophoresis*, 36(24):3050-3060.

Ashrafian, H., Sounderajah, V., Glen, R., Ebbels, T., Blaise, B.J., Kalra, D., ... Salek, R.M. 2021. Metabolomics: The Stethoscope for the Twenty-First Century. *Medical Principles and Practice*, 30(4):301-310.

Azizan, K.A. & Baharum, S.N. 2012. Metabolic profiling of *Lactococcus lactis* under different culture conditions. *Molecules*, 17(7):8022-8036.

Bani Rashaid, A., Jackson, G. & Harrington, P. 2014. Quantitation of amino acids in human hair by trimethylsilyl derivatization gas chromatography. *Mass Spectrometry. Enliven: Biological Analytical Techniques*, 1(1):002.

Barceló-Barrachina, E., Santos, F., Puignou, L. & Galceran, M. 2005. Comparison of dimethylformamide dialkylacetal derivatization reagents for the analysis of heterocyclic amines in meat extracts by gas chromatography–mass spectrometry. *Analytica chimica acta*, 545(2):209-217.

Bartel, J., Krumsiek, J. & Theis, F.J. 2013. Statistical methods for the analysis of high-throughput metabolomics data. *Computational and structural biotechnology journal*, 4(5):e201301009.

Bartle, K.D. & Myers, P. 2002. History of gas chromatography. *TrAC Trends in Analytical Chemistry*, 21(9-10):547-557.

Beale, D.J., Pinu, F.R., Kouremenos, K.A., Poojary, M.M., Narayana, V.K., Boughton, B.A., ... Dias, D.A. 2018. Review of recent developments in GC–MS approaches to metabolomics-based research. *Metabolomics*, 14(11):1-31.

Bhumireddy, S.R., Rocchetti, G., Pallerla, P., Lucini, L. & Sripadi, P. 2021. A combined targeted/untargeted screening based on GC/MS to detect low-molecular-weight compounds in different milk samples of different species and as affected by processing. *International Dairy Journal*, 118:105045.

Bi, H., Guo, Z., Jia, X., Liu, H., Ma, L. & Xue, L. 2020. The key points in the pre-analytical procedures of blood and urine samples in metabolomics studies. *Metabolomics*, 16:1-15.

Brown, C.E. 1998. Coefficient of variation. In. *Applied multivariate statistics in geohydrology and related sciences*: Springer. pp. 155-157.

Burgess, K., Rankin, N. & Weidt, S. 2014. Chapter 10 - Metabolomics. In: Padmanabhan, S., ed. *Handbook of Pharmacogenomics and Stratified Medicine*. San Diego: Academic Press. pp. 181-205.

Carrasco-Pancorbo, A., Nevedomskaya, E., Arthen-Engeland, T., Zey, T., Zurek, G., Baessmann, C., ... Mayboroda, O.A. 2009. Gas chromatography/atmospheric pressure chemical ionization-time of flight mass spectrometry: analytical validation and applicability to metabolic profiling. *Analytical Chemistry*, 81(24):10071-10079.

Casas, M.E. & Matamoros, V. 2021. Analytical challenges and solutions for performing metabolomic analysis of root exudates. *Trends in Environmental Analytical Chemistry*, 31:e00130.

Chong, J., Wishart, D.S. & Xia, J. 2019. Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Current protocols in bioinformatics*, 68(1):e86.

Christou, C., Gika, H.G., Raikos, N. & Theodoridis, G. 2014. GC-MS analysis of organic acids in human urine in clinical settings: a study of derivatization and other analytical parameters. *Journal of Chromatography B*, 964:195-201.

Crews, B., Wikoff, W.R., Patti, G.J., Woo, H.-K., Kalisiak, E., Heideker, J. & Siuzdak, G. 2009. Variability analysis of human plasma and cerebral spinal fluid reveals statistical significance of changes in mass spectrometry-based metabolomics data. *Analytical chemistry*, 81(20):8538-8544.

Dayalan, S., Xia, J., Spicer, R.A., Salek, R. & Roessner, U. 2019. Metabolome Analysis. In: Ranganathan, S., Gribskov, M., Nakai, K. & Schönbach, C., eds. *Encyclopedia of Bioinformatics and Computational Biology*. Oxford: Academic Press. pp. 396-409.

Denery, J.R., Nunes, A.A. & Dickerson, T.J. 2011. Characterization of differences between blood sample matrices in untargeted metabolomics. *Analytical Chemistry*, 83(3):1040-1047.

Dettmer, K., Aronov, P.A. & Hammock, B.D. 2007. Mass spectrometry-based metabolomics. *Mass spectrometry reviews*, 26(1):51-78.

Doğan, H.O., Şenol, O., Bolat, S., Yıldız, Ş.N., Büyüktuna, S.A., Sarıismailoğlu, R., ... Hekim, S.N. 2021. Understanding the pathophysiological changes via untargeted metabolomics in COVID-19 patients. *Journal of medical virology*, 93(4):2340-2349.

El-Maghrabey, M.H., Kishikawa, N. & Kuroda, N. 2020. Utility of isotope-coded derivatization in gas chromatographic-mass spectrometric analyses with special emphasis on metabolomics. *Medical Mass Spectrometry*, 4(1):24-35.

Engel, B., Suralik, P. & Marchetti-Deschmann, M. 2020. Critical considerations for trimethylsilyl derivatives of 24 primary metabolites measured by gas chromatography–tandem mass spectrometry. *Separation Science Plus*, 3(9):407-418.

Erarpat, S., Bodur, S., Öztürk Er, E. & Bakırdere, S. 2020. Combination of ultrasound-assisted ethyl chloroformate derivatization and switchable solvent liquid-phase microextraction for the sensitive determination of l-methionine in human plasma by GC–MS. *Journal of separation science*, 43(6):1100-1106.

Feizi, N., Hashemi-Nasab, F.S., Golpelichi, F., Saburouh, N. & Parastar, H. 2021. Recent trends in application of chemometric methods for GC-MS and GC×GC-MS-based metabolomic studies. *TrAC Trends in Analytical Chemistry*, 138:116239.

Ferreira, A.M.C., Laespada, M.E.F., Pavón, J.L.P. & Cordero, B.M. 2013. In situ aqueous derivatization as sample preparation technique for gas chromatographic determinations. *Journal of Chromatography A*, 1296:70-83.

Fiehn, O. 2000. Metabolite profiling for plant functional genomics. *Nature biotechnology*, 18(11):1157-1161.

Fiehn, O. 2016. Metabolomics by gas chromatography–mass spectrometry: Combined targeted and untargeted profiling. *Current protocols in molecular biology*, 114(1):30.34. 31-30.34. 32.

Fritsche-Guenther, R., Gloaguen, Y., Bauer, A., Opialla, T., Kempa, S., Fleming, C.A., ... Kirwan, J.A. 2021. Optimized workflow for on-line derivatization for targeted metabolomics approach by gas chromatography-mass spectrometry. *Metabolites*, 11(12):888.

Gao, X., Pujos, E., Martin, J.F., Galan, P., Juste, C., Jia, W. & Sebedio, J.L. 2009. Metabolite analysis of human fecal water by gas chromatography/mass spectrometry with ethyl chloroformate derivatization. *Analytical biochemistry*, 393(2):163-175.

Gao, X., Pujos-Guillot, E. & Sébédio, J.-L. 2010. Development of a quantitative metabolomic approach to study clinical human fecal water metabolome based on trimethylsilylation derivatization and GC/MS analysis. *Analytical chemistry*, 82(15):6447-6456.

Glicksberg, L. & Kerrigan, S. 2020. Derivatization. In: Levine, B.S. & Kerrigan, S., eds. *Principles of Forensic Toxicology*. Cham: Springer International Publishing. pp. 163-175.

Grace, S.C. & Hudson, D.A. 2017. *Processing and visualization of metabolomics data using R*. (Metabolomics: fundamentals and applications).

Griffin, J.L. 2006. The Cinderella story of metabolic profiling: does metabolomics get to go to the functional genomics ball? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1465):147-161.

Gullberg, J., Jonsson, P., Nordström, A., Sjöström, M. & Moritz, T. 2004. Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of Arabidopsis thaliana samples in metabolomic studies with gas chromatography/mass spectrometry. *Analytical biochemistry*, 331(2):283-295.

Gumbi, B.P., Moodley, B., Birungi, G. & Ndungu, P.G. 2019. Target, Suspect and Non-Target Screening of Silylated Derivatives of Polar Compounds Based on Single Ion Monitoring GC-MS. *International Journal of Environmental Research and Public Health*, 16(20):1-15.

Halket, J.M., Przyborowska, A., Stein, S.E., Mallard, W.G., Down, S. & Chalmers, R.A. 1999. Deconvolution gas chromatography/mass spectrometry of urinary organic acids–potential for pattern recognition and automated identification of metabolic disorders. *Rapid communications in mass spectrometry*, 13(4):279-284.

Halket, J.M. & Zaikin, V.G. 2003. Derivatization in mass spectrometry—1. Silylation. *European Journal of Mass Spectrometry*, 9(1):1-21.

Han, E., Park, Y., Kim, E., In, S., Yang, W., Lee, S., ... myong Song, J. 2011. Simultaneous analysis of Δ9-tetrahydrocannabinol and 11-nor-9-carboxy-tetrahydrocannabinol in hair without different sample preparation and derivatization by gas chromatography–tandem mass spectrometry. *Journal of pharmaceutical and biomedical analysis*, 55(5):1096-1103.

He, Y., Zhang, Z., Ma, P., Ji, H. & Lu, H. 2018. GC-MS profiling of leukemia cells: an optimized preparation protocol for the intracellular metabolome. *Analytical Methods*, 10(10):1266-1274.

Hendriks, M.M., van Eeuwijk, F.A., Jellema, R.H., Westerhuis, J.A., Reijmers, T.H., Hoefsloot, H.C. & Smilde, A.K. 2011. Data-processing strategies for metabolomics studies. *TrAC Trends in Analytical Chemistry*, 30(10):1685-1698.

Hibbert, D.B. 2007. Systematic errors in analytical measurement results. *Journal of Chromatography A*, 1158(1-2):25-32.

Hites, R.A. 1997. Gas chromatography mass spectrometry. In. *Handbook of instrumental techniques for analytical chemistry*. pp. 609-626.

Huang, B., Pan, X., Liu, J., Fang, K., Wang, Y. & Gao, J. 2010. New discoveries of heating effect on trimethylsilyl derivatization for simultaneous determination of steroid endocrine disrupting chemicals by GC–MS. *Chromatography*, 71(1):149-153.

Hušek, P., Švagera, Z., Hanzlíková, D., Řimnáčová, L., Zahradníčková, H., Opekarová, I. & Šimek, P. 2016. Profiling of urinary amino-carboxylic metabolites by in-situ heptafluorobutyl chloroformate mediated sample preparation and gas chromatography-mass spectrometry. *Journal of Chromatography A*, 1443:32-211.

Johnson, C.H., Ivanisevic, J. & Siuzdak, G. 2016. Metabolomics: beyond biomarkers and towards mechanisms. *Nature reviews Molecular cell biology*, 17(7):451-459.

Jolliffe, I. 2005. Principal component analysis. In. *Encyclopedia of statistics in behavioral science*. 3. pp. 1580-1584.

Karaman, I. 2017. Preprocessing and pretreatment of metabolomics data for statistical analysis. In. *Metabolomics: From Fundamentals to Clinical Applications*. 956. pp. 145-161.

Kaspar, H., Dettmer, K., Gronwald, W. & Oefner, P.J. 2008. Automated GC–MS analysis of free amino acids in biological fluids. *Journal of Chromatography B*, 870(2):222-232.

Khodadadi, M. & Pourfarzam, M. 2020. A review of strategies for untargeted urinary metabolomic analysis using gas chromatography–mass spectrometry. *Metabolomics*, 16(6):1-14.

Killiny, N., Hijaz, F., El-Shesheny, I., Alfaress, S., Jones, S.E. & Rogers, M.E. 2017. Metabolomic analyses of the haemolymph of the Asian citrus psyllid *Diaphorina citri*, the vector of huanglongbing. *Physiological Entomology*, 42(2):134-145.

Kim, H., Kwon, J., Choi, S.Y. & Ahn, Y.G. 2019. Method development for the quantitative determination of short chain fatty acids in microbial samples by solid phase extraction and gas chromatography with flame ionization detection. *Journal of Analytical Science and Technology*, 10(1):1-6.

Kiseleva, O., Kurbatov, I., Ilgisonis, E. & Poverennaya, E. 2021. Defining blood plasma and serum metabolome by GC-MS. *Metabolites*, 12(1):15.

Knapp, D.R. 1979. *Handbook of analytical derivatization reactions*. 1. John Wiley & Sons.

Koek, M.M., Muilwijk, B., van der Werf, M.J. & Hankemeier, T. 2006. Microbial metabolomics with gas chromatography/mass spectrometry. *Analytical chemistry*, 78(4):1272-1281.

Kouremenos, K.A., Harynuk, J.J., Winniford, W.L., Morrison, P.D. & Marriott, P.J. 2010. One-pot microwave derivatization of target compounds relevant to metabolomics with comprehensive two-dimensional gas chromatography. *Journal of Chromatography B*, 878(21):1761-1770.

Kranenburg, R.F., Verduin, J., Stuyver, L.I., de Ridder, R., van Beek, A., Colmsee, E. & van Asten, A.C. 2020. Benefits of derivatization in GC–MS-based identification of new psychoactive substances. *Forensic Chemistry*, 20:100273.

Krzywinski, M. & Altman, N. 2014. Visualizing samples with box plots. *Nature methods*, 11(2):119-120.

Kvitvang, H.F., Andreassen, T., Adam, T., Villas-Bôas, S.G. & Bruheim, P. 2011. Highly sensitive GC/MS/MS method for quantitation of amino and nonamino organic acids. *Analytical chemistry*, 83(7):2705-2711.

Kyle, P. 2017. Toxicology: GCMS. In. *Mass spectrometry for the clinical laboratory*: Elsevier. pp. 131-163.

Lai, Z. & Fiehn, O. 2016. Mass spectral fragmentation of trimethylsilylated small molecules. *Mass Spectrometry Reviews*, 37(3):245-257.

Lamichhane, S., Sen, P., Dickens, A.M., Hyötyläinen, T. & Orešič, M. 2018. An overview of metabolomics data analysis: current tools and future perspectives. *Comprehensive analytical chemistry*, 82:387-413.

León, Z., García-Cañaveras, J.C., Donato, M.T. & Lahoz, A. 2013. Mammalian cell metabolomics: Experimental design and sample preparation. *Electrophoresis*, 34(19):2762-2775.

Li, B., Tang, J., Yang, Q., Cui, X., Li, S., Chen, S., ... Zhu, F. 2016. Performance evaluation and online realization of data-driven normalization methods used in LC/MS based untargeted metabolomics analysis. *Scientific reports*, 6(1):1-13.

Li, Z., Lu, Y., Guo, Y., Cao, H., Wang, Q. & Shui, W. 2018. Comprehensive evaluation of untargeted metabolomics data processing software in feature detection, quantification and discriminating marker selection. *Analytica chimica acta*, 1029:50-57.

Liebenberg, C., Luies, L. & Williams, A.A. 2021. Metabolomics as a tool to investigate HIV/TB co-infection. *Frontiers in molecular biosciences*, 8:692823.

Lin, D.-L., Wang, S.-M., Wu, C.-H., Chen, B.-G. & Liu, R.H. 2008. Chemical derivatization for the analysis of drugs by GC-MS: A conceptual review. *Journal of Food and Drug Analysis*, 16(1):1-10.

Lin, D.L., Wang, S.M., Wu, C.H., Chen, B.G. & Liu, R.H. 2016. Chemical derivatization for forensic drug analysis by GC-and LC-MS. *Forensic Scientific Review*, 28(1):17-35.

Lisec, J., Schauer, N., Kopka, J., Willmitzer, L. & Fernie, A.R. 2006. Gas chromatography mass spectrometry–based metabolite profiling in plants. *Nature protocols*, 1(1):387-396.

Little, J.L. 1999. Artifacts in trimethylsilyl derivatization reactions and ways to avoid them. *Journal of Chromatography A*, 844(1-2):1-22.

Liu, Y., Mei, B., Chen, D. & Cai, L. 2021. GC-MS metabolomics identifies novel biomarkers to distinguish tuberculosis pleural effusion from malignant pleural effusion. *Journal of Clinical Laboratory Analysis*, 35(4):e23706.

Lkhagva, A., Shen, C.C., Leung, Y.S. & Tai, H.C. 2020. Comparative study of five different amine-derivatization methods for metabolite analyses by liquid chromatography-tandem mass spectrometry. *Journal of Chromatography A*, 1610:460536.

Loots, D.T., Meissner-Roloff, R.J., Newton-Foot, M. & van Pittius, N.C.G. 2013. A metabolomics approach exploring the function of the ESX-3 type VII secretion system of *M. smegmatis*. *Metabolomics*, 9(3):631-641.

López-Bascón, M., Calderón-Santiago, M., Argüello, H., Morera, L., Garrido, J. & Priego-Capote, F. 2019. Comprehensive analysis of pig feces metabolome by chromatographic techniques coupled to mass spectrometry in high resolution mode: Influence of sample preparation on the identification coverage. *Talanta*, 199:303-309.

Lord, H. & Pfannkoch, E. 2012. Sample preparation automation for GC injection. In. *Comprehensive Sampling and Sample Preparation*. 2. pp. 597-612.

Loyo, R.M., Zarate, E., Barbosa, C.S. & Simoes-Barbosa, A. 2021. Gas chromatography-mass spectrometry (GC/MS) reveals urine metabolites associated to light and heavy infections by *Schistosoma mansoni* in mice. *Parasitology international*, 80:102239.

Luies, L. & Loots, D.T. 2016. Tuberculosis metabolomics reveals adaptations of man and microbe in order to outcompete and survive. *Metabolomics*, 12(3):40.

Manchester, M. & Anand, A. 2017. Metabolomics: strategies to define the role of metabolism in virus infection and pathogenesis. *Advances in virus research*, 98:57-81.

Marcos, J. & Pozo, O.J. 2015. Derivatization of steroids in biological samples for GC–MS and LC–MS analyses. *Bioanalysis*, 7(19):2515-2536.

Maštovská, K. & Lehotay, S.J. 2003. Practical approaches to fast gas chromatography–mass spectrometry. *Journal of Chromatography A*, 1000(1-2):153-180.

Miggiels, P., Wouters, B., van Westen, G.J., Dubbelman, A.-C. & Hankemeier, T. 2019. Novel technologies for metabolomics: More for less. *TrAC Trends in Analytical Chemistry*, 120:115323.

Miyagawa, H. & Bamba, T. 2019. Comparison of sequential derivatization with concurrent methods for GC/MS-based metabolomics. *Journal of Bioscience and Bioengineering*, 127(2):160-168.

Mojsak, P., Rey-Stolle, F., Parfieniuk, E., Kretowski, A. & Ciborowski, M. 2020. The role of gut microbiota (GM) and GM-related metabolites in diabetes and obesity. A review of analytical methods used to

measure GM-related metabolites in fecal samples with a focus on metabolites' derivatization step. *Journal of Pharmaceutical and Biomedical Analysis*, 191:113617.

Moldoveanu, S.C. & David, V. 2018. Derivatization methods in GC and GC/MS. In. *Gas Chromatography-Derivatization, Sample Preparation, Application*: IntechOpen London, UK. pp. 1-33.

Moros, G., Chatziioannou, A.C., Gika, H.G., Raikos, N. & Theodoridis, G. 2017. Investigation of the derivatization conditions for GC–MS metabolomics of biological samples. *Bioanalysis*, 9(1):53-65.

Munir, M.A. & Badri, K.H. 2020. The importance of derivatizing reagent in chromatography applications for biogenic amine detection in food and beverages. *Journal of analytical methods in chemistry*, 2020,

Muthubharathi, B.C., Gowripriya, T. & Balamurugan, K. 2021. Metabolomics: small molecules that matter more. *Molecular Omics*, 17(2):210-229.

Nguyen, T.V., Alfaro, A.C., Venter, L., Ericson, J.A., Ragg, N.L., McCowan, T. & Mundy, C. 2023. Metabolomics approach reveals size-specific variations of blackfoot abalone (*Haliotis iris*) in Chatham Islands, New Zealand. *Fisheries Research*, 262:106645.

Noctor, G., Bergot, G., Mauve, C., Thominet, D., Lelarge-Trouverie, C. & Prioul, J.-L. 2007. A comparative study of amino acid measurement in leaf extracts by gas chromatography-time of flight-mass spectrometry and high performance liquid chromatography with fluorescence detection. *Metabolomics*, 3(2):161-174.

Nuzzo, R.L. 2016. The box plots alternative for visualizing quantitative data. *Journal of Physical Medicine and Rehabilitation*, 8(3):268-272.

Oliver, S.G., Winson, M.K., Kell, D.B. & Baganz, F. 1998. Systematic functional analysis of the yeast genome. *Trends in biotechnology*, 16(9):373-378.

Orata, F. 2012. Derivatization reactions and reagents for gas chromatography analysis. *Advanced gas chromatography-Progress in agricultural, biomedical and industrial applications*, 91:83-104.

Paiva, A.C., Oliveira, A.M.d., Crucello, J., Facanali, R. & Hantao, L.W. 2021. Practical Considerations in Method Development for Gas Chromatography-Based Metabolomic Profiling. In. *Separation Techniques Applied to Omics Sciences*: Springer. pp. 139-157.

Pang, Z., Zhou, G., Ewald, J., Chang, L., Hacariz, O., Basu, N. & Xia, J. 2022. Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nature protocols*, 17(8):1735-1761.

Parkinson, D.R. 2012. Analytical Derivatization Techniques. In. 2. pp. 559-595.

Parlapally, S., Cherukupalli, N., Bhumireddy, S.R., Sripadi, P., Anisetti, R., Giri, C.C., ... Reddy Vudem, D. 2016. Chemical profiling and anti-psoriatic activity of methanolic extract of Andrographis nallamalayana JL Ellis. *Natural Product Research*, 30(11):1256-1261.

Parsons, H.M., Ekman, D.R., Collette, T.W. & Viant, M.R. 2009. Spectral relative standard deviation: a practical benchmark in metabolomics. *Analyst*, 134(3):478-485.

Parvatam, R., Singh, R. & Sharma, R. 2023. Comparison of different derivatising reagents in identification of milk metabolites using GC–MS. *International Dairy Journal*, 138:105535.

Pasikanti, K.K., Ho, P. & Chan, E. 2008. Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. *Journal of Chromatography B*, 871(2):202-211.

Pechlivanis, A., Chatziioannou, A.C., Veskoukis, A.S., Kouretas, D., Mougios, V. & Theodoridis, G.A. 2014. GC-MS analysis of blood for the metabonomic investigation of the effects of physical exercise and allopurinol administration on rats. *Journal of Chromatography B Analytical Technology Biomedical Life Sciences*, 966:127-131.

Peisl, L., Schymanski, E.L. & Wilmes, P. 2018. Dark matter in host-microbiome metabolomics: Tackling the unknowns–A review. *Analytica Chimica Acta*, 1037:13-27.

Pélabon, C., Hilde, C.H., Einum, S. & Gamelon, M. 2020. On the use of the coefficient of variation to quantify and compare trait variation. *Evolution letters*, 4(3):180-188.

Pereira Braga, C. & Adamec, J. 2019. Metabolome Analysis. In: Ranganathan, S., Gribskov, M., Nakai, K. & Schönbach, C., eds. *Encyclopedia of Bioinformatics and Computational Biology*. Oxford: Academic Press. pp. 463-475.

Pietrogrande, M.C., Bacco, D. & Mercuriali, M. 2010. GC–MS analysis of low-molecular-weight dicarboxylic acids in atmospheric aerosol: comparison between silylation and esterification derivatization procedures. *Analytical and bioanalytical chemistry*, 396:877-885.

Pino, L.K., Searle, B.C., Huang, E.L., Noble, W.S., Hoofnagle, A.N. & MacCoss, M.J. 2018. Calibration using a single-point external reference material harmonizes quantitative mass spectrometry proteomics data between platforms and laboratories. *Analytical chemistry*, 90(21):13112-13117.

Pinu, F.R., Villas-Boas, S.G. & Aggio, R. 2017. Analysis of intracellular metabolites from microorganisms: quenching and extraction protocols. *Metabolites*, 7(4):53.

Pinu, F.R., Villas-Boas, S.G. & Martin, D. 2019. Pre-fermentative supplementation of fatty acids alters the metabolic activity of wine yeasts. *Food Research International*, 121:835-844.

Poole, C.F. 2013. Derivatization reactions for use with the electron-capture detector. *Journal of Chromatography A*, 1296:15-24.

Putri, S.P., Ikram, M.M.M., Sato, A., Dahlan, H.A., Rahmawati, D., Ohto, Y. & Fukusaki, E. 2022. Application of gas chromatography-mass spectrometry-based metabolomics in food science and technology. *Journal of Bioscience and Bioengineering*, 133(5):425-435.

Quéro, A., Jousse, C., Lequart-Pillon, M., Gontier, E., Guillot, X., Courtois, B., ... Pau-Roblot, C. 2014. Improved stability of TMS derivatives for the robust quantification of plant polar metabolites by gas chromatography–mass spectrometry. *Journal of Chromatography B*, 970:36-43.

Reinecke, C.J., Koekemoer, G., Van der Westhuizen, F.H., Louw, R., Lindeque, J.Z., Mienie, L.J. & Smuts, I. 2012. Metabolomics of urinary organic acids in respiratory chain deficiencies in children. *Metabolomics*, 8:264-283.

Ren, S., Hinzman, A.A., Kang, E.L., Szczesniak, R.D. & Lu, L.J. 2015. Computational and statistical analysis of metabolomics data. *Metabolomics*, 11:1492-1513.

Řezanka, T., Pádrová, K. & Sigler, K. 2016. Derivatization in gas chromatography of lipids. *Encyclopedia of Lipidomics*:1-9.

Rimayi, C., Odusanya, D., Mtunzi, F. & Tsoka, S. 2015. Alternative calibration techniques for counteracting the matrix effects in GC–MS-SPE pesticide residue analysis–A statistical approach. *Chemosphere*, 118:35-43.

Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N. & Willmitzer, L. 2000. Simultaneous analysis of metabolites in potato tuber by gas chromatography–mass spectrometry. *The plant journal*, 23(1):131-142.

Roessner, U. & Bowne, J. 2009. What is metabolomics all about? *BioTechniques*, 46(5):363-365.

Rohloff, J. 2015. Analysis of phenolic and cyclic compounds in plants using derivatization techniques in combination with GC-MS-based metabolite profiling. *Molecules*, 20(2):3431-3462.

Rompa, M., Kremer, E. & Zygmunt, B. 2003. Derivatisation in gas chromatographic determination of acidic herbicides in aqueous environmental samples. *Analytical and bioanalytical chemistry*, 377(4):590-599.

Saccenti, E., Hoefsloot, H.C., Smilde, A.K., Westerhuis, J.A. & Hendriks, M.M. 2014. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics*, 10(3):361-374.

Sajid, M. & Płotka-Wasylka, J. 2018. “Green” nature of the process of derivatization in analytical sample preparation. *TrAC Trends in Analytical Chemistry*, 102:16-31.

Santos, F. & Galceran, M. 2003. Modern developments in gas chromatography–mass spectrometry-based environmental analysis. *Journal of Chromatography A*, 1000(1-2):125-151.

Schoeman, J.C. & du Preez, I. 2012. A comparison of four sputum pre-extraction preparation methods for identifying and characterising *Mycobacterium tuberculosis* using GCxGC-TOFMS metabolomics. *Journal of microbiological methods*, 91(2):301-311.

Schrimpe-Rutledge, A.C., Codreanu, S.G., Sherrod, S.D. & McLean, J.A. 2016. Untargeted metabolomics strategies — challenges and emerging directions. *Journal of the American Society for Mass Spectrometry*, 27(12):1897-1905.

Schummer, C., Delhomme, O., Appenzeller, B.M., Wennig, R. & Millet, M. 2009. Comparison of MTBSTFA and BSTFA in derivatization reactions of polar compounds prior to GC/MS analysis. *Talanta*, 77(4):1473-1482.

Sindelar, M. & Patti, G.J. 2020. Chemical discovery in the era of metabolomics. *Journal of the American Chemical Society*, 142(20):9097-9105.

Smart, K.F., Aggio, R.B., Van Houtte, J.R. & Villas-Bôas, S.G. 2010. Analytical platform for metabolome analysis of microbial cells using methyl chloroformate derivatization followed by gas chromatography–mass spectrometry. *Nature protocols*, 5(10):1709-1729.

Smilde, A., Westerhuis, J., Hoefsloot, H., Bijlsma, S., Rubingh, C., Vis, D., ... Van Der Greef, J. 2010. Dynamic metabolomic data analysis: a tutorial review. *Metabolomics*, 6:3-17.

Smuts, I., Van der Westhuizen, F.H., Louw, R., Mienie, L.J., Engelke, U.F., Wevers, R.A., ... Reinecke, C.J. 2013. Disclosure of a putative biosignature for respiratory chain disorders through a metabolomics approach. *Metabolomics*, 9(2):379-391.

Sobolevsky, T.G., Revelsky, A.I., Miller, B., Oriedo, V., Chernetsova, E.S. & Revelsky, I.A. 2003. Comparison of silylation and esterification/acylation procedures in GC-MS analysis of amino acids. *Journal of Separation Science*, 26(17):1474-1478.

Sogin, E.M., Puskás, E., Dubilier, N. & Liebeke, M. 2019. Marine metabolomics: a method for nontargeted measurement of metabolites in seawater by gas chromatography–mass spectrometry. *Msystems*, 4(6):e00638-00619.

Sotelo-Orozco, J., Chen, S.-Y., Hertz-Picciotto, I. & Slupsky, C.M. 2021. A comparison of serum and plasma blood collection tubes for the integration of epidemiological and metabolomics data. *Frontiers in Molecular Biosciences*, 8:682134.

Špánik, I. & Machyňáková, A. 2018. Recent applications of gas chromatography with high-resolution mass spectrometry. *Journal of separation science*, 41(1):163-179.

Stein, S.E. & Scott, D.R. 1994. Optimization and testing of mass spectral library search algorithms for compound identification. *Journal of the American Society for Mass Spectrometry*, 5(9):859-866.

Su, Y., Xia, S., Wang, R. & Xiao, L. 2017. Phytohormonal quantification based on biological principles. *Hormone metabolism and signaling in plants*, 13:431-470.

Swenson, T.L., Jenkins, S., Bowen, B.P. & Northen, T.R. 2015. Untargeted soil metabolomics methods for analysis of extractable organic matter. *Soil Biology and Biochemistry*, 80:189-198.

Swenson, T.L. & Northen, T.R. 2019. Untargeted soil metabolomics using liquid chromatography-mass spectrometry and gas chromatography-mass spectrometry. In. *Microbial Metabolomics*. 1859: Springer. pp. 97-109.

t'Kindt, R., Morreel, K., Deforce, D., Boerjan, W. & Van Bocxlaer, J. 2009. Joint GC–MS and LC–MS platforms for comprehensive plant metabolomics: Repeatability and sample pre-treatment. *Journal of Chromatography B*, 877(29):3572-3580.

Tan, S.Z., Begley, P., Mullard, G., Hollywood, K.A. & Bishop, P.N. 2016. Introduction to metabolomics and its applications in ophthalmology. *Eye*, 30(6):773-783.

Tang, Y.-S., Zhang, M.-J., Zhao, J.-H. & Liu, L.-Y. 2023. Optimization of a quantitative protocol for the intermediate metabolites of the glycolysis pathway in human serum using gas chromatography–mass spectrometry. *New Journal of Chemistry*, 47(19):9364-9376.

Tumanov, S., Zubenko, Y., Obolonkin, V., Greenwood, D.R., Shmanai, V. & Villas-Boas, S.G. 2016. Calibration curve-free GC–MS method for quantitation of amino and non-amino organic acids in biological samples. *Metabolomics*, 12(4):1-13.

Valdez, C.A. & Leif, R.N. 2021. Analysis of organophosphorus-based nerve agent degradation products by gas chromatography-mass spectrometry (GC-MS): Current derivatization reactions in the analytical chemist's toolbox. *Molecules*, 26(15):4631.

van den Berg, R.A., Hoefsloot, H.C., Westerhuis, J.A., Smilde, A.K. & van der Werf, M.J. 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC genomics*, 7(1):1-15.

van Ravenzwaay, B., Cunha, G.C.-P., Leibold, E., Looser, R., Mellert, W., Prokoudine, A., ... Wiemer, J. 2007. The use of metabolomics for the discovery of new biomarkers of effect. *Toxicology letters*, 172(1-2):21-28.

Venter, L., van Rensburg, P.J., Loots, D.T., Vosloo, A. & Lindeque, J.Z. 2016. Untargeted metabolite profiling of abalone using gas chromatography mass spectrometry. *Food Analytical Methods*, 9:1254-1261.

Verma, P., Bajaj, A., Tripathi, R., Shukla, S.K. & Nagpal, S. 2021. Development and validation of the method for the detection of glimepiride via derivatization employing N-methyl-N-(trimethylsilyl) trifluoroacetamide using gas chromatography-mass spectrometry. *Egyptian Journal of Forensic Sciences*, 11(1):1-9.

Vignoli, A., Tenori, L., Morsiani, C., Turano, P., Capri, M. & Luchinat, C. 2022. Serum or plasma (and which plasma), that is the question. *Journal of Proteome Research*, 21(4):1061-1072.

Vilbaste, M., Tammekivi, E. & Leito, I. 2020. Uncertainty contribution of derivatization in gas chromatography/mass spectrometric analysis. *Rapid Communications in Mass Spectrometry*, 34(16):e8704.

Villaret-Cazadamont, J., Poupin, N., Tournadre, A., Batut, A., Gales, L., Zalko, D., ... Bertrand-Michel, J. 2020. An optimized dual extraction method for the simultaneous and accurate analysis of polar metabolites and lipids carried out on single biological samples. *Metabolites*, 10(9):338.

Villas-Bôas, S.G., Delicado, D.G., Akesson, M. & Nielsen, J. 2003. Simultaneous analysis of amino and nonamino organic acids as methyl chloroformate derivatives using gas chromatography-mass spectrometry. *Analytical biochemistry*, 322(1):134-138.

Villas-Bôas, S.G., Moxley, J.F., Åkesson, M., Stephanopoulos, G. & Nielsen, J. 2005. High-throughput metabolic state analysis: the missing link in integrated functional genomics of yeasts. *Biochemical Journal*, 388(2):669-677.

Villas-Bôas, S.G., Noel, S., Lane, G.A., Attwood, G. & Cookson, A. 2006. Extracellular metabolomics: a metabolic footprinting approach to assess fiber degradation in complex media. *Analytical biochemistry*, 349(2):297-305.

Villas-Bôas, S.G., Koulman, A. & Lane, G.A. 2007. Analytical methods from the perspective of method standardization. *Metabolomics*:11-52.

Villas-Bôas, S.G., Smart, K.F., Sivakumaran, S. & Lane, G.A. 2011. Alkylation or silylation for analysis of amino and non-amino organic acids by GC-MS? *Metabolites*, 1(1):3-20.

Villas-Bôas, S.G., Mas, S., Åkesson, M., Smedsgaard, J. & Nielsen, J. 2005. Mass spectrometry in metabolome analysis. *Mass Spectrometry Reviews*, 24(5):613-646.

Wang, H., Feng, Y., Gui, S., Zhang, Y. & Lu, F. 2017. A sample pretreatment method to suit the metabolomic analysis of *Bacillus licheniformis* based on GC-MS. *Analytical Methods*, 9(15):2299-2304.

Wang, L., Naser, F.J., Spalding, J.L. & Patti, G.J. 2019. A protocol to compare methods for untargeted metabolomics. In. *Metabolic Signaling*: Springer. pp. 1-15.

Wang, T.J.L., Martin G;, Vasan, R., S;, Cheng, S., Rhee, E.P., McCabe, E., Lewis, G.D., ... Fernandez, C. 2011. Metabolite profiles and the risk of developing diabetes. *Nature medicine*, 17(4):448-453.

Willers, C., van Rensburg, P. & Claassens, S. 2016. Can a metabolomics-based approach be used as alternative to analyse fatty acid methyl esters from soil microbial communities? *Soil Biology and Biochemistry*, 103:417-428.

Williams, C., Mbuyane, L.L., Bauer, F.F., Mokwena, L., Divol, B. & Buica, A. 2021. A gas chromatography-mass spectrometry method for the determination of fatty acids and sterols in yeast and grape juice. *Applied Sciences*, 11(11):5152.

Williamson, D.F., Parker, R.A. & Kendrick, J.S. 1989. The box plot: a simple visual method to interpret data. *Annals of internal medicine*, 110(11):916-921.

Wilson, E.B.K. & Walker, J. 2005. Chromatographic techniques. In. *Biochemistry and Molecular biology*. 7. pp. 433-476.

Winder, C.L., Cornmell, R., Schuler, S., Jarvis, R.M., Stephens, G.M. & Goodacre, R. 2011. Metabolic fingerprinting as a tool to monitor whole-cell biotransformations. *Analytical and bioanalytical chemistry*, 399:387-401.

Wishart, D.S. 2019. Metabolomics for investigating physiological and pathophysiological processes. *Physiological Reviews*, 99(4):1819-1875.

Worsfold, P., Townshend, A., Poole, C.F. & Miró, M. 2019. *Encyclopedia of analytical science*. 10. 3. Elsevier.

Woźniak, M.K., Banaszkiewicz, L., Wiergowski, M., Tomczak, E., Kata, M., Szpiech, B., ... Biziuk, M. 2020. Development and validation of a GC–MS/MS method for the determination of 11 amphetamines and 34 synthetic cathinones in whole blood. *Forensic Toxicology*, 38:42-58.

Xia, J., Psychogios, N., Young, N. & Wishart, D.S. 2009. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic acids research*, 37:652-660.

Xu, F., Zou, L. & Ong, C.N. 2009. Multiorigination of chromatographic peaks in derivatized GC/MS metabolomics: a confounder that influences metabolic pathway interpretation. *Journal of proteome research*, 8(12):5657-5665.

Xu, L. 2009. Chemical reactions in liquid-phase microextraction. *Journal of Chromatography A*, 1216(4):701-707.

Yang, Q., Shi, B.H., Tian, G.L., Niu, Q.Q., Tang, J., Linghu, D.D., ... Xu, L. 2021. GC–MS urinary metabolomics analysis of inherited metabolic diseases and stable metabolic biomarker screening by a comprehensive chemometric method. *Microchemical Journal*, 168:106350.

Yegles, M., Labarthe, A., Auwärter, V., Hartwig, S., Vater, H., Wennig, R. & Pragst, F. 2004. Comparison of ethyl glucuronide and fatty acid ethyl ester concentrations in hair of alcoholics, social drinkers and teetotallers. *Forensic Science International*, 145(2-3):167-173.

Yin, P., Lehmann, R. & Xu, G. 2015. Effects of pre-analytical processes on blood samples used in metabolomics studies. *Analytical and bioanalytical chemistry*, 407:4879-4892.

Yin, S., Guo, P., Hai, D., Xu, L., Shu, J., Zhang, W., ... Liu, Y. 2017. Optimization of GC/TOF MS analysis conditions for assessing host-gut microbiota metabolic interactions: Chinese rhubarb alters fecal aromatic amino acids and phenol metabolism. *Analytica Chimica Acta*, 995:21-33.

Zarate, E., Boyle, V., Rupprecht, U., Green, S., Villas-Boas, S.G., Baker, P. & Pinu, F.R. 2016. Fully automated trimethylsilyl (TMS) derivatisation protocol for metabolite profiling by GC-MS. *Metabolites*, 7(1):1.

Zeki Ö, C., Eylem, C.C., Reçber, T., Kır, S. & Nemutlu, E. 2020. Integration of GC-MS and LC-MS for untargeted metabolomics profiling. *Journal of Pharmaceutical and Biomedical Analysis*, 190:113509. 10.1016/j.jpba.2020.113509

Zhang, A. 2012a. Modern analytical techniques in metabolomics analysis. *Analyst*, 137(2):293-300.

Zhang, A. 2012b. Serum metabolomics as a novel diagnostic approach for disease: a systematic review. *Analytical and bioanalytical chemistry*, 404:1239-1245.

Zhang, A., Sun, H., Wang, P., Han, Y. & Wang, X. 2012. Future perspectives of personalized medicine in traditional Chinese medicine: a systems biology approach. *Complementary therapies in medicine*, 20(1-2):93-99.

Zhang, H., Tan, X., Kang, K., Wang, W., Lian, K. & Kang, W. 2019. Simultaneous determination of lactic acid and pyruvic acid in tissue and cell culture media by gas chromatography after in situ derivatization-ultrasound-assisted emulsification microextraction. *Analytical and bioanalytical chemistry*, 411:787-795.

Zhao, L., Ni, Y., Su, M., Li, H., Dong, F., Chen, W., ... Martin, F.P. 2017. High throughput and quantitative measurement of microbial metabolome by gas chromatography/mass spectrometry using automated alkyl chloroformate derivatization. *Analytical chemistry*, 89(10):5565-5577.

Zhao, Z., Hamid, N., Gutierrez-Maddox, N., Kam, R., Kantono, K., Wang, K., ... Lu, J. 2023. Metabolite profiling identifies chemical markers associated with the cytotoxic properties of roasted fermented avocado seeds. *Food Research International*:113131.

Zheng, H., Ni, Z., Cai, A., Zhang, X., Chen, J., Shu, Q. & Gao, H. 2018. Balancing metabolome coverage and reproducibility for untargeted NMR-based metabolic profiling in tissue samples through mixture design methods. *Analytical and bioanalytical chemistry*, 410(29):7783-7792.

Zhu, C., Zhang, Q., Zhao, X., Yang, Z., Yang, F., Yang, Y., ... Laghi, L. 2023. Metabolomic analysis of multiple biological specimens (feces, serum, and urine) by 1H-NMR spectroscopy from dairy cows with clinical mastitis. *Animals*, 13(4):741.

Zhu, K., Gu, B., Kerry, M., Mintert, M., Luong, J. & Pursch, M. 2017. Elimination of N, O-bis (trimethylsilyl) trifluoroacetamide interference by base treatment in derivatization gas chromatography mass spectrometry determination of parts per billion of alcohols in a food additive. *Journal of Chromatography A*, 1490:74-79.