ANNEXURE A

METABOLOMICS: TECHNOLOGY, APPROACHES AND DATA MINING

A.1 METABOLOMICS APPROACHES AND TECHNOLOGY

A.1.1 INTRODUCTION

This study involves *metabolomics* which, from a methodology point of view, is a relatively recent development in biological investigations. For this reason, and to explain the rationale behind the methodology used in this study, which was the first of its kind on the metabolomics platform at the Centre for Human Metabonomics, North-West University, it was decided that an overview of metabolomics approaches and technology in this thesis is useful and necessary.

Since the extensive development of system biology in the last part of the twentieth century, scientists began to look at biological systems in a more holistic way. The holistic or global view of biological systems is the key to 'omics' research. In addition to holistic approaches, a paradigm shift started to occur whereby it is believed that the genetic blueprint alone cannot explain biological systems or phenotypes, especially in relation to their environment and all related perturbations. In the current 'omics' era, systems biology approaches are used at multiple biological levels to elucidate the effects of internal (transcription, mRNA degradation, posttranslational modification, protein function and metabolite concentrations and fluxes) and external (environment, molecular signals, etc) parameters (collectively called "epigenetics"). In the following illustration (Figure A.1) the main omics fields are shown, which contribute to the full understanding of biological systems and gene function (Madsen *et al*., 2010; Weckwerth, 2007).

Metabolomics plays an important part in this cascade. The metabolome is the final downstream product of the genome and is closest to the functional phenotype of the cell which may give us more direct answers than the proteome or transcriptome (Dettmer *et al*., 2007; Dunn, 2008; Roessner & Bowne, 2009; Sumner *et al.,* 2007). Many studies have shown that alterations of metabolites (under specific conditions) can be observed even when alterations in the concentrations of proteins and transcripts are not detectable (Dunn, 2008). Furthermore, analysis of the metabolome looks more attractive than the analysis of the proteome or transcriptome from the view point that it consists of < 10 000 estimated small molecules (metabolites) in contrast to the

millions or tens of thousands of proteins, transcripts and genes (Tolstikov *et al*., 2007). Although *metabolomics* can simply be described as the study of the metabolome, its meaning often differs among researchers (Dunn, 2008; Goodacre *et al*., 2004; Nicholson *et al*., 1999). Therefore, it is necessary to clearly define the terminology used in this thesis. In the following section the definitions of relevant terms used for the study designing and reporting are given.

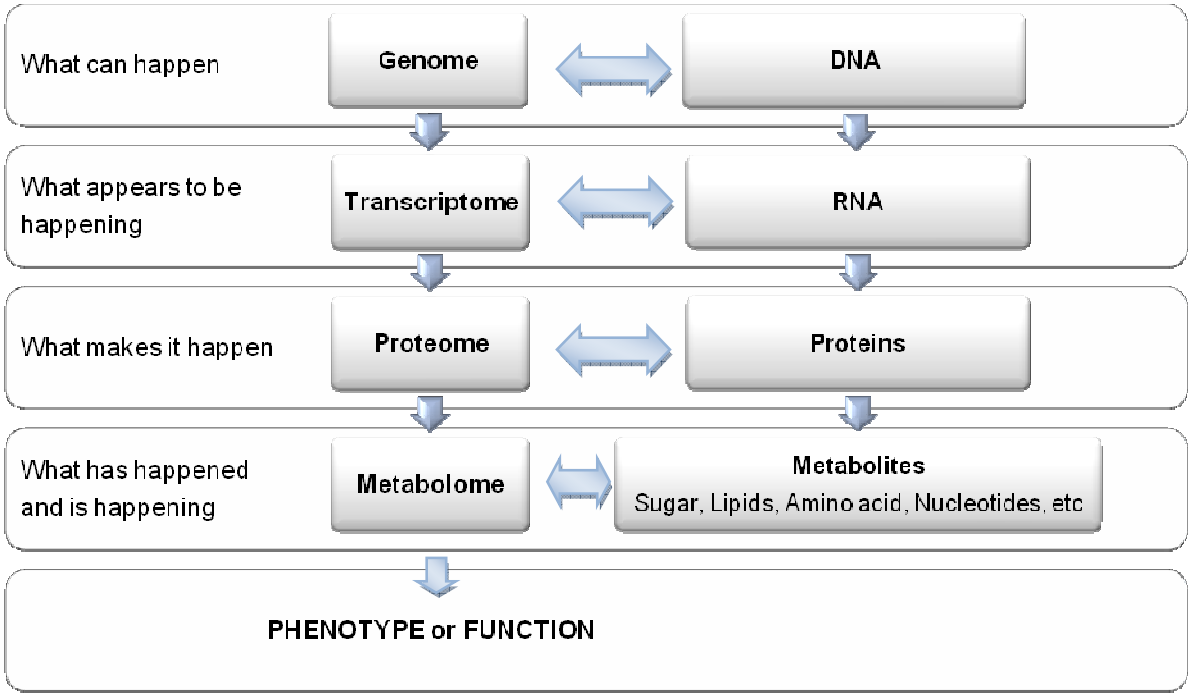


Figure A.1: The 'omics' cascade. Comprehensive data on each level of this cascade contribute to the initiative that is systems biology (adapted from Dettmer *et al*., 2007).

A.1.1.1 Terminology

Due to the variations in the terminology used among researchers, the definition for '*metabolomics*' can be broken up into the following key descriptive elements to give a comprehensive definition as found in the literature.

Metabolomics: – (ambitiously) aims to asses / detect / quantify metabolic changes;
– in a comprehensive / global / holistic and;
– non-selective / unbiased manner;
– in a biological sample;
– in order to infer biological functions and;
– provide the detailed biochemical responses of cellular systems;

– at a given time (snapshot).

(Favé *et al*., 2009; Fiehn & Kind, 2007; Lu *et al*., 2008; Ryan & Robards, 2006; Tolstikov *et al*., 2005; Want *et al*., 2007; Weckwerth & Morgenthal, 2005; Weckwerth, 2007.)

Metabonomics: This term is very similar to *metabolomics* in that it aims to quantify and identify numerous metabolites in the metabolome. The difference is mainly in the models that are used. It is fit to use this term for medical, toxicological and pharmaceutical models where the focus is on the dynamic multi-parametric metabolic responses of living systems to pathophysiological stimuli (sick vs. healthy, wildtype vs. mutant) or drugs (treated vs. untreated). The term *metabolomics* is preferred in plant and other studies where the focus is more often than not the elucidation of gene function or the identification of novel metabolites (Dunn, 2008; Goodacre *et al*., 2004; Nicholson *et al.,* 1999).

Metabolome: It is generally defined as the total (quantitative) collection of low molecular weight compounds – or metabolites – present in a cell or organism which participate in metabolic reactions needed for growth, maintenance and normal function (Dunn, 2008). The term *sample metabolome* is also commonly used in light of the definition for metabolomics and refers to all the metabolites in a given sample or tissue type. Therefore when studying red blood cells, which do not contain certain metabolites and metabolic pathways, it is better to use the term 'sample metabolome'. Moreover, not all the metabolites in the organism are quantified when using a single tissue type (such as red blood cells). Hence, the use of sample metabolome ensures that no false conclusions are made.

Exometabolome: Also known as the metabolic footprint. It describes the influence of the intracellular metabolic network on its external environment by the uptake of extracellular (exogenous) metabolites and secretion of intracellular (endogenous) metabolites. Both blood and urine can be described as metabolic footprints and provide a cumulative picture of mammalian cellular metabolism for the complete organism. Similarly, when working with cell cultures, the growth medium is seen as the exometabolome (Dunn, 2008; Kell *et al.,* 2005). When investigating the metabolome of cells or tissue samples, the level of metabolites at any given time represents a composite of both catabolic and anabolic processes and is only a snapshot of the metabolome at that particular time (Ryan & Robards, 2006). When looking at it this way urine, and to a lesser extent blood, is more advantageous as it provides a cumulative picture (over a period of time).

Publications by Nicholson *et al.* (1999) and Fiehn *et al*. (2000) were of the first to propose that metabolic analyses should not be restricted to a handful of compounds but that a more holistic metabolic screening approach should be used when studying metabolic differences. Hendrik Kacser also stated that to understand the whole, one must study the whole (Goodacre *et al*., 2004). The applications of metabolomics have expanded rapidly over the last decade especially in the field of functional genomics, where gene knockout or knockdown models are commonly used to study the role of genes in biological systems (Fiehn, 2002; Griffin, 2006; Roessner & Bowne, 2009). *Quite frequently the repercussions of a single genetic alteration are not limited to a single metabolite or metabolic pathway which means that a (classical) targeted approach is not satisfactory.* In addition, when a silent mutation is studied or a mutation of which the effect is not known, it might be better to look at the broad view than to limit the scope (Fiehn, 2002; Madsen *et al*., 2010).

A.1.2 METABOLOMICS APPROACHES

The first and most important step to study metabolic responses is to thoroughly plan the study and to define the objectives carefully as there are several approaches that can be followed (Figure A.2). For example, when the primary biochemical response of a genetic alteration is *a priori* known and narrowed to one or two metabolites, it is best to use the classical, hypothesis-driven approach of targeted analyses (Fiehn, 2002). When the biochemical response can be narrowed down to a single set of metabolites (common in chemical property or pathway) it is best to use the hypothesis-driven approach of metabolic profiling (Dettmer *et al*., 2007). However, when the metabolic response of a genetic alteration is not known, nor predicted, it is best to use an explorative, hypothesis-generating approach to *search* for metabolic differences (Dunn, 2008). In light of the above, it is evident that there are mainly three different approaches to study metabolic responses, namely targeted analyses, metabolic profiling and metabolic fingerprinting (Dettmer *et al*., 2007; Shulaev, 2006; Steinfath *et al*., 2008). The main differences and definitions of these approaches will be discussed in the following paragraphs.

Targeted analysis is the ‘absolute’ quantification and identification of a single or highly-related small set of *a priori* known metabolites after significant sample preparation to separate the target metabolites from the sample matrix (Dunn, 2008; Fiehn, 2002; Fiehn & Kind, 2007; Koek *et al.,* 2010). The term ‘absolute’ is used in the context that the target metabolite is quantified in the best possible way with selected technology (and associated methodology). With mass spectrometry, stable isotope labelled internal standards is generally used to achieve ‘absolute’ quantification. These isotopes are very expensive and isotopes of only a few compounds are commercially

available (Fiehn & Kind, 2007; Halket *et al*., 2005; Shulaev, 2006; Steinfath *et al*., 2008). Obviously, this method also ignores any unknown metabolites that might be influenced by the intervention (Favé *et al*., 2009)

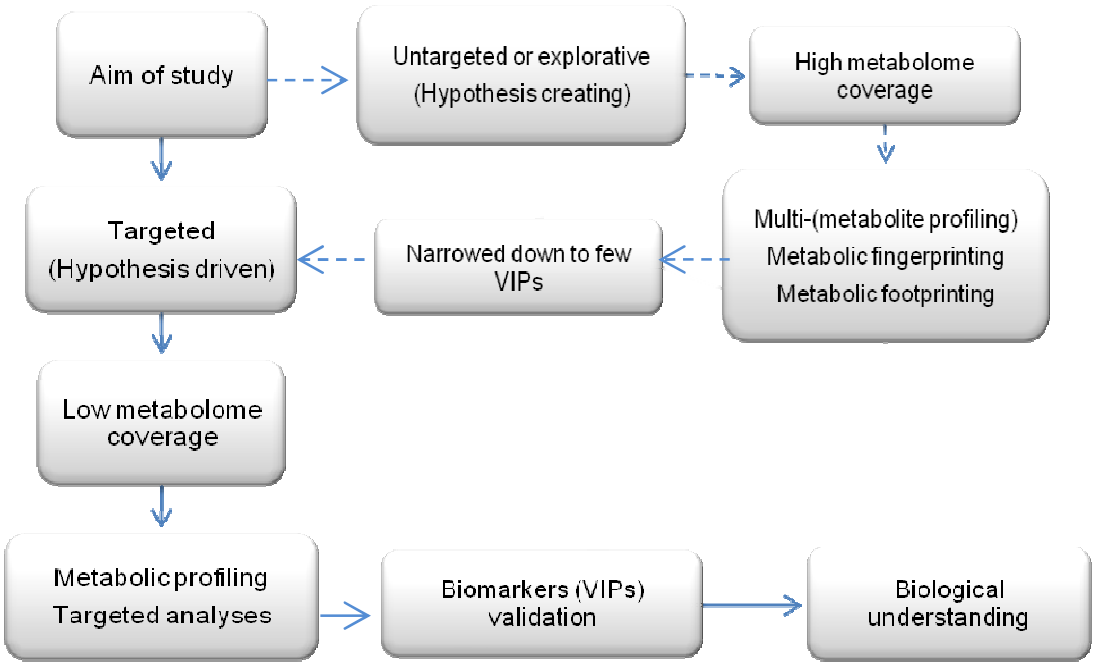


Figure A.2: Approaches to study metabolic responses in biological systems. The classical approach of targeted analyses (solid lines) remains important as it delivers quality data. A metabolomics approach must be seen as a detour (perforated lines) that should be exercised only when required and its findings confirmed with targeted analyses (Want *et al.,* 2007; Woo *et al.,* 2009). In accordance metabolomics approaches are best used to generate hypotheses which can then be tested with more targeted approaches (adapted from Dettmer *et al*., 2007).

Metabolic profiling is the hypothesis-driven, semi-untargeted detection and identification of a wide range of metabolites, generally related by pathway or metabolite class(es), employing single or multiple analytical platforms. It is more of a hypothesis-driven approach, rather than a hypothesis-generating one, where certain metabolites or classes of metabolites are selected for analysis depending on the questions asked (Dettmer *et al*., 2007; Koek *et al.,* 2010). It is therefore not a "true omics" approach, but the assembly of numerous profiles (or different pathways), like building blocks, can make it a powerful metabolomics approach. Thus it must be seen as a compromise between truly quantitative targeted analysis and completely unbiased metabolomics (Dettmer *et al*., 2007; Fiehn, 2002; Goodacre *et al*., 2004; Ryan & Robard, 2006; Shulaev, 2006; Steinfath *et al*., 2008).

Metabolic fingerprinting is a hypothesis-generating approach and is the high-throughput collection of a global snapshot, or fingerprint, of the endogenous metabolome of a crude sample with minimal sample preparation (Teahan *et al*., 2006). Identification and quantification is limited and the strategy is mainly employed as a tool for the discrimination of samples from different biological origins or status. This approach is also essential when the metabolic response to stimuli is more wide-ranging and the total metabolic response (patterns or fingerprints) is necessary to differentiate between experimental groups. Since this approach attempts to measure and take into account a wide range of metabolites, it can be considered as a "true omics" approach (Dettmer *et al*., 2007; Dunn, 2008; Fiehn & Kind, 2007; Shulaev, 2006; Steinfath *et al*., 2008; Sumner *et al*., 2007).

Metabolic footprinting is similar to metabolic fingerprinting, except that a global snapshot of the exometabolome is taken (footprint). This strategy is more high-throughput as there is (arguably) no requirement for quenching of metabolism or metabolite extraction as is the case for intracellular metabolomes (Dunn, 2008; Kell *et al.,* 2005; Roessner & Bowne, 2009; Sumner *et al*., 2007).

An aspect that should be kept in mind when a study is designed and an appropriate approach is selected, is that *quantity* and *quality* in metabolic research most often do not go together. Targeted analyses give quality data for a selection of targeted metabolites, as they are absolutely quantified, but lack quantity as the focus is on only one or two metabolites. Metabolic fingerprinting, in contrast, gives quantity data, as a large part of the metabolome is covered but lacks quality since metabolites are only relatively quantified. Therefore, the methodology and results from the targeted analysis of two compounds with LC-MS will look completely different from the method and results of a LC-MS method measuring 500 compounds. Hence, it is good practice to test and verify metabolomics findings with targeted and metabolic profiling methods as shown in Figure A.2 (Want *et al*., 2007; Woo *et al*., 2009).

A.1.3. ANALYTICAL PLATFORMS (TECHNOLOGIES)

The choice of analytical platform(s) for metabolomics investigations is often difficult and is generally a compromise between speed, selectivity and sensitivity (Favé *et al*., 2009; Fiehn, 2008; Sumner *et al*., 2007; Want *et al.,* 2010). For metabolomics investigations, the analytical platform must be able to accurately measure numerous known and unknown compounds that span a diverse chemical spectrum and a large dynamic concentration range (Ryan & Robards, 2006; Tolstikov *et al*., 2007). One of the best suited technologies for this is nuclear magnetic resonance (NMR) spectroscopy which is not only very high-throughput and robust, but also gives absolute

quantification as few variables influence the measured response (Dunn, 2008; Xin *et al.,* 2007). Mass spectrometry (MS) is another powerful metabolomics platform which provides multi-analyte detection with high sensitivities and specificities, especially when combined with high resolution chromatographic systems (Shuleav, 2006). The most commonly known and used hyphenated MS platform is gas chromatography mass spectrometry (GC-MS). It has been used in a vast number of applications and is favoured for metabolomics investigations (next to NMR). However, liquid chromatography mass spectrometry (LC-MS) is also becoming more attractive for metabolomics investigations as it is an information-rich platform which can analyse a much wider range of chemical species than GC-MS. Nevertheless even with these well established analytical platforms, it is currently impossible to get a global view of the metabolome with a single platform due to the complex nature of the metabolome.

The metabolome is chemically complex in the sense that it consists of various types of compounds with large dynamic ranges and biological variances (Tolstikov *et al*., 2007). Unlike the linear four-letter code for genes and the linear 20-letter code for proteins, the chemically complex nature of the metabolome complicates its analysis and coverage (Fiehn, 2002). Furthermore, the properties and biases of each analytical platform complicate matters even more. For example, some compounds in complex mixtures, such as plant extracts, will preferentially form positive ions, others negative ions, and some will be difficult to ionize during LC-MS analysis (Fiehn, 2008). It is thus advantageous to make use of a multi-platform (or multi-setting) approach when maximum coverage of the metabolome is needed, especially when each type of technology exhibits a bias towards certain compound classes (Dunn, 2008; Fiehn & Kind, 2007; Morris & Watkins, 2005; Roessner & Bowne, 2009; Shuleav, 2006; Tolstikov *et al*., 2007; Weckwerth & Morgenthal, 2005; Werner *et al*., 2006).

To use more than one analytical platform is common practice among metabolomics researchers, who use many complementary combinations to support the study design. For example, the researchers at Metabolon (Durham, NC, USA) use two different platforms, namely LC-MS and GC-MS, for metabolomics studies as small molecules can be very polar as well as very non-polar (Robinson *et al.,* 2008). To get broad coverage of the metabolome, Woo *et al*. (2009) used four different sample preparation techniques for GC-MS analysis, each technique focussing on different metabolites/classes. Hence, this is also a multi-platform approach or multiple metabolic profiling. Williams *et al*. (2006) used three analytical platforms to analyse as much of the metabolome as possible in tissue samples from normal and Zucker (fa/fa) obese rats. NMR, LC-MS and GC-MS (with silylation) were employed in this case. Other researchers use multiple analytical platforms in

parallel in order to detect as many compounds as possible (Lu *et al*., 2008). Van der Werf *et al*. (2007) developed a comprehensive metabolic platform which uses three GC-MS methods and three LC-MS methods.

With the wide variety of analytical approaches it is often difficult to select the most appropriate. There is no obviously ideal or completely wrong combination of platforms for a metabolomics study where no *a priori* information is available or where the scope is to cover as much of the metabolome as possible. Nevertheless, the choice of analytical platform(s) should complement the investigation at hand and should be justifiable when put on trial. For this the advantages, bottlenecks and drawbacks of each platform must be known. In the following sections, the advantages and disadvantages of GC-MS and LC-MS will be reviewed, as these platforms were mainly used in this study.

A.1.3.1 GC-MS: overview, advantages and limitations

Except for the advantage that GC-MS is one of the oldest and longest standing analytical technologies, there are also other advantages (summarised in Table A.1). GC-MS generally provides greater resolution and sensitivity than LC-MS and is predominantly suited for the smaller metabolites. These include compound classes appearing mainly in the primary metabolism such as amino acids, fatty acids, carbohydrates and organic acids (Gullberg *et al*., 2004; Shuleav, 2006). GC-MS has overall better repeatability and reproducibility compared to LC-MS and is also much more cost effective (t'Kindt *et al*., 2009). It is also (arguably) better suited for targeted analysis and metabolic profiling. The biggest advantage of this platform is perhaps the universal electron impact (EI) libraries (Shuleav, 2006) that assist in metabolite identification.

Unfortunately, there are also a few bottlenecks and drawbacks when it comes to this platform. The first and most important disadvantage is that it requires the derivatization of polar, non-volatile metabolites in order to make them apolar, thermally stable and volatile (Dettmer *et al*., 2007; Fiehn, 2008; Gullberg *et al*., 2004). Not only is this more laborious and time consuming in comparison to LC-MS, it also means that GC-MS has a bias toward volatile metabolites and metabolites that have active hydrogens in functional groups, such as –COOH, -OH, -NH, and –SH (Dettmer *et al*., 2007; Fiehn, 2008; Gullberg *et al*., 2004). Another major limitation of GC-MS is the upper mass limit of metabolites which exists due to volatility constraints and the mass added with derivatization. This means that molecules such as trisaccharides, many secondary metabolites and membrane lipids cannot be analysed or detected with GC-MS (Fiehn, 2008; Tolstikov *et al*., 2005; Want *et al*., 2007).

Table A.1: Summarised advantages and disadvantages of GC-MS vs. LC-MS. (Agilent, 2009; Shuleav, 2006; Want *et al*., 2007; Weckwerth & Morgenthal, 2005).

GC-MS	LC-MS
High chromatographic resolution	Lower chromatographic resolution
EI does not suffer from ionization suppression	ESI can suffer from ionization suppression
Libraries + fragmentation patterns = relatively easy ID (unknowns more difficult as molecular ion mass lacks)	No libraries = relatively difficult ID but molecular ion mass helps as it can give few candidates
More suitable for targeted analysis & metabolic profiling (such as organic acid analysis)	More suitable for discovery metabolomics i.e. fingerprinting but also profiling (amino acid analysis)
Low running costs	Relatively high running costs
Limited to volatile and derivatizable metabolites + limited mass range	Can analyse a much wider range of metabolites
More sample preparation required (due to derivatization)	Minimum sample preparation required
Shorter running times?	Relatively longer running times?
Retention time shifts limited	Retention time shifts occur more often
Less matrix effects and noise	Higher noise levels and matrix effects

? depending on setup, columns and focus.

Another disadvantage of GC-MS is that the peak areas of the metabolite derivatives containing -NH_2, -NH and –SH groups, cannot strictly be used to extract quantitative conclusions about the physiological state of the investigated systems (Kanani *et al*., 2008). Their inclusion in the bioinformatics analysis is expected to significantly distort the final results as the observed changes in their peak area profiles might be due to experimental and not to biological factors. On top of all this, the stability of derivatives, the influence of moisture during derivatization and the costs of derivatization reagents are a great concern and disadvantage to this otherwise excellent analytical platform. Moreover, while EI and EI libraries are one of the biggest advantages in GC-MS, these can also be restrictive when dealing with unknown compounds. Since the molecular ion is often lost with EI, it is difficult to identify unknown compounds purely on fragments. There is to date no “universal” derivatization agent available that can lead to one derivative for every metabolite, independent of its chemical class. This is the main reason why GC-MS metabolomics are

restricted and why researchers are turning more and more to LC-MS when performing true untargeted analysis (Kanani *et al*., 2008). Despite the limitations, GC-MS remains one of the most used analytical platforms in metabolomics as it is complementary to LC-MS.

A.1.3.2 LC-MS: overview, advantages and limitations

LC-MS is slowly becoming a powerful tool in metabolic research despite a few hurdles (Jonsson *et al*., 2005). It is an information-rich technique which can analyse a much wider range of chemical species (than GC-MS). It can separate and detect metabolites that are not volatile and which have not been derivatized. Consequently, LC-MS is best suited for a discovery-based metabolomics approach when researching unknown metabolites and is complementary to NMR and GC-MS (Jonsson *et al*., 2005; Tolstikov *et al*., 2005; Want *et al*., 2007). Moreover, it is especially well suited to compounds belonging to secondary metabolism (t'Kindt *et al*., 2009; Shuleav, 2006; Tolstikov *et al*., 2005) – organic compounds not part of the primary metabolism and not directly involved in growth and development (Wikipedia, 2010). The most attractive advantage of LC-MS is that no laborious sample preparation and derivatization is required, which makes LC-MS metabolomics (arguably) higher throughput than GC-MS metabolomics. In addition, with fewer sample preparation steps comes less bias towards certain compounds, technical variance and sample loss. A major advantage of the LC-MS platform is that the molecular ion is almost always produced by ESI (electrospray ionization), which aids in metabolite identification. Furthermore, with ESI it is possible to analyse both the positively and negatively charged molecules simultaneously in a single run which gives a broader coverage of the metabolome (Dettmer *et al*., 2007).

As with GC-MS, this platform also has several weaknesses and shortfalls with the first being its running costs, which are much higher than GC-MS (Agilent, 2009). Except for the high running costs, the lower chromatographic resolution, repeatability and reproducibility are of the greatest disadvantages of LC-MS (t'Kindt *et al*., 2009). Unlike EI, ESI easily suffers from matrix effects such as ionization suppression and enhancement caused by the presence of salts and other compounds being ionized at the same time (Burton *et al*., 2008; Fiehn, 2008; Halket *et al*., 2005; Morgenthal *et al*., 2007; Roberts *et al*., 2008). Another major drawback of LC-MS-MS particularly is that there are no universal fragmentation spectra libraries available, mainly because these spectra can vary between different mass analysers and brands (Dettmer *et al*., 2007; Halket *et al*., 2005; Shuleav, 2006). This makes the identification of compounds difficult, especially when the molecular ion alone is used for this purpose (regardless of the accurate mass given by time-of-flight mass analysers). In addition to this, ESI also leads to the fragmentation of certain compounds in a complex mixture, which complicates matters even more.

With the wide choice of mass analysers on the market, it is often hard to choose a suitable instrument for the intended work, not to mention the choice of ionization sources. The two main types of mass analysers, namely quadrupoles (Q) and time-of-flight (TOF) mass analysers, function quite differently from each other. The oldest and perhaps most used is the quadrupole which has comparatively high pressure tolerance, good dynamic range, and excellent stability, all at a relatively low cost (Timischl *et al*., 2008). Since the development of ESI in the early '90s, the triple quadrupole (QQQ) became a popular and essential instrument in the clinical field and the screening of inborn errors of metabolism. The true power of this instrument comes when it is used in a (semi) targeted manner where, for example, a single product ion is used to monitor a whole class of metabolites (Lu *et al.,* 2008). However, due to limitations in scanning and detection speed, this type of mass analyser does not give favourable sensitivity when untargeted analysis is performed and hence, is not truly suitable for metabolic fingerprinting.

On the other hand the increasing instrument-of-choice, time-of-flight (TOF) mass analyser, offers high resolution, fast scanning capabilities, and mass accuracy in the order of five parts per million (ppm) (Timischl *et al*., 2008). Since all ions are given the same kinetic energy and are then detected by their time of flight, no ions are "missed" due to changing scanning windows. This instrument is therefore suitable for untargeted analysis of compounds and metabolomics types of investigations. The quadrupole-TOF (Q-TOF) mass analyser is highly recommended when it comes to metabolomics research as it gives the best of both worlds. It combines the stability of a quadrupole analyser with the high efficiency, sensitivity, and accuracy of a TOF reflectron mass analyser. The Q-TOF and TOF are equal in sensitivity, but when it comes to the lower masses (< 200 m/z), the Q-TOF (~ 2 ppm) outperforms the TOF (~ 4 ppm) with mass accuracy. Q-TOF analysers also offer significantly higher sensitivity and accuracy over tandem quadrupole instruments when acquiring full fragment mass spectra (Want *et al*., 2007).

With the high sensitivity of the TOF and Q-TOF instruments, they are able to scan for positive and negative ions simultaneously in a single run. This gives a higher coverage of the metabolome but at a price. The sensitivity of the instrument is compromised for higher coverage when both the positive and negative ions are simultaneously detected in a single run, similar to the effect seen when the quadrupole is used for untargeted analysis. The use of dual-ionization is thus often questioned, especially as only little additional information is obtained from the negative ionization when the instrument and mobile phases are predominantly set for positive scan (and *vice versa*) (Llorach *et al.,* 2009). Hence, when comparing the total amount of extra metabolites gained by negative ionization with these settings and the amount lost due to compromised sensitivity, it is

(arguably) better to perform only positive or negative ionization in order to use the full potential of the selected instrument. When it is mandatory to perform both, it is better to do it in separate runs where the mobile phases can be adjusted for maximum performance (ionization). Llorach *et al*. (2009) reported that positive ionization produces overall more information (number of ions) than negative ionization.

A.1.4 METABOLOMICS (FINGERPRINTING) METHODOLOGY

It is evident from the above sections that metabolomics methods depend greatly on choices and compromises – compromises between quantity and quality, precision and accuracy, speed and efficiency (Fiehn, 2008). Thus, compromises are unavoidable, but the approach and method of choice should complement the work being done (Koek *et al.,* 2010). This is also true when selecting sample collection and preparation methods. The choice of sample collection and preparation methods is also a compromise between efficiency, speed, coverage and repeatability. While it is accepted that compromises have to be made with regard to quality and quantity in metabolomics (Morris & Watkins, 2005), the number of studies focusing on sample preparation and metabolite extraction protocols indicate that this area is a confounding factor for the quality of metabolomics (Fiehn, 2008; Gullberg *et al*., 2004). Metabolomics protocols are also very different from former classical protocols. Repeatability, robustness, practicality and comprehensiveness are more important than absolute recovery, absolute quantification and detection limits to keep with the general definition of omics technologies (Fiehn, 2002; Fiehn & Kind, 2007; Gullberg *et al*, 2004; Teahan *et al*., 2006; Weckwerth & Morgenthal, 2005). Hence precision of relative metabolite levels are more important than accurate metabolite levels. However, it must be noted that repeatability is not a guarantee for accurate results as incorrect results can also be repeatable (Morgenthal *et al*., 2007; Shurubor *et al.,* 2005).

A.1.4.1 Sample collection and storage

The choice of sample collection and preparation methods greatly depends on the study aim, metabolomics and analytical approaches, as well as the type of samples that need to be analysed. Bio-fluids, such as blood and urine, are generally easier to collect than tissue samples (biopsies), especially when working with human subjects. The information from bio-fluids is also effectively an integration of individual metabolic changes occurring within each of the organism's organs – i.e. systemic information and not organ-based information (Want *et al.,* 2010). Despite the ease of collection and preparation of bio-fluids for metabolic analyses, it is often desired to perform metabolic fingerprinting of tissue samples to study certain diseases or sites of toxicity (Viant, 2007).

Since metabolite turnover is extremely rapid in comparison to mRNA and protein turnover, it is of the utmost importance to 'quench' or stop all metabolic activity and preserve metabolite concentrations after sample collection. Several methods can be employed for this purpose. Snap-freezing in liquid nitrogen is a commonly used method and preferred over heating (or heat shock) methods to stop enzymatic activity. As it is impossible to stop all enzyme activity and metabolite degradation, samples are ideally kept at -80 ℃ after quenching to temporarily prevent metabolic decay. According to some reports, samples can be stored at this temperature for many months as it was shown that no significant metabolic changes could be detected even after nine months of storage (Deprez *et al*., 2002). However, shorter storage periods are obviously more ideal than longer periods (Deprez *et al*., 2002; Erban *et al*., 2007; Shuleav, 2006; Viant, 2007).

A.1.4.2 Sample preparation

The main prerequisites for a metabolomics (metabolic fingerprinting) sample preparation step is that it should be as simple, high-throughput and universal (comprehensive) as possible (Dettmer *et al*., 2007; Lu *et al*., 2008; Shuleav, 2006; Teahan *et al*., 2006). Comprehensive extraction and analysis of as many metabolites as possible from biological tissue is a considerably more complex task than extracting and analysing a limited number of target compounds (Gullberg *et al*., 2004). As mentioned, the extraction protocol always involves a compromise between efficiency and speed. Furthermore, despite the protocol used, there will always be a degree of analyte loss (Dettmer *et al*., 2007; Gullberg *et al*., 2004) limiting true absolute quantification. Since no extraction protocol gives perfect metabolite extraction without any disadvantages, it is obviously a choice that must complement the experiment (t'Kindt *et al.,* 2009). Moreover, increasing the sample preparation time by multiple extractions clearly disagrees with the scope of high-throughput analysis, which is detrimental for metabolomics studies that involve the analysis of many samples. Therefore, the goal must be to develop and use a simple and high-throughput method which gives the highest extraction efficiency and repeatability for as many classes of compounds as possible (Gullberg *et al*., 2004).

A.1.4.2.1 Metabolite extraction from tissue samples and cells

Liquid-liquid extraction protocols are generally used to extract metabolites from tissue samples (Dettmer *et al*., 2007). The choice of solvents (or buffers), solvent ratios and volumes strongly affects the number of metabolites that can be extracted and detected in metabolomics analyses. For most metabolites methanol (MeOH) and water is adequate to extract it from cells. However, it was shown that the addition of chloroform resulted in the detection of (on average) 7 to 16 % more compounds as it aids in the extraction of lipophilic compounds, such as fatty acids (Gullberg *et al*.,

2004). A recent paper from Wu *et al*. (2008) described a high-throughput and universal extraction method that is useful for NMR and MS based metabolomics. Several solvent combinations were tested and it was concluded that the methanol, water and chloroform combination was the best in terms of extraction yield and reproducibility (Wu *et al*., 2008). The method optimised and tested by the authors was in fact the Bligh & Dyer (1959) method that extracts polar and apolar metabolites from tissues in a two-phase manner. The Bligh & Dyer (1959) protocol has also been adapted several times so that a single phase is obtained rather than the two-phase extraction. In fact, it was shown that the use of a single phase containing both the lipophilic and hydrophilic compounds is sometimes more favourable in comparison to using two separate phases. Bligh & Dyer (1959) determined that a 2:0.8:1 MeOH:water:chloroform ratio is ideal for monophasic metabolite extraction while a 2:1.8:2 ratio is ideal for biphasic solution extraction. Gullberg *et al*. (2004) adapted this and used a ratio of 6:2:2 (MeOH:water:chloroform) to obtain a single phase containing both the polar and apolar metabolites.

According to Fiehn (2002) and t'Kindt *et al*. (2009) it might be advantageous for metabolomics approaches to extract metabolites from frozen tissues that contain the original amount of water in comparison to extracting freeze-dried samples. The main reason for this is that freeze-drying may potentially lead to the irreversible adsorption of metabolites on cell walls and membranes. In contrast, others promote the use of freeze-drying before homogenization as the variable water content in tissue samples can cause lower extraction reproducibility and enhanced degradation of the metabolites (t'Kindt *et al*., 2009). The temperature at which extraction of metabolites from tissue samples is done also differs among reports. It is evident form literature that higher temperatures (such as 70 ℃) give a higher metaboli te yield. While enzyme activity is limited at high temperatures (Gulberg *et al*. 2004; t'Kindt *et al*., 2009) it was also shown that certain metabolites are lost. In contrast, many researchers prefer metabolite extraction at -25 ℃ by working in acetone containing dry ice (Yang *et al.,* 2008). Although enzyme activity and metabolite loss is limited at this temperature, the yield is also low. Others proposed working at room temperature when using methanol to extract metabolites as methanol 'freezes' enzymatic activity (Kanani *et al*., 2008).

Liquid-liquid extraction of metabolites from tissue samples or cells is greatly dependent on the disruption (homogenization) of the tissue or cells. Although tissue disruption for research has come a long way, the commonly used and classical methods are not suited for metabolomics research. For metabolomics, it is critical to ensure that all samples are homogenized to the same extent so that the same amount of metabolites is released. The use of semi-automated

homogenization systems such as vibration mills is therefore mandatory for metabolomics research. Another name for vibration mills is 'bead beaters' and as the name implies, it uses beads to disrupt cell and tissue samples by vibrating or shaking at relatively high speeds. Many different types of beads (in terms of material and sizes) are available to suit the sample type. For instance, 3-5 mm Ø steel beads are commonly used to disrupt animal tissue such as liver and muscle while 0.5 mm Ø glass beads are used to disrupt cell samples.

A.1.4.2.2 Deproteinization of blood plasma and serum

Many of the problems and effort mentioned in the previous section can be omitted by studying the exometabolome such as blood and urine instead of the intracellular metabolome. Blood is the major vehicle by which metabolites are transported around the body. This makes the chemical analysis of plasma or serum very attractive in view of the fact that it can provide a wealth of information relating to the biochemical status of the organism under investigation (Daykin *et al*., 2002). As mentioned, plasma and serum are seen as the exometabolome which reflects the influence of the intracellular metabolic network on its external environment by the uptake of exogenous metabolites and secretion of endogenous metabolites (Viant, 2007). However, unlike urine, the complexity of plasma and serum hinders its direct use, and ignorant use of it can create more questions than answers. Metabolites that are commonly abundant in plasma are glucose (blood sugar), free cholesterol, saturated free fatty acids and a range of amino acids (especially alanine which serves as a three-carbon carrier between skeletal muscle and the liver) (Fiehn & Kind, 2007; Palazogly & Fiehn, 2009). Since blood is the vehicle for metabolites, it is easily influenced by the diet of an individual. For instance, although blood sugar is finely regulated, there is almost always an increase in concentration after a meal. Furthermore, the amount of free fatty acids and amino acids as alternative energy sources can also vary significantly when an individual is fasting. Therefore, careful planning and execution of the experiment is mandatory when using blood in metabolomics investigations.

To complicate things even more, some metabolites exist free in solution while others are bound to proteins or within the 'organised aggregates of macromolecules' such as lipoprotein complexes. Metabolites that are known to be bound to plasma proteins include tyrosine, phenylalanine, histidine, lactate and ketone bodies (Nicholson *et al*., 1989; Bell *et al*., 1988). Others that might be bound include citrate, lysine and threonine. It is commonly believed that clotting of the blood results in the loss of many metabolites that participate in clotting due to the entrapment of them in the protein mesh. As a result plasma is more commonly used than serum (Fiehn & Kind, 2007). However, results from Teahan *et al*. (2006) on human blood indicate a minimum difference

between them and it was shown that serum in fact contained an insignificantly small increase in triglyceride resonances.

Sample preparation of plasma and serum for metabolomics is much simpler in comparison to that of tissue samples, as metabolite extraction is not necessary. However, blood plasma and serum contain many proteins such as albumin and lipoprotein complexes. The only preparation thus necessary for plasma and serum is 'deproteinization'. There are a number of protein removal methods in the literature, all of which work differently and give different results. Since numerous metabolites are bound to proteins, a good 'deproteinization' method must be able to discard most of the plasma proteins while releasing as many of the bound metabolites as possible. This means that denaturing of the proteins to precipitate it would (arguably) be more advantageous in comparison to solid-phase extraction (SPE), where the protein structure remains intact. The selection of an optimal protocol will thus rely on the metabolome coverage and small amount of interfering proteins it gives at the end. Daykin *et al*. (2002) tested a number of plasma deproteinization methods for NMR analysis. They compared ultra-filtration (with 10 kDa cut-off), solid-phase extraction chromatography, acetone extraction, perchloric acid and acetonitrile protein precipitation at normal and low pH. They concluded that the precipitation of plasma with acetonitrile (ACN) at physiological pH, which is commonly used, proved to be a useful method for the quick and easy release of many small molecules such as isoleucine, valine, lactate, alanine, methionine and citrate. It also gave the highest number of metabolites with the best signal-to-noise ratio (Daykin *et al*., 2002).

A.1.4.2.3 Preparation of urine

One advantage of urine is that it can be used directly without any laborious sample preparation. This is especially true when using NMR and LC-MS. This means less technical variation or bias toward certain compounds (not to mention sample loss). The use of urine without any laborious sample preparation also keeps with the definition of *high-throughput* metabolomics and metabolic fingerprinting. However, urine occasionally contains high molecular weight molecules such as peptides and proteins which can cause problems for the chromatography column as well as for ionization of certain compounds. This is especially true for rodent urine (Bell, 1932; Pasikanti *et al.,* 2008; Want *et al.,* 2010). Hence, it is becoming common practice to prepare urine in a similar way to plasma and serum. Large molecules, such as proteins, are precipitated with organic solvents and removed with centrifugation. As with the blood, most prefer acetonitrile when removing proteins from the urine.

When it comes to metabolic profiling of urine using GC-MS, organic acid analyses remain widely used and preferred over other less laborious methods (Reinecke *et al*., 2011). This approach makes use of a semi-targeted extraction step using ethyl acetate and diethyl ether. All organic acids are forced into the organic phase by a lowered pH, which is then separated from the polar phase (Annexure B). When referring back to the definitions earlier in this chapter, it is obvious that this approach is in reality *metabolic profiling* as it is the "semi-targeted detection and identification of a wide range of metabolites, generally related by pathway or metabolite class" (Dettmer *et al*., 2007; Koek *et al.,* 2010). Despite the "limited and semi-targeted view" of this approach, it is able to detect and identify approximately 200 - 400 organic acids in urine samples. Taking note that the urine exometabolome is not large (probably <1000 compounds excluding xenobiotics), and considering the definition of *metabolomics*, it can arguably be considered a true 'omics' approach in the sense that it gives good coverage of the metabolome of the *sample* being investigated. If this or a similar method were used on tissue samples which contain a larger variety of compounds, it can then be argued that it is not a true "omics" approach when used without another complementary analytical platform. In light of the fact that organic acid analyses can be laborious and time consuming, the use of deproteinized urine for derivatisation and GC-MS analysis is becoming more common (Want *et al.,* 2010).

A.1.4.3 Technical variance and the use of internal standards in MS-based metabolomics

Any bias toward certain compounds must be avoided with metabolomics studies as it defines importance by the relative changes in metabolite abundances in comparative experiments. There must be confidence that the measured intensity of any metabolite in a sample is close to the true concentration and independent of matrix effects that are a notorious part of crude extract analysis (Fiehn, 2002; Teahan *et al*., 2006). Since variance plays an important part in chemometrics, all technical variance and bias must be known and, where possible, 'controlled' (Trygg *et al.,* 2007). There are mainly three types of variation in samples: relevant biological variation (which can be induced or 'amplified' by interventions), unwanted biological variation (such as diet-induced variation) and technical variance from experimental factors (van der Berg *et al*., 2006). For metabolomics and chemometrics experiments to give accurate and reliable results, one must always ensure that the relevant biological variance is amplified over all other variances (Sysi-Aho *et al.,* 2007). Unsupervised multivariate methods, such as PCA (Section A.4.5.2), focus on the highest amount of variance in the data (Scholz & Selbig, 2007). If the highest variance in the data is from diet, gender differences, unknown diseases, etc, the relevant biological variances that are studied (e.g. treatment vs. no treatment) are masked and harder to interpret. When the relevant biological variance cannot exceed that of the unwanted variance, it is best to fix the latter in order

to focus on the first. For example, if PCA show that gender differences dominate over treatment, it is best to evaluate the treatment effect in one gender group or both groups separately.

Besides unwanted biological variance, technical variance remains the most important threat to validity in biomarker research (Teahan *et al*., 2006). Technical variance can be described as any artificial (and unwanted) variance/bias induced during sample preparation and analysis. The introduction of technical variance (or experimental bias) in the sample preparation step means that it is 'hardwired' and cannot be eradicated by analytical or statistical approaches, no matter how advanced (Teahan *et al*., 2006). The induction of technical variation is, therefore, one of the greatest drawbacks in metabolomics which must be avoided to ensure that the statistics tells the true story at the end. Hence, it is of the utmost importance to standardise all experimental and analytical procedures in order to limit and correct unwanted technical variation and bias (Annexure B).

All encountered experimental biases in metabolomics can be classified in one of the following two categories:
Type A – When occurring, these biases affect the entire sample uniformly, and hence vary the measured signal of all the metabolites to the same extent.
Type B – When occurring, these biases affect individual metabolites differently and hence vary their measured signal to a different extent.

Type A biases are common among analyses that extract information from a physical quantity that cannot be directly measured through its proportional relationship with a measurable one. Thus, one step receptive and sensitive to this type of bias is the extraction step. The varying amount of supernatant recovery (between samples) after centrifugation, or the varying extent of drying (between samples) are considered Type A biases. These biases can be accounted for by the addition of a known quantity of internal standard(s) to each biological sample before extraction (Steinfath *et al*., 2008).

Type B biases cannot, however, be corrected or accounted for by the use of internal standards alone. Specific data correction and/or experimental optimisation methodologies that depend on the source of these biases and how they affect the final outcome have to be developed (Kanani *et al*., 2008). The derivatization step for example is not only a source of Type A errors but to a greater extent Type B errors, as derivatization favours some metabolites more than others. Ideally, the addition of a stable isotope for every anticipated metabolite would help correct for Type

B bias. However, MS-based metabolomics contains too many peaks to be absolutely quantified using thorough calibrations and the inclusion of an IS for every anticipated metabolite (Fiehn, 2008). For this reason it has become quite ordinary to include various internal standards representing each metabolite class (such as amino acids, sugars, fatty acids and organic acids) in order to correct for Type A, as well as Type B bias. However, to use these internal standards correctly would require the identification of all metabolites beforehand in order to normalise the detected amino acids with the included amino acid IS, and the sugars with the included sugar IS, etc. This also means that many unknown and unidentified metabolites will not be corrected (normalised) with the proper IS.

However, it must be stressed that absolute quantification is not reachable with metabolomics approaches using MS platforms for a number of reasons, including the dependence of response on a number of variables. As mentioned, MS-based metabolomics contains too many peaks to absolutely quantify everyone (Fiehn, 2008; Steinfath *et al*., 2008). Moreover, as MS-based metabolomics require metabolite extraction steps, the true concentration estimation of the metabolites would require 100 % extraction efficiency, which is impossible for many metabolites (Gullberg *et al*., 2004). Hence it is accepted that compromises have to be made with respect to the quantitative accuracy in metabolomics (Fiehn, 2008), which makes the selection of standards for metabolomics investigation more difficult (Dettmer *et al*., 2007).

Recently, a new method of normalisation with multiple internal standards was developed, called NOMIS (normalisation using optimal selection of multiple internal standards; Sysi-Aho *et al*., 2007). This approach removes the need to identify all compounds beforehand as it takes into account all variance. In other words, this method assumes that the technical variance measured in the compounds can be modelled as a function of the variation of the standard compounds (Sysi-Aho *et al*., 2007). Despite recent trends and developments, many researchers still keep with the approach of using only one internal standard to normalise their data (Boudonk *et al*., 2009; Gullberg *et al*., 2004; Kanani *et al*., 2008; Xin *et al*., 2007). These researchers assume that a single standard can 'capture' enough technical variance for data correction and are adequate for metabolomics experiments for two reasons. Firstly, metabolomics uses normalised data or relative quantities instead of absolute quantities, and secondly, additional errors will always be present for metabolites that are normalised/quantified using internal standards that are not specific to them (Fiehn & Kind, 2007; Gullberg *et al*., 2004), such as when using standards of each metabolite class.

A.1.4.4 Derivatization strategies for GC-MS analysis

GC-MS requires derivatization to make polar metabolites apolar, thermally stable and volatile. There are several derivatization methods which include alkylation, acylation and silylation, which add specific groups to any functional groups containing active hydrogens, such as –COOH, -OH, -NH, and –SH (Dettmer *et al*., 2007; Gullberg *et al*., 2004; Fiehn, 2008).

A.1.4.4.1 Oximation and silylation

Silylation (specifically trimethylsilylation) is perhaps the most commonly used and most preferred derivatization technique for the GC-MS analysis of a relatively wide range of compounds. With trimethylsilylation, trimethylsilyl (TMS) groups are added to compounds, especially small metabolites of the primary metabolism (Fiehn *et al*., 2000; Kanani *et al*., 2008). Silylation not only makes the target metabolites thermal, stable and volatile, but also aids in compound identification via available EI libraries. When it comes to metabolic fingerprinting with GC-MS, a two-step derivatization procedure is often used in order to get a broad coverage of the sample metabolome (Fiehn *et al*., 2000; Kanani *et al*., 2008). The two-step procedure includes oximation (or methoximation) followed by silylation (Figure A.3). The oximation step is a valuable addition to the normal silylation procedure because of the fact that aldehyde and ketone groups react with methoxyamine to form methoxime derivatives (Kanani *et al*., 2008). This means that compounds such as keto acids (containing both ketone and carboxylic acid groups) are added only one TMS group instead of two, thereby preventing the formation of multiple peaks per single compound. For example, direct silylation of reducing sugars such as fructose and glucose leads to a number of different peaks related to cyclic and open-chain structures that cannot be completely controlled with altered reaction conditions. Fortunately, this cyclization of reducing sugars can be restricted with the additional oximation step (Fiehn *et al*., 2000). Furthermore, this additional step is also valuable for the detection of certain α-keto acids such as pyruvic acid, as oximation protects against decarboxylation (Fiehn *et al*., 2000).

According to Kanani *et al*. (2008), when it comes to GC-MS metabolomics, most technical variation comes from the derivatization step which makes the optimisation and standardisation of this step most important. Incomplete and inconstant derivatization of compounds can cause a great amount of confusion when analysing the data and is commonly the case with metabolites containing -NH_2, -NH and –SH groups (referred to as Category 3 metabolites) (Kanani *et al*., 2008). These authors have gone so far as to state that the peak areas of Category 3 metabolites cannot be used to extract quantitative information regarding the physiological state of the investigated system and should be excluded from bioinformatics analysis as they can significantly distort the final results

(Kanani *et al*., 2008). According to Kanani *et al*. (2008) asparagine, glutamate and pyroglutamate are considered Category 3 metabolites. Glucose, fructose, fructose-6-phosphate and glucose-6-phosphate are considered Category 2 metabolites, which contain -OH and/or -COOH and ketone groups but not -NH_2, while fumarate, isocitrate, citrate, pyruvate, malate and succinate are Category 1 metabolites containing only -OH or -COOH but no ketones.

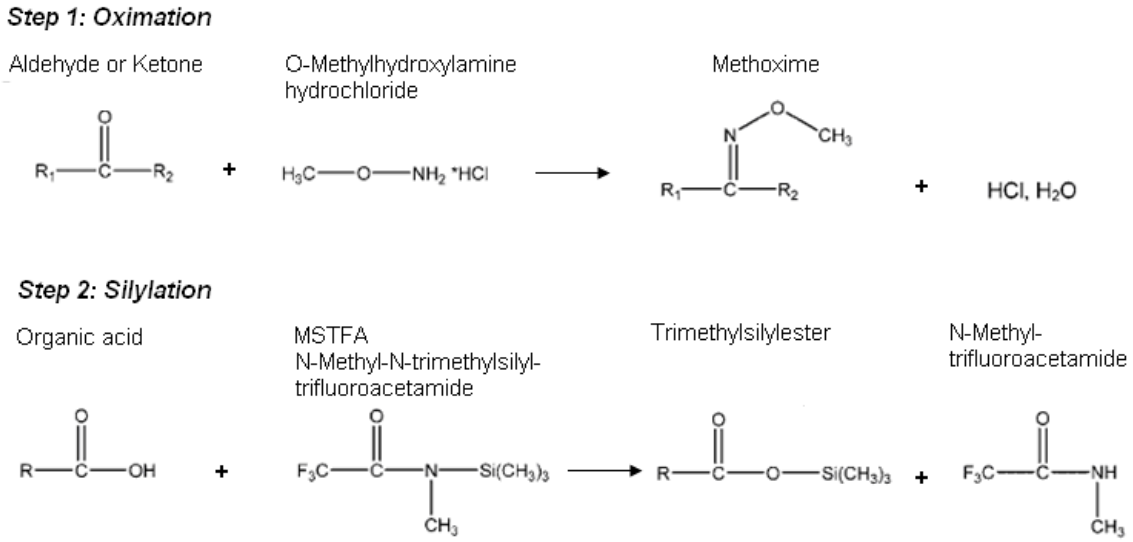


Figure A.3: The oximation and silylation reaction. For metabolomics both oximation and silylation are performed to limit multiple peaks for keto-acids and similar compounds (adapted from Dettmer *et al*., 2007).

Kanani *et al*. (2008) have shown that derivatization of Category 1 and 2 metabolites can be completed within a certain amount of time, while derivatization of Category 3 metabolites continues even after a day of silylation. Thus, these observations indicate that the derivatization time selected in the experimental protocol "*cannot correspond to complete silylation for all metabolites*" since 30 hour silylation times is not practical and can cause the degradation of other metabolites. Ascorbic acid, for example, can be oxidized yielding dehydroascorbic acid, and pyroglutamic acid is often formed by glutamic acid breakdown (Morgenthal *et al*., 2007). Therefore, it is important to find the best derivatization time that gives satisfactory derivatization while limiting metabolite degradation. Also, not only is the time frame important, but derivatization temperature (for the same reasons). Therefore, the choices are always between a long derivatization time and low temperature, or relatively high temperatures for shorter periods to ensure complete derivatization. Long derivatization time and high temperatures are seldom used (Gullberg *et al*., 2004). This rule also applies to oximation.

Compounds belonging to the Category 1 metabolites do not participate in the oximation process and neither are they directly affected by the duration of the oximation step. However, temperature might play a bigger part with these metabolites. Higher oximation temperatures might have a positive effect on the solubility of analytes in pyridine while it can also have a negative effect on others, causing them to degrade (Gullberg *et al*., 2004). Taking both effects into account, the oximation step is commonly performed at 60 ℃ for 6 0 minutes or at room temperature for 17 hours (A *et al.,* 2005; Atherton *et al*., 2009; Dettmer *et al*., 2007; Gullberg *et al*., 2004). Alternatively, both options can be simultaneously used by incubating the samples for 60 minutes at 60 ℃ followed by 16 hours at room temperature (Gullberg *et al*., 2004). This would increase solubility of compounds which are commonly difficult to dissolve, while still preserving many other compounds that are heat sensitive. Dettmer *et al*, (2007) used twelve reference compounds from different metabolite classes to study silylation (including oximation) efficiency and found that methoximation with methoxyamine for 17 hours at room temperature gave optimum results. Gullberg *et al*, (2004) on the other hand, specifically focused on the oximation of glucosamine and α-ketoglutaric acid, and found that oximation for 60 minutes at 60 ℃ followed by 16 hours at room temperature to be sufficient to ensure oximation of all compounds without the degradation of others, such as α-ketoglutaric acid. In contrast to this, it was seen that the temperature of silylation did not have any dramatic effect on the two tested compounds (Gullberg *et al*., 2004).

Silylation is performed after oximation. O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) is one of the commonly used and cost effective silylation reagents on the market giving reliable derivatization. However, its more expensive equivalent, MSTFA, is the most volatile of the TMS acetamides. Some researchers suggest that it is superior to BSTFA with respect to reactions with amines and amino acids (Dettmer *et al*., 2007; Fiehn *et al*., 2000; Koek *et al.,* 2010). When it comes to metabolomics studies, MSTFA is more often used because of its supposed higher reactivity. Seeing that metabolic fingerprinting requires the silylation of many chemical compounds in comparison to the limited number in more targeted analyses, it is obvious why most would take extra precautions to ensure sufficient derivatization. For similar reasons trimethylchlorosilane (TMCS) is also added to a final volume of up to 1 % (v/v) which acts as a catalyst for the silylation reaction.

In order to get complete and reliable derivatization without the degradation of certain compounds, it is also important to investigate the influence of time and temperature (in opposite directions) on the derivatization reaction (Gullberg *et al*., 2004; Morgenthal *et al*., 2007). Using twelve reference compounds from different metabolite classes to study derivatization efficiency, it was found that

silylation with MSTFA for 60 minutes at room temperature was the optimum derivatization procedure (A *et al.,* 2005; Atherton *et al*., 2009; Dettmer *et al*., 2007; Gullberg *et al*., 2004; Pierce *et al*., 2006). While many researchers use these conditions, there are also numerous others that prefer silylation for 30 minutes at 37 - 40 ℃ (Fie hn & Kind, 2007; Fiehn *et al*. 2000; Palazoglu & Fiehn, 2009; Shellie *et al*., 2005; Williams *et al*., 2006).

Obviously the selection of derivatization conditions should be optimised to complement the work being done and should be constant for all samples being investigated in order to limit technical variance. However, even with optimised and constant derivatization conditions the derivatization reaction is hardly ever stopped by the addition of quenching chemicals or fractionation schemes and continues over hours (Erban *et al*., 2007; Fiehn, 2008). This means that the actual derivatization duration between samples differs significantly. For example, if a run is about an hour, this means that the derivatization time of two adjacent samples differs by an hour, not to mention the difference of samples at the end of the batch. Although this approach is normally preferred when there is fear of metabolite loss due to drying or low re-suspension coefficients, the downside is that major variation are introduced into the experiment as some compounds (like the Category 3 metabolites) will be more completely derivatized, while instable derivatives may decompose and silylation artefacts or side products accumulate (Erban *et al*., 2007).

Another limitation that should be avoided in order to limit technical variance is moisture. Not only does moisture inactivate the derivatization agents, but it also results in the generation of polysiloxanes and highly toxic and corrosive hydrogen fluoride which can consequently produce hydrogen gas when it comes into contact with metal (Dettmer *et al*., 2007; Morgenthal *et al*., 2007). Another limitation of silylation is that certain amino acids such as arginine, belong to the more problematic classes of compounds. Arginine, for example, is converted into ornithine upon silylation (Dettmer *et al*., 2007; Fiehn, 2008).

A.1.4.4.2 Methylation of lipids

One of the most efficient ways to analyse lipid-containing mixtures is by direct infusion ESI (also called shotgun lipidomics) which can identify and relatively quantify over 450 phospholipids (Dettmer *et al*., 2007; Roberts *et al*., 2008). Most lipids are detected in the negative mode, while several others can be detected in the positive ion mode. There are also numerous neutral lipids which can be detected as lithium, ammonium or sodium adducts in the positive ion mode (Dettmer *et al*., 2007; Roberts *et al*., 2008). ESI in the positive mode allows the detection of acylcarnitines, phosphatidylcholines, lysophophotidylethanolamines, phosphatidyl ethanolamines and

sphingomyelins, while ESI in the negative mode allows the detection of free fatty acids, phosphatidic acid, phosphatidyl serine, phosphatidyl glycerol, phosphotidyl inositol, and phosphatidyl ethanolamines (Dettmer *et al*., 2007).

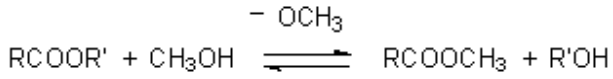


Figure A.4: The base-catalyzed methylation reaction. O-acyl lipids are methylated to yield fatty acid methyl esters (FAMEs)

The complexity and diversity of the lipidome also requires the use of multiple platforms to cover it sufficiently (Morris & Watkins, 2005). Since lipidomics is a field on its own, it is not always within the scope of a study to profile all the lipids, but rather to quickly fingerprint a reasonable part of the lipidome. For this, GC-MS remains the platform of choice (Roberts *et al*., 2008). The methylation of *o*-acyl lipids to yield fatty acid methyl esters (FAMEs) is often the preferred method when GC-MS is the analytical platform of choice (Dettmer *et al*., 2007). One major advantage of methylation is that the mass added to target compounds is low, which does not put the compounds at risk of being outside the mass range of the detector (Fiehn, 2008). Another advantage of current methylation protocols is that lipids do not have to be hydrolyzed to obtain the free fatty acids before preparing the esters, as most lipids can be transesterified directly. There are two non-commercial methylation methods commonly used, which are acid- and base-catalyzed methylation. Base-catalyzed methylation of lipids is a well documented method which gives great repeatability and superior reliability (White & Ringelberg, 1998). In short, this method makes use of anhydrous methanol to transesterify *o*-acyl lipids very rapidly in the presence of a basic catalyst as shown in Figure A.4.

A.2 METABOLOMICS DATA PROCESSING AND STATISTICAL ANALYSIS

A.2.1 INTRODUCTION

The “omics-era” is recognised with the total or global view of the biology in research models, which differs from the classical narrow and focused view. One major difference (and bottleneck) of this approach is the large data sets that are generated. With the larger data sets, not only is it necessary for the data processing methods to be adapted, but also the statistics. Commonly used,

"classical" univariate statistics are not always sufficient to extract the necessary biological information from the data, which means that multivariate statistics have to be incorporated into the omics platform. The importance of data extraction and processing of omics data has become a hot topic in recent years as faulty manipulation can lead to erroneous biological information and interpretation. The field of multivariate statistics plays an important role here and is the centre of discussion when it comes to the biological information mined from these large data sets. Hence, due to the importance of data processing and chemometric analysis, it is important to give a quick overview of all the steps involved in data extraction, processing, and multivariate statistics as illustrated in Figure A.5, which will also be the outline of Section A.4.

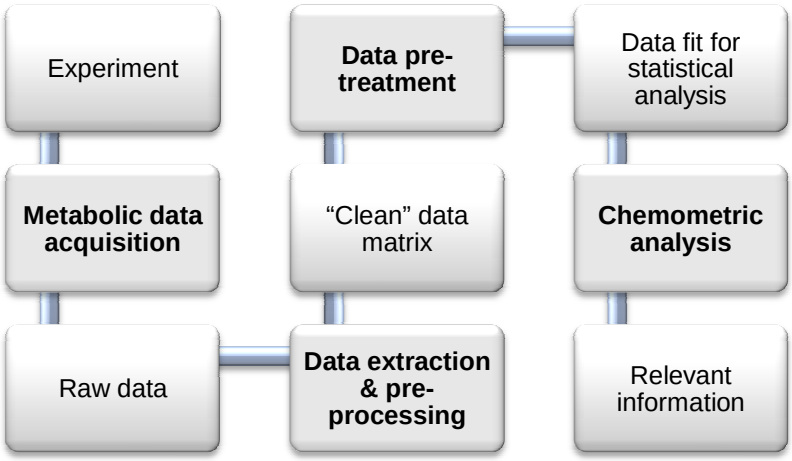


Figure A.5: The process of data mining to acquire relevant biological information from complex metabolomics data. After metabolic data have been acquired, the relevant data must be extracted, aligned and cleaned from the raw data which contain noise, irrelevant peaks, etc. The aligned data matrix, containing the clean data, are then pretreated or scaled so that the weight of the more abundant compounds are lowered. After data pre-treatment, the data is then fit for multivariate data analysis (chemometric analysis) (Craig *et al*., 2006). Figure adapted from Dettmer *et al*. (2007), Madsen *et al*. (2010) and van der Berg *et al.,* (2006).

The general workflow of a metabolomics experiment is as follows: first, a biological experiment is planned and executed, followed by the analysis of the metabolome of specific samples. After data acquisition via analytical instruments, the data are usually in the form of multiple chromatographs and mass spectra (raw data). Some data pre-processing (cleanup steps) can be performed on the raw (unprocessed) data as allowed by software packages, but these raw data files ultimately have to be converted into data matrices for further processing and statistical analysis. This step is commonly called *data extraction* (Dettmer *et al*., 2007; Nordstrom, 2008) and is by definition the

transformation of (mass) spectral data into vectors of uniform length (Dettmer *et al*., 2007). This term differs from the term *data mining* which refers to the extraction of relevant biological information from large data sets and is therefore a comprehensive term which include all the steps in Figure A.5 (Nordstrom, 2008). Following data extraction, more *data pre-processing* can be performed which includes more cleaning steps and data reduction, or blocking where necessary. After data pre-processing, *data normalisation* and *pre-treatment* is performed to put all samples and variables on "level footing" for comparison, respectively. Depending on the question being asked, data mining can take several different routes (Nordstrom, 2008; Steinfath *et al*., 2008). After completing these steps the data is then fit for statistical analysis to find relevant biological information from the complex data sets.

A.2.2 RAW DATA CLEANUP AND EXTRACTION

A.2.2.1 First order data cleaning

After analyses of the biological samples, the raw data are in the form of several chromatograms and mass spectra which might differ depending on the technology used. Since LC-MS and GC-MS raw data differ considerably, the data extraction approach and software are also different and sometimes specialised. Raw LC-MS data contains three dimensions which are retention time, molecular mass of detected compounds, and intensities. GC-MS data also consists of these three dimensions but instead of a single mass (or ion), a compound is represented by a mass spectrum containing fragment ions, making it more complex. However, despite the differences, similar steps are taken to clean the raw data before data extraction. The main steps involved in the first order cleaning of the raw data include smoothing (to reduce or filter noise), and baseline correction (Dettmer *et al*., 2007; Jonsson *et al*., 2005; Steinfath *et al*., 2008; Woo *et al*., 2009). These steps are essential and are nowadays mostly automated in the instrument-specific software (Sumner *et al*., 2007).

A.2.2.2 Data extraction

All the chromatograms and mass spectral data need to be converted into a more universal format (i.e. vectors, or matrices of numerical and string characters) in order to perform mathematical and statistical analysis on the data. The conversion of the raw data into matrices or vectors also requires the reduction of the dimensionality of the data. For example, to convert raw GC-MS data into data matrices without reducing the dimensionality means that each detected compound would be listed as a series of features (fragment ions and retention time) with intensities, rather than just a single entry with intensity. For this reason it is common practice to identify (or annotate) the GC-

MS peaks before converting the data into matrices as it will mean that each compound would only consist of a name and intensity. This classical way of extracting data works well with targeted analyses, where only a few peaks are of interest, but is unsatisfactory for the larger data sets as it is time consuming and requires *a priori* knowledge (library entry) of all the detected compounds. Metabolomics data is therefore more challenging and requires the automated, unbiased and high-throughput conversion of numerous data sets into organised data matrices (Fiehn *et al*., 2000; Want *et al*., 2007).

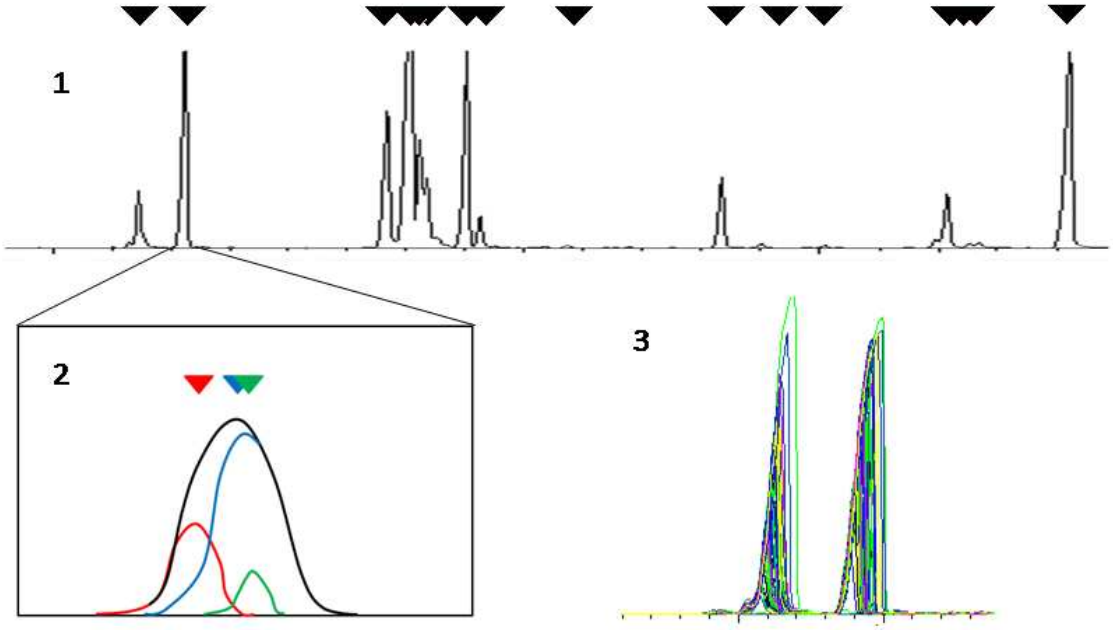


Figure A.6: The three main steps involved in data extraction. Peak picking (1) and peak deconvolution (2) is normally performed first, followed by alignment (3).

There are mainly three steps involved in data extraction (Figure A.6), namely peak picking (or feature extraction), deconvolution and alignment (in case of metabolomics data). Peak picking (or feature extraction) mainly involves the detection of chromatographic peaks among the background and baseline which represent the elution of compounds (Lidić, 2009; Theodoridis *et al*., 2008). The second process, deconvolution, goes hand-in-hand with peak picking and is performed simultaneously. Since metabolomics involves the analysis of crude and complex biological samples, many compounds co-elute and are thus obscured. The separate detection and quantification of these co-eluting compounds (deconvolution) are also an important property of data extraction (Fiehn *et al*., 2000). Peak, feature or compound alignment is the last step in data ex-traction and ensures that a specific variable represents the same detected compound in all

samples and that all the samples contain the same amount of variables for statistics (Jonsson *et al*., 2005; Nordstrom, 2008; Theodoridis *et al*., 2008; Want *et al*., 2007).

Many of the data extraction software packages (vendor-linked or freeware) are designed for classical use (targeted analyses) and are not suited for metabolomics data as they are only automated to some extent. Many of these packages can perform automated peak picking and deconvolution of a single sample or in some cases multiple samples, but lack the ability to align peaks, features or compounds to create an organised data matrix (Sumner *et al*., 2007). When using these software packages (such as AMDIS), the alignment is performed manually or with in-house written algorithms which is not high-throughput or analytically optimal. However, since the onset of metabolomics, many automated and high-throughput data extraction software packages surfaced which perform all three these steps, some better than others (Dettmer *et al*., 2007; Sumner *et al*., 2007). Most of these software packages use different algorithms for peak picking, deconvolution, quantification and alignment, which means that the data matrices created by the different metabolomics software are quite different and can lead to different results and conclusions (Lidić, 2009; Theodoridis *et al*., 2008). It is therefore often necessary to compare new software packages or combinations to find a data extraction package that would best fit the conditions of the study (Peters *et al.,* 2009) and the metabolomics work of an institute.

A.2.2.2.1 LC-MS data extraction

LC-MS data is less complex than GC-MS data which means that it is also easily manageable without the use of software. However, each compound contains information in three dimensions. These are RT, mass, and intensity which must first be reduced into two dimensions in order to compare samples, or to create a data matrix of numerous samples. Idborg *et al* (2005b) compared three data reduction methods to reduce the dimensionality of LC-MS data for further statistical analysis. The first method of data reduction is to collapse the mass axis. By doing this, the data is actually looked at in a TIC (total ion chromatogram) way – i.e. summing the intensities of all masses at each RT or scan. Thus, the “TIC chromatogram” of each sample is one vector in the matrix which can then be used as a fingerprint. In other words, each entry (peak) is listed only as an RT and intensity without using the mass. Only when a peak is statistically different between the sample groups would the mass under the peak be examined. With this approach, many alignment problems are overcome as only relative RT alignment must be performed. An obvious problem is when large or unseen RT drifts occur (Idborg *et al*., 2005b; Sumner *et al*., 2007). Another problem that can exist with this approach is when one compound under a peak increases in intensity while a co-eluting compound decreases (a common feature of ionization suppression), resulting in the

TIC peak being of the same size and not statistically different from the other. Jonsson *et al*. (2005) and Jonsson *et al*. (2004) used a similar approach in which the chromatograms were binned in time segments. Instead of summing the mass intensities to reduce the data within a bin, alternating regression was used, which is similar so PCA.

Another way of reducing the data is by collapsing the data in the other direction, the time axis. With this approach only the mass information is used regardless of its RT. Thus, the intensities of similar masses are summed even though they eluted at very different times (Idborg *et al*., 2005b; Nordstrom, 2008). This method gives similar results to direct infusion, but without the ion suppression. With this method the problem of RT drifts is avoided. However, one problem that may hinder this approach is when drifts occur in the mass axis (Idborg *et al*., 2005b; Sumner *et al*., 2007). These two approaches in data dimension reduction can give sufficient information and once the statistics gives a result, it is always possible to return to the actual raw data and confirm. A third method compared by Idborg *et al* (2005b) and also used by most software packages, is the extraction of each detected compound as a feature (RT and mass) which is then aligned across samples (Lange *et al*., 2008; Nordstrom, 2008; Tautenhahn *et al*., 2008). One problem with this (and actually with LC-MS metabolomics in general) is that compounds that may fragment are given as separate compounds in the matrix. This will not always influence the classification but will make the results harder to interpret (Idborg *et al*., 2005b). Since the latter method makes use of compound specific information by using both the RT and mass dimensions, it is obviously the method of choice. However, many researchers use the TIC method as it is much simpler to use, especially as RT shift and alignment are less troublesome compared to mass drift and alignment.

One problem that often exists when extracting LC-TOF-MS data with or without software is that the detected masses listed in the matrix contain accurate masses of up to four decimal places. If this accurate mass is used in this form and a specific compound is detected with slight differences on the fourth decimal in two samples, it will be listed as two different compounds and not aligned as the same compound. The obvious solution to this problem is to round these masses to a more preferred decimal such as two decimals and then unite (re-align) similar masses as it is in fact probably only one compound. However, how much rounding is acceptable? According to Dettmer *et al*. (2007) metabolites usually consist of carbon, hydrogen, nitrogen, and oxygen and to a lower extent sulphur and phosphorus atoms. While hydrogen and nitrogen increase the decimal part of the accurate mass, oxygen, sulphur, and phosphorus atoms decrease it. Therefore, the accurate mass of some metabolites rich in oxygen and phosphorus (e.g. fructose 1,6-bisphosphate $C_6H_{14}O_{12}P_2$) will be slightly less than the integer nominal mass. In addition, since metabolites are

considered small molecules, the decimal part of their accurate mass will rarely exceed 0.4 m/z for the majority of metabolites with molecular mass below ~500 m/z. Therefore, the optimum set of low-resolution bins for the majority of primary metabolites except heavy lipids is (m_{nom} − 0.1 m/z; m_{nom} + 0.4 m/z) and (m_{nom} + 0.4 m/z; m_{nom} + 0.9 m/z) where m_{nom} is a nominal mass. It is therefore safe to accept that the masses can be rounded to at least two decimal places (which is still a conservative approach) but can also be rounded to the first decimal (which is acceptable but holds risks).

In addition to the accurate mass problem, another dilemma that exists with LC-MS data is the fact that multiple salt adducts (such as sodium and potassium) can be formed for a single compound, thus complicating the data produced by increasing the number of peaks detected (Werner *et al.,* 2008). Many methods have been developed to deal with this problem including certain statistical methods such as principal component variable grouping (PCVG) (Ivosev *et al.,* 2008; Mohamed *et al*. 2009; Werner *et al.,* 2008). Fortunately, some of the newer software packages nowadays also deal with this (to some extent) such as MassHunter and MZmine. Nevertheless ignorance of this phenomenon can lead to many problems in data interpretation and should motivate the inclusion of such a deconvolution step in the data extraction procedure.

The choice of data extraction software/algorithm is not always an easy one as one package may have solutions for one of the above problems, whereas another might perform better with another problem. Furthermore, the high-throughput and ease of use of different software also play an important part in the selection process. Most instrument manufacturers produce their own software packages for data extraction but are often designed to work solely with data generated from their particular instruments. These include for example, MarkerLynx (Waters), MassHunter with Mass Profiler Professional (Agilent), and MarkerView (Applied Biosystems/MDS SCIEX). However, researchers desire freedom to modify parameters and data in a transparent way which is not always possible with these software packages. Hence more and more open-access and free software packages/algorithms are appearing which are not instrument-dependent and also more transparent (as the algorithms used are known and modifiable). Examples of these include MZmine, XCMS and MET-IDEA (Broeckling *et al.,* 2006; Fiehn, 2008; Tautenhahn *et al*., 2008; Want *et al*., 2007). Recently a number of papers have surfaced that deal with LC-MS data extraction. Lange *et al*. (2008) and Peters *et al.* (2009) assessed the alignment procedures of available software packages, namely OpenMS, msInspect, SpecArray, XAlign, XCMS, MZmine, MetAlign and Markerlynx. These software packages overall performed well but it was really OpenMS, XAlign and XCMS that were superior.

A.2.2.2.2 GC-MS data extraction

The structure and dimension of GC-MS data differ from LC-MS data as the molecular ion of compounds are not recorded but numerous fragment ions. The structure of this raw data is therefore much more complex. As with LC-MS, the GC-MS data dimensionality must also be reduced to create data matrices. While detected compounds in LC-MS data can be given as features with intensities, this is not possible with GC-MS. For this reason it is common practice to identify (annotate) GC-MS peaks before converting the data into matrices as it would mean that each compound would only consist of a name and intensity. As mentioned, this classical way of extracting data works well with targeted analyses, where only a few peaks are of interest, but is unsatisfactory for the larger data sets as it is time consuming and requires a library entry of all detected compounds (which is not always possible). However, this trend has remained with many researchers even when extracting data from metabolomics work, and AMDIS (Automated Mass Spectral Deconvolution and Identification Software) from NIST (US National Institute of Standards and Technology) has remained the workhorse for this (Lidić, 2009). Since metabolomics samples is very complex, chances are good that many peaks will remain unidentified and others semi-identified (uncertain) with this approach due to spectral contamination of co-eluting compounds and matrix effects. This is especially the case with low-abundant or novel metabolites (Fiehn, 2008). These unidentified and semi-identified peaks are often omitted form the data matrices. This approach not only results in the loss of potentially important biological information but also produces extra unwanted variance in the data (due to the missing values).

While AMDIS remains a favourite among researchers for extracting GC-MS data (Lidić, 2009), one has to recognise that it also has some major limitations for untargeted and high-throughput metabolomics work. Firstly, it cannot create a report of all the detected components in a sample but only those that have been identified (library hits). Secondly, while AMDIS is capable of analysing multiple samples at a time, it is only capable of generating a linear report from those results and not a data matrix of all the samples. This means that the alignment is often performed manually or with in-house written algorithms in software packages such as R or Matlab. In light of this, a few complementary software packages were released with the main purpose of aligning AMDIS results, such as SpectConnect, which also filters peaks that are not consistently detected in these batch comparisons (Styczynski *et al.*, 2007). Furthermore, many free and open-access software, as well as vendor-linked software packages, make use of AMDIS to process GC-MS data before they merely align the results. In fact, many of the untargeted data extraction software such as MET-IDEA (Broeckling *et al.,* 2006) are not capable of deconvolution and are dependent on AMDIS and other software for this (Fiehn, 2008). Whether they were designed with this gap that

AMDIS just happened to fill or built around AMDIS is not clear. Nonetheless, if so, it is perhaps due to the fact that AMDIS is universally recognised as one of the best deconvolution software packages with an algorithm exceeding most others. Stein (1999) developed this deconvolution algorithm based on the work of Dromey *et al*. (1976) in which model ions and the least-squares method is used to detect co-eluting compounds.

A.2.3 DATA PRE-PROCESSING

Since most multivariate statistical methods mainly focus on the (total) variance in the data, it is necessary to have as little technical variance as possible so that the multivariate statistics can identify and show the relevant biological variation. Technical variance can be introduced in almost every step of the metabolomics workflow of which the sample preparation, analysis and data extraction steps are key to this type of variance. Although this unwanted variance can be limited with the use of standardised conditions and methods, it is impossible to limit all such variance. It is therefore mandatory to clean the data matrix further before performing statistical analyses (Burton *et al*., 2008). Cleaning of the data matrix is also collectively called data pre-processing and includes a number of steps (Shuleav, 2006; Steinfath *et al*., 2008):

- Further compound alignment if necessary (Dettmer *et al*., 2007; Woo *et al*., 2009)
- Subtraction of blanks from samples
- Removal of uncommon variables not detected in most of the samples
- Removal/correction of known artefacts from data such as over-deconvoluted peaks in GC-MS data or salt adducts of LC-MS peaks (Jonsson *et al*., 2005)
- Filtering of components with small variance (between sample groups) (Burton *et al*., 2008; Jonsson *et al*., 2005; Scholz & Selbig, 2007)
- Removal of unstable variables (with high CVs) and removal of outlier variable values (by replacing them for example with the mean of group)
- Replace missing values. One method is to assign the group mean (Bijlsma *et al.,* 2006; Morgenthal *et al*., 2007; Steuer *et al*., 2007).
- Discard outlier samples (t'Kindt *et al*., 2009; Steuer *et al*., 2007; Urbanczyk-Wochniak *et al*., 2007)

In many cases it is necessary to re-align compounds after data extraction. With LC-MS data, a mass (or retention) drift often exceeds that of the alignment parameters in the software. In such cases, these parameters must be adapted to allow the larger drift, or re-alignment must be done manually. With GC-MS data analysed with AMDIS, certain compounds are often identified with

much uncertainty resulting in the addition of a question mark ('?') to its name. Since this unknown is not aligned with the confidently identified compound, it takes some manual labour to investigate whether this compound is in fact what its name implies and then to re-align it correctly. In some cases it is obvious, for example, when four out of five samples contain data for compound 'A' while the fifth sample does not contain data for compound 'A' but for compound '?A'. If the retention time correlates with that of compound 'A', one can assume that it is in fact the same compound. This is just an example of further (manual) compound alignment which not only takes time and effort but leaves the analyst to make educated guesses (which can also lead to erroneous results).

A blank sample is often included in the metabolomics experiment to monitor possible contaminant compounds in the sample preparation and mobile phase solvents. It is also of use to monitor chromatography carry-over when large batches of samples are analysed. It is then common practice to subtract the pre- or post batch blank from the rest of the samples (thus 'blanking' the signal to remove noise and contaminants). More data reduction steps are commonly performed in addition to the removal of contaminant compounds. Uncommon variables that are not detected in most of the samples (in a group) can also be removed from the data matrix (Fiehn & Kind, 2007; Steinfath *et al*., 2008). These variables carry no relevant biological information and just add more unwanted variance to the data. They are often a result of chemical contamination or compounds that are not extracted or analysed repeatedly. The removal of uncommon variables requires some manual labour but can also be performed automatically.

Certain known artefacts must also be removed from the data. Salt adducts of compounds analysed with LC-MS are an example of artefacts that must be removed from the data as they can complicate the outcome. The abundance of a salt adduct (M+Na) of a specific compound can be added to the abundance of the salt-less compound (M+H). Depending on the samples and data extraction software used, peptides and protein peaks can also be present and should be removed. GC-MS artefacts include over-deconvolution (where a single compound peak was divided into two components) or where a single compound was not fully derivatized and gives two separate peaks (Morgenthal *et al*., 2007). These duplicate variables must be corrected by adding the abundances or by removing the duplicates. It is also advantageous to remove variables that contain virtually no (inter-group) variance and thus no relevant information (Scholz & Selbig, 2007). A threshold for the amount of variance that should exist for variables to be used can be set. Such a threshold can greatly improve the outcome of the data (Maggio *et al*., 2008) but can also have the potential to ruin the final result. Since multivariate statistics focus on metabolite relationships (covariance), one can risk removing important information.

It is also advisable to remove unstable compounds with high intra-group (within a group) variance (Fiehn & Kind, 2007). This large unwanted variation indicates that these variables were probably not extracted, prepared and analysed precisely (repeatable) and can complicate the results. The coefficient of variance (CV) of each variable can be very useful to screen the stability of the various variables quickly (Mohamed *et al.,* 2009). It is even more useful to do this screening with the help of quality control (QC) samples analysed throughout the experiment. Since the QC samples are chemically similar and treated similarly, any compound that has a high CV would indicate that it is not trust-worthy and should be removed. Not only should unstable variables be removed but also outlier variables in the data set. A simple yet effective way of screening for outliers in the data set is to compare the mean and median of a row of variables. When a large difference is detected between the mean and median, it then needs closer inspection (Steinfath *et al*., 2008). If only one value was causing this instability (outlier value), it can be replaced by the mean of other values in the group. If there is no clear outlier value it could then be that the data are non-parametrically distributed (Fiehn & Kind, 2007).

However, before such screening for unstable and outlier variables is performed, it is (often) advisable to perform missing value imputation first. Missing (or zero) values can result from deconvolution and alignment failings, but also from low abundant compounds that were under the detection limit of the analytical instrument (Steinfath *et al*., 2008). In the latter case it is appropriate to keep the zero values but most of the time such variables are automatically removed from the list of variables as they often fall under those that are uncommonly detected (Steinfath *et al*., 2008). However, most of the other “missing” values should be filled in order to limit unwanted, irrelevant variance in the data. Several missing value imputation methods exist but the most common method of replacing the missing values is to assign the mean or median of the whole batch to that sample. Another simple method is to use the mean of its nearest neighbours (Steinfath *et al*., 2008; Steuer *et al.,* 2007).

The last part in data cleaning involves the screening for outlier samples and removing them from the data (Fiehn & Kind, 2007). Unfortunately there is no absolute method and criteria that can help one decide whether a sample is an outlier (Steinfath *et al*., 2008). This decision must be made by the individual self. Numerous statistical software packages use different measures and criteria for outliers. In multivariate statistics, Hotteling’s T^2 (Maggio *et al*., 2008; t’Kindt *et al.,* 2009) and the Mahalanobis distance measure are often used to screen for outliers. Several statistical packages have also been developed, of which the LIBRA (Library for Robust Analysis) toolbox in Matlab is

one example. This toolbox makes use of two distance measures simultaneously to screen for outliers. A PCA score plot with Mahalanobis distance on the x-axis and the orthogonal distance on the y-axis are used to identify possible outliers.

The data cleanup steps described is only the basics in data pre-processing. There are numerous other steps and more advanced methods described in the literature which deal with identified and unidentified experimental and instrumental effects. Some important papers on this matter are those of Burton *et al*. (2008) and Watson *et al*. (2006). After the data have been thoroughly cleaned to remove unwanted variation and irrelevant information, one can proceed to the last part before chemometrics analyses.

A.2.4 DATA NORMALISATION AND PRE-TREATMENT

A.2.4.1 Data Normalisation

The main function of data normalisation is to reduce the influence of experimental factors (non-biological contributions) (Jonsson *et al.,* 2004; Scholz & Selbig, 2007; Sumner *et al.,* 2007). Normalisation is normally performed column-wise when each column in the data matrix represents a sample, and each row a variable (detected com-pound). For this reason it is also called *sample* normalisation and is not to be mixed with the row operation of *metabolite* scaling (which will be discussed in Section A.4.4.2). Hence, normalisation makes the different samples more comparable with each other (Craig *et al*., 2006). There are a number of ways in which data can be normalised which fit the experimental setup. These can be categorised into the three categories described in the following paragraphs.

A.2.4.1.1 Normalisation to sample weight

One way of normalisation is to correct the data according to sample weight. For example, if a sample was concentrated to twice the concentration of the other samples before analysis, all the detected compounds of the concentrated sample should be halved for comparison purposes. Another example of where this type of normalisation is used is when different tissue sizes are used for metabolite extraction and metabolic fingerprinting. If the volume of the solvents is not adapted according to weight, the larger pieces of tissue could give higher amounts of metabolites and, if not corrected, erroneous information. Data from urine samples are often normalised with the creatinine values of the urine (Warrack *et al*., 2009) as creatinine is indicative of the urine concentration. Blood and twenty-four hour urine samples are also commonly normalised by the volume obtained or used (Warrack *et al*., 2009).

A.2.4.1.2 Normalisation to internal standard(s)

The most common sample normalisation method employed by analysts is the normalisation of data according to an internal standard (t'Kindt *et al.,* 2009; Xin *et al.,* 2007). This is done by dividing all (relevant) peak areas (abundances) by the internal standard peak area (abundance). Hence, if sample loss, incomplete derivatization or lowered detection response of the analytical instrument occurs, normalisation with the internal standard(s) can correct for this and make the samples more comparable. However, one of the main discussion points in metabolomics is the use of multiple internal standards. Since it is not recommended to use a single internal standard in metabolomics investigations due to the different responses of certain types of compounds, many researchers make use of more than one standard. A previous trend among the metabolomics scientists was to include a stable isotope internal standard of each class of compound (such as amino acids, fatty acids, sugars, etc) and to normalise each class with the relevant standard. This meant that all detected compounds had to be identified before-hand which was both time consuming and laborious. Also, this method does not allow the normalisation of unidentified (or unclassified) compounds.

This brought forth a few other methods of normalisation using multiple standards. Katajamaa & Oresic (2005) described a method in which the best standard for a series of compounds based on its elution position and mass is used for normalisation. Another new method that was recently developed by Sysi-Aho *et al*., (2007) called NOMIS (Normalization using Optimal selection of Multiple Internal Standards), has become very popular among metabolomics researchers (as seen at the 2010 Metabolomics conference in Amsterdam). This method assumes that the technical variance measured in the compounds can be modelled as a function of the variation of the standard compounds. Redestig *et al*., (2009) improved this method and called it CCMN (Cross-Contribution robust Multiple standard Normalisation) which takes into account the fact that the compounds themselves can cause direct variation on the internal standards. Examples of this are ion suppression in LC-MS, and insufficient chromatographic separation in which the concentration changes in one compound can cause variance of a different compound (Redestig *et al*., 2009).

A.2.4.1.3 Normalisation to total signal

Another way of normalisation is to normalise the detected peaks to the total signal so that the total sum of peaks are set, for example, to 10 000 (Atherton *et al*., 2009) or so that the largest peak is set to 1 (Scholz & Selbig, 2007). This method thus exploits the ratios between all the peaks and is also called unit vector normalisation (Scholz & Selbig, 2007), or scaling to total response (Craig *et al*., 2006; Goodacre, 2006). Although this type of normalisation can be very efficient, it also holds

much risks as contaminant peaks (and xenobiotics in urine) can result in the increase of the total signal, consequently resulting in the reduction of the other detected compounds after normalisation (Craig *et al*., 2006; Warrack *et al*., 2009). A better way of sample normalisation using the total signal was described recently by Warrack *et al*., (2009), which is especially useful when working with urine. This method, called MSTUS, makes use of the "total useful signal" rather than the total signal. Thus the total intensity of components that are common in all samples are used, which removes the influence of xenobiotics, chemical contamination and artefacts. It was shown that this approach gives similar results to normalisation obtained when osmolality is used to normalise urinary metabolites (Warrack *et al*., 2009).

A.2.4.2 Data Pre-treatment

In contrast to sample normalisation which is a column-wise process, data pre-treatment is a row operation in which the variables/metabolites (and not the samples) are made comparable (Scholz & Selbig, 2007). The main function of the data pre-treatment step is to reduce the impact of the high intensity variables on the results so that the lower intensity variables have more weight in the multivariate statistical results (van der Berg *et al*., 2006). If the data is not pre-treated before multivariate statistical analyses, the results will show insignificant but variable components with high intensities and neglect (possibly) significant but low abundance components or biomarkers (Dettmer *et al*., 2007). Therefore, depending on the research question, it is advisable to scale or transform the data with one or more of the methods described below (Trygg *et al.,* 2007).

Although any method or combination of pre-treatment methods can be used on the data, it is always recommended to make use of prior knowledge regarding the experimental outline, analytical setup and chemometric limitations (van der Berg *et al*., 2006; Scholz & Selbig, 2007; Steuer *et al*., 2007). For example, even though a small intensity or variance may contain great biological information, low abundant values are usually strongly corrupted by background noise and are therefore not always very useful (Scholz & Selbig, 2007). Furthermore, different data pre-treatment methods give different results (Madsen *et al*., 2010; van der Berg *et al*., 2006). Hence, it is often required that standardised procedures for data pre-treatment be used on all batches and that the choice of data pre-treatment method(s) be justifiable (Goodacre, 2006).

Data pre-treatment methods can be separated into three groups:

- Centering: Converts all values to fluctuations around zero and is normally used with the other pre-treatment methods (prior to PCA or other similar methods).

- Scaling: Converts data into differences in concentration relative to a scaling factor (by dividing each variable by a scaling factor).
- Transformation: Nonlinear conversions of the data to correct for heteroscedasticity.

In the following sections, some data pre-treatment methods (relevant to this study) will be discussed in short in terms of their goals, advantages and disadvantages.

A.2.4.2.1 Centering

With centering the mean of each variable is subtracted from all the samples and thus converts all concentration fluctuations around zero instead of around the mean. Hence, it adjusts for differences in the offset between high and low abundant metabolites, making the metabolites more comparable. For more robustness, centering with the median is sometimes more favourable to limit the influence of outlier variables on the centering method. In certain cases the mean or median of only the control samples is used to centre the data. However, centering alone does not reduce the influence of high abundant compounds greatly, and is not sufficient when the data is heteroscedactic – when the variables have different variances (Scholz & Selbig, 2007; Sumner *et al*., 2007; van der Berg *et al*., 2006).

A.2.4.2.2 Auto-scaling

Auto-scaling uses the standard deviation as the scaling unit and is basically the opposite of centering as it focuses on the correlations (similarities) instead of the covariances. After centering, the variables are divided by their standard deviation (across all samples) so that all compounds have a standard deviation of one. After auto-scaling, all variables have the same variance and thus an equal weight in the statistics. This type of scaling is especially useful when two or more different types of data are combined in one matrix prior to chemometrics analyses and is also very popular with other omics data (Scholz & Selbig, 2007; van der Berg *et al*., 2006). However, its use with metabolomics data has been limited (Gullberg *et al*., 2004) due to the fact that it inflates measurement errors and noise (which is commonly present in metabolomics data).

A.2.4.2.3 Pareto scaling

Pareto scaling is similar to auto-scaling except that the square root of the standard deviation is used as the scaling unit. This makes it a kind of intermediate scaling as it falls between auto-scaling and no scaling. Since the square root of the standard deviation is used, it still increases the importance of low abundant compounds but not as much as with auto-scaling making it one of the preferred scaling methods in the metabolomics field (Burton *et al*., 2008; Ivosev *et al.,* 2008;

Ramadan *et al.,* 2006; t'Kindt *et al*., 2009; Trygg *et al.,* 2007). The data therefore does not become completely dimensionless. However, this benefit of Pareto scaling can also be a disadvantage as the influence of the large fold changes in the data (large compound peaks with large variance) are not always limited enough for smaller variance to influence the chemometrics outcome (van der Berg *et al*., 2006).

A.2.4.2.4 VAST scaling

VAST scaling is short for variable stability scaling and is an extension of auto-scaling that make use of the coefficient of variance (CV) as a second scaling unit. Following auto-scaling, the variables are divided by their respective CVs. By using the CV to scale the data, the influence of "unstable" variables with high CVs is reduced in the chemometrics analysis. This makes it more advantageous to use in metabolomics data when auto-scaling is required as the influence of noise and compounds that were not repeatable detected are reduced in the multivariate statistics (Keun *et al.,* 2003; Scholz & Selbig, 2007). This scaling method is even more powerful when used in a supervised way. Supervised VAST scaling can be performed when the average of the CVs of each variable within an experimental group is used to scale the variable instead of the 'total' CV across all samples (Keun *et al.,* 2003; van der Berg *et al*., 2006).

A.2.4.2.5 Log transformation

The main function of data transformation is to remove heteroscedasticity and reduces high intensity compounds more than the lower intensity compounds (Dettmer *et al*., 2007). Log transformation down-weighs outliers and ensures a more Gaussian-type frequency distribution (Fiehn & Kind, 2007). Since transformation reduces the impact of the larger values more than the smaller values, transformation has a pseudo scaling effect (van der Berg *et al*., 2006). Log transformation is, along with Pareto scaling, one of the most used data pre-treatment methods and a great asset for metabolomics data (Jonsson *et al*., 2005). The procedure of log transformation is as follows: after the log of each variable in each sample has been acquired the data is centred before multivariate statistics. There have also been some deviations from this order of transformation as some researchers first centre the data before log transformation (Steuer *et al*., 2007; Scholz & Selbig, 2007). This change in order is done so that zero values are replaced with actual values before log transformation as log 0 = ∞ (Idborg-Bjorkman *et al*., 2003). It is also common to add an infinitesimal constant or unity to all intensity values such as 1 (Jonsson *et al*., 2005). Since log 1 = 0, the impact of the added constant to the data does not create false signals for absent compounds. One disadvantage of using log transformation is that the standard deviation of the lower abundant compounds is inflated (Sumner *et al*., 2007; van der Berg *et al*., 2006). Due

to some of the limitations in the use of log transformation, the glog (generalized log) transformation method was introduced. It takes advantage of the fact that the variance does not depend on mean values at low-signal intensities. Thus, more precise cluster analysis can be performed for signals close to the limit of detection (Purohit *et al*., 2004).

A.2.4.2.6 Power and Fourth root transformation

Power transformation functions very similarly to log transformation by removing heteroscedasticity, but has fewer problems than the log transformation method. This method makes use of the square root instead of the log of each variable and is also followed by centering of the data before chemometric analyses. Since the high intensity compounds are reduced more than the lower intensity compounds, this method also has a pseudo scaling effect. Like Pareto scaling, this pre-treatment method does not make the data completely dimensionless and allows the more abundant compounds to remain dominant in the statistics. A similar type of transformation commonly used by researchers is called 'fourth root transformation' which fits basically between power and log transformation. As the name implies, this method makes use of the fourth root (twice the square root) followed by centering of the data (Idborg *et al*., 2005b). In many cases, this method can be more powerful than log transformation, especially for metabolomics data as it generally gives satisfactory reduction of the high intensity compounds without inflating the unstable lower intensity compounds to overshadow the results.

A.2.4.2.7 Data pre-treatment combinations in the literature

A quick overview of the literature will indicate the diverse combinations (Craig *et al*., 2006) and choices of data pre-treatment methods. Jonsson *et al*. (2005) log transformed their data after a value of one was added to each variable in the matrix and was followed by mean centering and auto-scaling. Morgenthal *et al*. (2007) divided each variable by the median followed by log transformation. By doing this, the influence of potential outliers is reduced. Choi *et al*. (2006) only centred their NMR data before chemometric analysis and did not scale it at all. Before PCA, Llorach *et al*. (2009) log transformed, mean centred and Pareto scaled their data while it was only log transformed and mean centred before PLS-DA, orthogonal signal correction (OSC)-PLS and OSC-PLS-DA. Van den Berg *et al*. (2006) used four data pre-treatment methods separately on the data and then ranked the VIPs according to their overall result in the four different methods.

A.2.5 STATISTICAL ANALYSIS OF THE DATA USING CHEMOMETRICS AND UNIVARIATE METHODS

After the data has been cleaned, normalised and pre-treated, it is finally fit for the final step which is data analysis. In order to get the relevant information from the complex data matrices, the data has to be simplified and this is where chemometrics plays a key role. The term chemometrics refers to the multivariate statistical analysis of (bio)-chemical data (Goodacre, 2006; Wishart, 2007) and is in most cases nothing more than a dimensionality (data) reduction technique that reduces the number of variables while keeping most of the variance/information (Scholz & Selbig, 2007). Chemometric methods not only help visualise the data but also give the important metabolites (IMs) that are responsible for grouping. Therefore, it has two basic functions which are; to reduce the dimensions of the complex data sets; and to show (dis)similarities between experimental or biological different groups/samples.

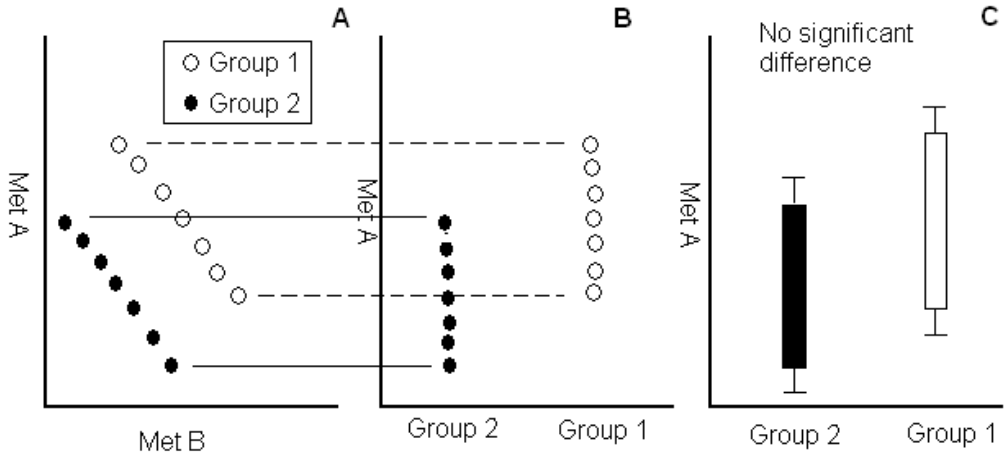


Figure A.7: Multivariate vs. univariate statistics in grouping and validation. There is in reality no difference between group 1 and group 2 when the levels of metabolite A are solitarily compared between them (Fig A.7B – C), as there is a huge overlap in the metabolite levels. In other words, if the groups were unknown and only the levels of metabolite A (or B) compared with univariate tests, it would be impossible to differentiate between samples from group 1 and group 2. However, when the relationship of metabolite A and B are used (such as in multivariate tests), it becomes possible to differentiate between the two groups.

Not only has chemometrics become a valuable tool in all omics-type studies, but also in numerous targeted studies where only a few variables (metabolites) are measured. But why use chemometrics on targeted data when you can still use univariate statistics on each detected variable? The answer to this question lies in the term ‘coherent perception’ which is a psychological term used in, for instance, the recognition of human faces. When only one variable is used (like the eyes) it would be hard, sometimes impossible, to discriminate between faces, while

the use of numerous single variables add up to a holistic perception of the shape (Weckwerth & Morgenthal, 2005). This is also true in metabolomics especially as one metabolite can influence the levels of another directly or indirectly, giving a compounding effect (which is better captured by multivariate statistics as shown in Figure A.7). Similar to the example of face recognition, it is sometimes hard or impossible to differentiate between experimental/biological groups when the levels of only one metabolite are measured or univariately compared as indicated in Figure A.7. In many cases it is the bio-signature (covariance of several compounds) that is important in the discrimination of samples. For this reason, it is also necessary to keep in mind that univariate statistics cannot always be used to validate multivariate findings as many researchers did in the past. However, the use of univariate tests as additional, secondary data mining methods can be useful (if used to complement the multivariate findings and not validate it).

A.2.5.1 Univariate statistics

Univariate statistical methods are commonly used in biological research and range from the simple comparison of group means and standard deviations to somewhat more advanced methods such as t-tests and analysis of variance (ANOVA). Student's t-test and ANOVA are especially useful in biological research for indicating whether the values (means) of two or more experimental groups differ significantly or not. However, the (sole) use of these statistical methods and the limitation of the researcher's input have led to numerous irrelevant results. In most cases it is the statistical significance that matters even if there is no biological significance such as when the blood pressure of a hundred volunteers was lowered by 5 mmHg after taking an experimental drug. The statistics would indicate a significant difference but will have no real biological meaning. It is because of this that another univariate test, *Effect size*, has been introduced to support the findings of the t-tests and ANOVAs. As its name implies, the 'size of the effect' (of the drug), or difference between groups, are determined with this test. It therefore helps to interpret whether a statistical difference is also *practical* (Ellis & Steyn, 2003). The combination of these tests can thus be valuable in determining significant and practical biomarkers or IMs in metabolomics research.

A.2.5.2 Chemometrics techniques

There are numerous chemometrics techniques that can be used to analyse the large data sets, each with a different approach and final result. As with all the other steps in metabolomics, the choice of chemometric method depends on the data at hand and the type of answers that are desired (Dettmer *et al*., 2007). For example, if one merely desires to group the samples to show their similarities, it might be sufficient to use clustering techniques, while principal component analysis (PCA) can be used when the variables that give group separation are also important

(Goodacre, 2006). In most cases, the final result of a metabolomics experiment is not to cluster similar samples to find a sample class but to find metabolites that are different between them. For this reason a certain chemometric method is often a stepping stone for another (supervised method) and it is common to use more than one chemometric method.

All chemometrics techniques can be separated into two groups, i.e. supervised and unsupervised. Unsupervised methods are most commonly used and do not require any additional information besides the clean data. The best known unsupervised method is PCA. Another unsupervised method that is especially useful to cluster similar samples is hierarchical cluster analysis (HCA) which groups samples according to their similarity in distance. Several distance measures can be used of which the Euclidean distance, Mahalanobis distance or correlation is mostly used. With HCA the result is usually visualised as a dendogram or a tree (Goodacre, 2006; Sumner *et al*., 2007). The supervised methods require additional training information (such as group identifications) for 'supervising' the data processing and are much more powerful than unsupervised methods. Supervised methods are especially useful when the focus is to discover important compounds that differ between experimental or biological groups (Dettmer *et al*., 2007; Jonsson *et al*., 2005; Scholz & Selbig, 2007; Sumner *et al*., 2007). Projection to Latent Structures (or partial least squares) Discriminant Analysis (PLS-DA) is one of the best known supervised methods. Another is Discriminant Function Analysis (DFA) which is mathematically similar to MANOVA but different in terms of emphasis. With MANOVA the aim is to find whether groups are significantly different on the linear combinations of the measured metabolites, while DFA aims to predict group membership and also why the groups are different (Goodacre, 2006; Goodacre, 2008; Steuer *et al*., 2007).

A.2.5.2.1 Principal Component Analysis (PCA)

As mentioned earlier, PCA is perhaps one of the best known and most used chemometrics techniques (Sumner *et al*., 2007). The main objective of PCA is to reduce the dimensionality of the data while retaining as much information (variance) as possible (Wishart, 2007). The first principal component (PC) is thus the linear combination of the original variables that explains the greatest amount of variation, whereby each successive PC explains the remaining variance in the data set (Llorach *et al*., 2009; Scholz & Selbig, 2007; Steinfath *et al*., 2008). In other words, PCA summarises the variation in the data in a smaller number of latent components (Madsen *et al*., 2010). These new components are then used to visualise the samples in this new space via a scatter plot as illustrated in Figure A.8. Similar samples would cluster together while those that are different would not. Most often it is required to determine the variables that played a key role in the

clustering which means that the coefficients can be used to visualise the loadings indicated in Figure A.8B (Ivosev *et al.,* 2008; Ramadan *et al.,* 2006; Steuer *et al.,* 2007).

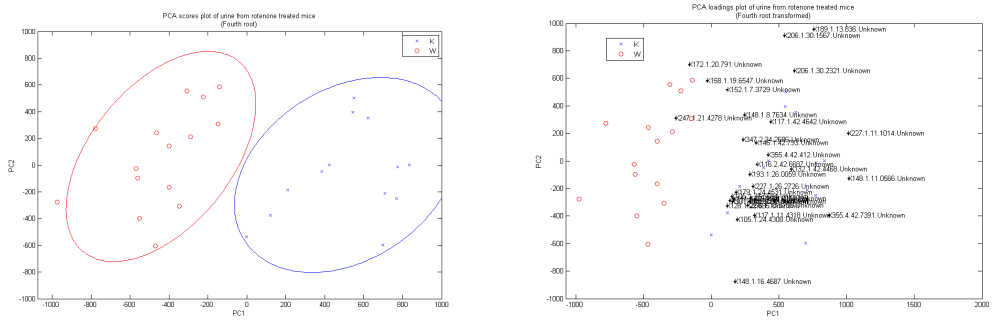


Figure A.8: PCA scores plot (A) and loadings plot (B). The loadings plot indicated here is in fact a bi-plot of the scores and loadings (variables) which simplify interpretation. The direction of the loadings (from 0) indicates their weight in the groups. In this example, most of the loadings are situated on the right side of the 0 mark alongside with the blue sample scores. This means that these compounds (loadings) were more abundant in these samples in comparison to the samples in the left. Or, in other words, it means that the samples on the left (red) contained lower levels of these compounds (Steuer *et al.,* 2007).

Although PCA is frequently used to reveal and visualise disjoined groups in the data, it does not perform any actual partitioning of the points into distinct sets. Also, it frequently fails to give optimal projections of a data set. Thus a clustering method still needs to be used. “Strictly speaking, PCA is rather a data visualisation tool that projects most variable components of a multidimensional data on a 2D or 3D plot, which can be comprehended by the human mind…” (Dettmer *et al*., 2007). In addition, there is also no accurate threshold value for defining clusters and one can only be guided by visual interpretation (Fiehn, 2002). PCA is always recommended as a starting point for analysing multivariate data and will rapidly provide an overview of the information hidden in the data. But as there are always irrelevant and technical variance in the data, PCA cannot be the only chemometric technique used (Issaq *et al.,* 2009; Scholz & Selbig, 2007; Trygg *et al.,* 2007). It is therefore recommended that the data be re-analysed with supervised methods as it is more powerful (Want *et al*., 2007).

A.2.5.2.2 Independent Component Analysis (ICA)

While PCA is a data reduction technique, ICA is rather a feature extraction technique (Scholz & Selbig, 2007). ICA is similar to PCA but is less strict to the choice of components. ICA assumes that the measured metabolite concentrations are statistically independent. This means that the values of one component provide no information about the values of the other components (which

is different from PCA). Thus the new components are not orthogonal to each other (Scholz *et al.,* 2004). With this it is possible to detect instrument drift as a separate component which PCA cannot do. As a result, ICA gives more biologically reasonable grouping in comparison to clustering with PCA (Scholz *et al.,* 2004; Scholz & Selbig, 2007; Weckwerth & Morgenthal, 2005). Since ICA does not reduce the multivariate data into fewer latent structures, a data reduction method (such as PCA) has to be incorporated when this technique is used on metabolomics data (Morgenthal *et al*., 2007; Steinfath *et al*., 2008).

A.2.5.2.3 Partial least squares (projections to latent structures) – discriminant analysis (PLS-DA)

Unlike the unsupervised PCA method, PLS-DA maximises the covariance between the predicting data set, which is X of the numerical value of metabolites, and Y of the class assignment (Woo *et al*., 2009). The objective with discriminant analysis is to find a model that separates classes of observations on the basis of the values of their X-variables. Provided that each class is "tight" and occupies a small and separate volume in X space, one can find a plane – a discriminant plane – in which the projected observations are well separated according to class. In other words, PLS-DA is performed in order to sharpen the separation between groups of observations, by rotating PCA components such that a maximum separation among classes is obtained, and to understand which variables carry the class separating information (Want *et al*., 2007). So, instead of asking “what is there” as in PCA, one rather asks, “what is the relation to or difference between”. Knowledge of PLS and PLS-DA has increased substantially lately and we know now that these methods are very much affected by systematic (technical) variation in X that is not related to Y. This variation among samples results in a higher number of PLS components which complicates model interpretation. The use of *a priori* information bears the risk that ‘noise’ in the measurements is overestimated, especially in the case of a low number of samples or replicates (Weckwerth & Morgenthal, 2005).

A.3 ANNEXURE SUMMARY

Recent advances in the field of metabolomics and chemometrics, as discussed in this annexure, hold great promise for elucidating the relationship of MTs and the mitochondrion. However, as noted and discussed in detail, this technology has many choices, limitations and risks that could influence the result and/or hamper our understanding of the relationship of MTs with the mitochondrion. The purpose of this literature study was thus not only to review recent advances in the field but also to clarify and support the selection and standardization of approaches, methods and conditions that are given in the following annexures.

ANNEXURE B

METABOLOMICS: METHODOLOGY CONSIDERATIONS, DEVELOPMENT AND STANDARDISATION

B.1 INTRODUCTION

Although metabolic profiling has been around for some time, the field of metabolomics is fairly new and requires *standardisation*[$] and *validation*[#] of selected methods (Koek *et al.,* 2010). Since this is the first comprehensive metabolomics study at our institution (Centre for Human Metabonomics, NWU), this annexure will give every detail in method development, procedure standardisation and validation.

$ The context of *standardisation* in this study must be understood in terms of establishing a standard operating procedure which is not necessarily ideal or up-to-date with the methodology of the field. The standard operating procedure is used for all samples to ensure repeatable results for comparison of the experimental groups.

Validation in this study refers to the process of checking and confirming that a selected method satisfies certain criteria (such as metabolome coverage, repeatability, etc.) in comparison to other methods or approaches tested. The selected method is not validated using external sources or data.

B.2 STANDARDISING THE EXTRACTION AND LC-MS ANALYSIS OF POLAR METABOLITES

B.2.1 LC-MS ANALYSIS OF POLAR METABOLITES

As mentioned in Chapter 3, a reverse phase LC-Q-TOF system was selected for untargeted analyses of polar metabolites which mainly belong to the secondary metabolism. It was also decided prior to this study to only focus on the positive ions. This allows for superior sensitivity while still covering 'enough' of the metabolome (Want *et al.,* 2010). The organic acids and free fatty acids that are mainly detected in negative mode (Dettmer *et al.,* 2007), are analyzed with the silylation-GC-MS platform. Recommended and default data acquisition settings from Agilent for metabolomics work was used throughout this study. Instrument specifications and the data acquisition (chromatographic, ionization and scan) settings are described in Chapter 4.

B.2.2. SELECTION OF INTERNAL AND EXTERNAL STANDARDS

The newly developed NOMIS (Sysi-Aho *et al.,* 2007) and CCMN (Redestig *et al.,* 2009) normalisation methods mainly correct for analytical variance, such as ionization suppression and other (retention) time-related occurrences. Hence, it would appear that it is adequate to include standards that are spread across the chromatogram. In light of this, it was decided to use more than one standard per analysis in this study. This would firstly allow non-linear chromatographic alignment (Sumner *et al.,* 2007; Steinfath *et al.,* 2008) when necessary and would also allow CCMN normalisation which is an improved version of NOMIS normalisation (Redestig *et al.,* 2009). When there is not a constant shift of retention times, non-linear alignment is then performed (using regression algorithms). Literature on the normalisation methods are given in Annexure A while Annexure C will discuss their uses in more detail.

Due to the cost of stable isotopes, it was decided to use cheaper, more available standards for this purpose. The potential standards had to fulfil a few important prerequisites which were: (1) it should be pure, (2) be water soluble (for good extraction efficiency), (3) have no ionization problems, (4) should not fragment with the selected ionization settings, (5) should not be part of the metabolome being investigated and (6) should preferably not elute with many other metabolites as it may cause ionization suppression or enhancement. It was also important that the standards of choice had a linear response to the instrument in the concentration range it will be used, regardless of the slope by which the signal dropped with decreasing concentrations. However, since the response of the standards alone will differ quite a lot when analyzed within samples (due to ionization suppression), this response is more of a guideline rather than a calibration curve. By adding standards before and after sample preparation, the recovery can be checked and corrected for (Morgethal *et al.,* 2007). Therefore it was decided to use both internal (standards added before sample preparation) and external standards (added just before analysis) (Koek *et al.,* 2010). The external standard help to monitor the analytical instrument's response more precisely as the internal standards can technically not be used for this purpose. Variation in the abundances of the internal standards is mainly due to sample preparation while variance in the external standard abundances is mainly due to instrument response.

B.2.2.1 Materials and Method

The following compounds were purchased and analyzed in order to find suitable standards for chromatographic alignment and data normalisation:

Table B.1: List of compounds tested for the selection of standards.

Compound	Supplier	Catalogue #	Purity
Allopurinol	Sigma	A-8003	unknown
Acetaminophen (paracetamol)	Sigma-Aldrich	A7085	~99%
Caffeine	Fluka	44818	≥ 99%
Norleucine	Fluka	74560	≥99%
Norvaline	Sigma	N7627	unknown
3-hydroxynorvaline	Sigma	H4002	≥98%
6-aminocaproic acid	Calbiochem	1381	>98%
N,N-dimethyl-L-phenylalanine (DMPA)	Aldrich	27391-0	99%
Valproic acid	Sigma	P4543	≥98%

These compounds were dissolved in Burdick & Jackson water (cat # 364-4) to attain a 1 mM solution. Each compound solution was analyzed separately on a LC-Q-TOF (settings Section B.2.1) for the first screening. Representative liver, muscle, plasma and urine samples from MT1+2KO mice were also prepared (without any standards) and analyzed separately. This was done to confirm the absence of the candidate standards in these mice as well as any other compound with similar mass and retention time. In addition, these samples were also used to find favourable elution points where the chromatogram does not contain many (large) co-eluting compounds.

The second screening tested the extraction efficiencies (and water solubility) of the candidate standards. As mentioned, only the polar, water soluble metabolites were of interest for the LC-MS platform. After sample preparation (protein precipitation or metabolite extraction), the metabolites are normally dried to get rid of solvents and re-dissolved in water. It was therefore critical for the standard of choice not to be lost in the apolar (organic) phase and to re-dissolve effectively in water. Equal volume of the 1 mM standard solutions was mixed together followed by the metabolite extraction method described in Section B.2.3. The top polar phase were dried and re-suspended in the correct volume of water to ensure that the compound concentrations were similar as before.

B.2.2.2 Results and Discussion

Figure B.1A shows the overlaid chromatograms of all the compounds tested (except allopurinol). Although the same amount of each compound was analyzed, the respective peaks differ in height and width. Peaks belonging to 6-aminocaproic acid, norleucine, DMPA and caffeine were slightly

higher while the DMPA peak was also the widest, eluting over two minutes. Figure B.1B shows the overlaid chromatograms of the representative mice samples.

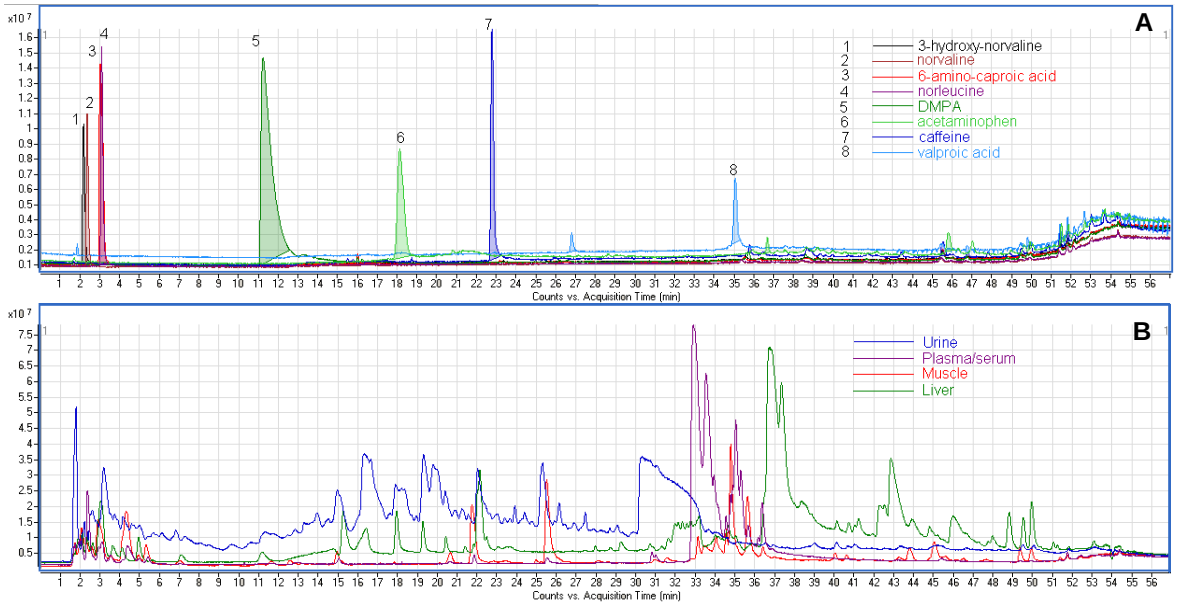


Figure B.1: Overlaid LC-MS chromatograms of all the tested standards (A) and representative mice samples (B). The different standards and samples are shown in different colours as indicated in the legend. The peaks representing each standard are also numbered according to elution order.

Since the same amount of each compound was analyzed, the larger peaks indicate that caffeine, DMPA, 6-aminocaproic acid and norleucine have somewhat better ionization capabilities. Despite this difference, all tested compounds showed a linear response with altering concentrations (results not shown). Furthermore, on inspecting the mass spectra of each candidate peak, it was clear that none of the candidates showed fragmentation (results not shown). The chromatograms of each candidate, except valproic acid, were also clean of significant contaminating peaks indicating that they were of sufficient purity. Of all the tested compounds, DMPA was the only one which did not elute very effectively from the selected column giving a wide peak. By comparing the compound elution patterns of the standards and the tissue samples (Figure B.1B) it seems that DMPA, acetaminophen and caffeine elute in relatively ‘clear’ regions of the chromatogram. Norvaline, norleucine, 6-aminocaproic acid and 3-hydroxynorvaline elute quite early and possibly along numerous other small, very polar molecules, such as amino acids. From these four early eluting compounds, norleucine elutes last, in a more favourable area of the chromatogram.

Not only was it important that these compounds were not naturally part of the metabolome being studied, it was also important that it had a unique mass as to limit problems later on. Since

compounds with similar mass often has similar structures and hydrophilicity (especially the lower masses), one can assume that such compounds would easily co-elute making their discrimination difficult. A database search was performed to compare the masses of the standards to known metabolites. The following standards were found to have the same monoisotopic mass as some natural metabolites in the mice tissue samples (Table B.2).

Table B.2: Mass similarities of standards and naturally abundant metabolites.

Standards	**Monoisotopic mass**	**Natural abundant metabolite**	**Monoisotopic mass**
6-aminocaproic acid	131.094635	Leucine	131.094635
Allopurinol	136.038513	Hypoxanthine	136.038513
Norleucine*	131.094635	Leucine	131.094635
Norvaline*	117.078979	Valine	117.078979

*The similarities in mass were *a priori* known

It was decided to remove 3-hydroxynorvaline, 6-aminocaproic acid and allopurinol from selection due to the discovered mass similarities and other factors (such as ionization and elution time). Interestingly, it was possible to resolve norleucine (131.2066) and norvaline (117.185) from naturally abundant leucine (131.0929) and valine (117.0776) (results not shown). Whether this result was due to the software bin and alignment settings were not clear. These compounds were not expected to differ in mass but was hoped to differ in retention time. Hence, it was decided to keep these candidates for further investigation. The resolution of norleucine and norvaline from the naturally abundant amino acids was observed when they were analyzed along with the pure standards of leucine and valine.

The extraction efficiencies of the standards were also compared and it was found that all the candidates except valproic acid and caffeine had high extraction efficiencies. After the extraction procedure, the valproic acid and caffeine peaks were reduced or completely absent which meant that these compounds were lost in the organic phase (results not shown). Norleucine and acetaminophen were finally selected as internal standards. Caffeine was selected as external standard as it showed good ionization under the conditions used. The resolution of norleucine from leucine was repeatedly seen when the data was extracted with MassHunter and was thus included as standard. It was reasoned that, though their resolution in biological samples may be limited which would hinder its use for normalisation, it could be valuable for non-linear chromatographic alignment. By using standards in the beginning, middle and near-end of the chromatogram, non-linear RT alignment would be more reliable.

B.2.3 METABOLITE EXTRACTION FROM TISSUE SAMPLES FOR METABOLIC FINGERPRINTING

Irrespective of the analytical technique used, sample preparation will remain the step that is most prone to errors and technical variation. Technical variance and bias has been identified as one of the most important threats to validity in biomarker research (Teahan *et al.,* 2006). The introduction of technical variance in the sample preparation step means that it is 'hardwired' and cannot be eradicated by analytical or statistical approaches, no matter how advanced (Teahan *et al.,* 2006). Hence the importance that each sample is prepared identical is obvious and so the need to standardise all steps. It is also important to quench the metabolism of the tissue samples quickly after harvesting. Snap-freezing in liquid nitrogen is a commonly used quenching method and was the preferred quenching method in this study.

Since metabolomics involves the analysis of many samples, the goal was to develop and use a simple and high-throughput method which gives the highest extraction efficiency and *repeatability*[§] for as many classes of compounds as possible (Gullberg *et al.,* 2004). A metabolite extraction method for NMR and MS-based metabolomics was recently published (Wu *et al.,* 2008) which is based on the Bligh & Dyer (1959) method, using methanol, water and chloroform to extract metabolites from tissue samples. This method was the primary choice for metabolite extraction but had to be tested and validated on these mice samples for repeatability. An important part of sample preparation that is often overlooked is the homogenization step, which can introduce a lot of technical variance. It is critical to ensure that all samples are broken and homogenized to the same extent so that the same amount of metabolites is released. Therefore, the use of semi-automated homogenization systems such as vibration mills is mandatory for metabolomics research. This meant that this semi-automated system also had to be standardized for each tissue type in order to give repeatable homogenization.

§ *Repeatability* in context of this study has the same meaning as precision. A repeatable method would give low technical variance, allowing replicate samples to give exactly the same result. Repeatability is different from the term reproducibility which refers to a method's ability to give similar results when repeated after a specific period, in another laboratory or when performed by another person.

According to Fiehn (2002) and t'Kindt *et al.* (2009) it might be advantageous for metabolomics approaches to extract metabolites from frozen tissues that contain the original amount of water in comparison to extracting freeze-dried samples. The main reason for this is because freeze-drying

may potentially lead to the irreversible adsorption of metabolites on cell walls and membranes. In contrast, others promote the use of freeze-drying before homogenization as the variable water content in tissue samples can cause lower extraction repeatability and enhanced degradation of the metabolites (t'Kindt *et al.,* 2009). However, since the selected method needs to be high-throughput, freeze-drying of tissue samples were not tested.

The temperature at which extraction of metabolites from tissue samples is done also differ among reports. Some suggest that higher temperatures (such as 70 ℃) give a higher metabolite yield while limiting enzyme activity (Gulberg *et al.,* 2004; t'Kindt *et al.,* 2009). Unfortunately, it was also shown that certain metabolites are lost at such high temperature. In contrast, other researchers prefer metabolite extraction at -25 ℃ by working in acetone containing dry ice (Yang *et al.,* 2008). Although enzyme activity and metabolite loss is limited at this temperature, the yield is also low. Others propose working at room temperature when using methanol to extract metabolites as methanol 'freezes' enzymatic activity (Kanani *et al.,* 2008). It is therefore also necessary to test and validate several of these extraction temperatures along with the biochemists' favourite of 0 – 4 ℃.

B.2.3.1 Materials and Methods

The proposed two step method from Wu *et al.* (2008) was validated with similar conditions as described in the paper. The exact procedure is described in Chapter 4. Soft tissue (such as liver) was homogenized for 2 minutes at 20 Hz while tougher tissue (such as muscle) was homogenized for 2 minutes at 30 Hz. For the LC-MS method the upper water (polar) phase was transferred to another tube and dried. The dried compounds were re-dissolved in pure water, containing 30 µg/ml caffeine (as external standard). A few factors were varied for method validation purposes. Only one factor was changed at any time. Homogenization was also validated in terms of the volume sample per volume beads and volume solvent as well as the vibration intensity and duration. The tested conditions were validated in terms of their repeatability and metabolome coverage (or preservation). This was done by splitting a single tissue sample into several pieces so that each of the tested conditions contained five to ten replicates.

The factors that were varied for validation of the extraction method are the following:
i) extraction temperature (70 ℃ vs. -25 ℃ vs. 0 ° C);
ii) replacement of chloroform with acetonitrile (ACN) or acetone;
iii) addition of chloroform to sample before and after homogenization;
iv) best sample wet weight range.

The amount of detected features (metabolome coverage), their intensities, method repeatability and the time it took were guidelines that were used to judge these altered conditions.

B.2.3.2 Results and Discussion

As it is difficult to monitor repeatability of homogenization without metabolite extraction, the "cumulative" results in Figure B.2 is an indication of the repeatability of the complete method. Furthermore, it is difficult to measure repeatability statistically, especially for multivariate data, as there is no clear and acceptable yardstick as such. One way of visualizing the repeatability of a method is to look at the total variance via PCA. After PCA, several distance measures can be used to identify potential outliers in the data which can give an indication of the method repeatability. However, ultimately the repeatability of the method is perhaps better shown in the data plot (also included in Figure B.2). The average coefficient of variance (CV) of all the detected compounds is also indicative of the repeatability of the method. However, this measure was not used here.

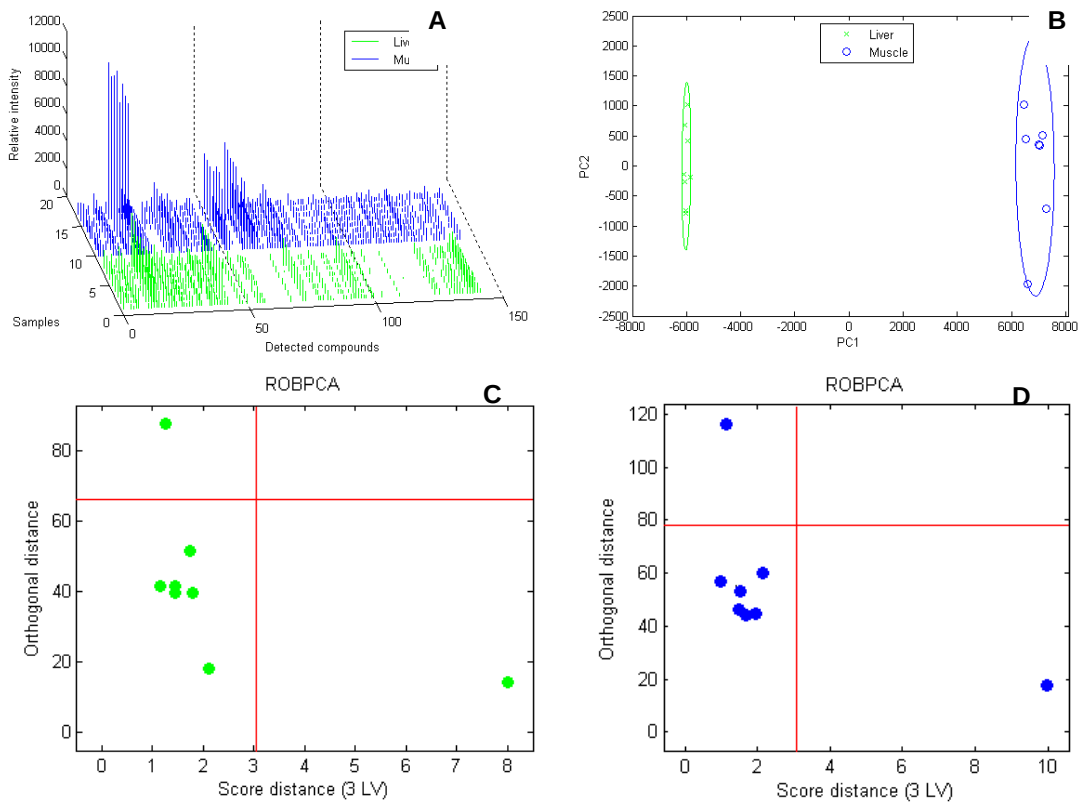


Figure B.2: Data and score plots indicating the repeatable metabolite extraction of liver (soft) and muscle (hard) tissue samples. Data plot A and the complementary scatter plots (B, C and D) shows the repeatability of both methods. The PCA score plot of both tissues after VAST scaling (B) shows some intra-group variance on the second PC axis. An orthogonal distance vs. Mahalanobis distance plot (C & D) on three latent variables (three PCs) with cut-offs (lines) shows outliers in the data.

All the detected compounds (also called features) are shown in the data plot (Figure B.2A). There is little difference between the replicates of similar tissue samples while the differences between the liver and muscle are obvious. Although sample preparation appear to be repeatable when looking at the data plot, it is difficult to really visualize the repeatability with the naked eye. Hence, the PCA score plots are also included which captures all the detected compounds (variance) into a single dot on the scatter plot. Samples that are metabolically very similar will thus be grouped in close vicinity (in this new two-dimensional space). In the PCA scores plot, the liver and muscle samples group separately from each other as should be expected while replicate samples group together. The intra-group variance of the liver and muscle samples (PC 2) are also very similar and was confirmed by the included orthogonal - and Mahalanobis distance plots. These distance-distance plots reveal possible outliers in the data when situated outside the cut-off lines. The data indicate that the method proposed by Wu *et al.* (2008) and selected homogenization settings are reliable and highly repeatable as little variance are introduced to the replicates. Also, in comparison to the other tested methods (Figure B.3) this method had sufficient coverage of the sample metabolome.

Several extraction temperatures were tested as mentioned in the previous section. Replacement of chloroform with different organic solvents was also tested. Figure B.3 shows only a few results from this validation experiment. Again, when looking at the data plot, the similarities between replicates are obvious in the detected compounds. Some differences are seen in the (amount of) the detected compounds of differently prepared samples albeit not as prominent like when compared to another tissue samples. Hence, one immediate conclusion that can be made is that the modification of extraction temperature and organic solvent did not have a great impact on the results. The distance-distance plot also shows somewhat similar repeatability between the tested conditions. However, the repeatability of the selected method (black dots) was superior in terms of intra-group variance as shown in the PCA scores plot and distance-distance plot.

The results obtained were in accordance to what others concluded in the literature. The appearance of certain features (compounds) was seen when extraction was performed at room temperature or 70 ℃ in comparison to 0 and -25 ℃. However, when comparing the amount of features between these temperatures, there was actually no significant difference. On closer look, the main difference that was noticed was that certain metabolites were lost (or reduced) with the higher extraction temperature while others were gained (or increased in intensity). Similar results were seen for the lower temperatures (data not shown). Therefore, it was a choice between compromises and after evaluating these results it was concluded that extraction at 0 – 4 ℃ would

be more suitable for this study as less effort and costs are involved when working at this temperature. In addition, this temperature ensures limited enzyme activity while still giving relatively good yield in comparison to the colder temperatures.

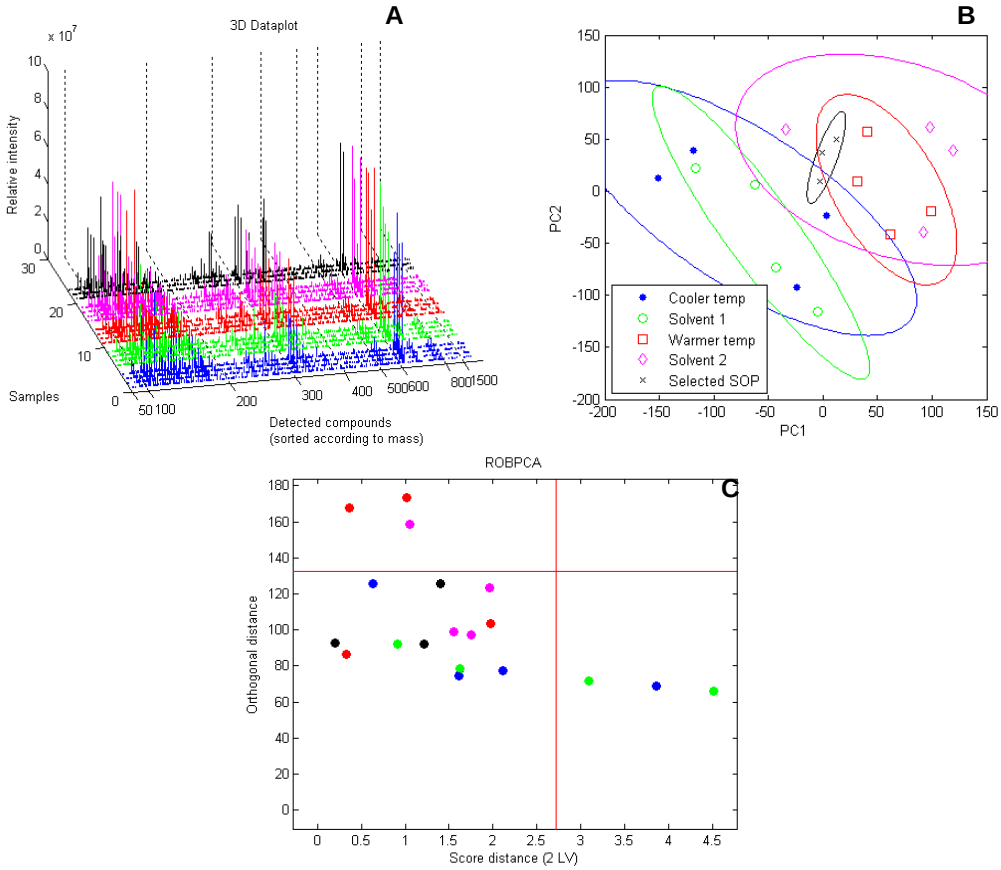


Figure B.3: Data and score plots indicating the metabolite extraction at different temperatures and different organic solvents. Data plot A and the complementary scatter plots (B and C) shows the sample metabolome coverage and repeatability, respectively. The PCA scores plot after VAST scaling (B) illustrate differences between the preparation methods. An orthogonal distance vs Mahalanobis distance plot (C) on two latent variables (two PCs) with cut-offs (lines) shows outliers in the data. Not all the tested extraction temperatures and solvents are shown.

The replacement of chloroform with acetone or acetonitrile also gave some differences in the data plots. As with the extraction temperatures, the amount of features, their intensities and the absence/presence of certain metabolites between the dataplots were witnessed. Since it could not be predicted which metabolites were important for this study, in order to preserve them, there was no possible reason to replace the chloroform with another solvent. Also, given that the water phase

is of importance for LC-MS fingerprinting and not the organic phase, the influence of the organic solvent on certain metabolites were seen as a minor factor. In contrast, the sequence of addition of the organic solvent to the sample had enormous effects on the data, especially the yield (data not shown). When the chloroform was added to the sample before homogenization, the metabolite yield was much lower in comparison to when it was added after homogenization. The reason for this is not clear but it was noticed that the tissue sample became 'stiff' (probably due to protein and lipid precipitation) and that homogenization of the sample became difficult.

The optimal sample weight range was experimentally confirmed to be between 5 – 50 mg (data not shown) as reported by Wu *et al.* (2008). This sample wet weight and associated solvent volumes ensure that there is enough empty space in the microcentrifuge tube for optimal collision energy during homogenization with the vibration mill. When the tissue sample was >50 mg, it was not completely homogenized as the increased volume of solvent (relative to the sample weight) hindered the steel beads from colliding with optimal energy.

A few other factors of the sample preparation step were also seen to influence the results and were standardised. Firstly, it was seen that autoclaved material, although sterile, were chemically contaminated (Erban *et al.,* 2007). Also, as mentioned by others, the use of plastic tubes when working with chloroform resulted in the release of plasticizers. As the efficiency of these plasticizers to ionize is very high, this could cause ionization suppression (Wu *et al.,* 2008). It is thus recommended to use glass vials where possible especially after the homogenization step. However, due to problems associated with most glass vials (like availability or centrifugation constraints) it was decided to use pre-treated polypropylene tubes. Overnight soaking of these tubes in methanol containing some chloroform reduced the leaching of plasticizers into the sample.

B.2.4. DEPROTEINIZATION OF PLASMA / SERUM SAMPLES FOR METABOLIC FOOTPRINTING

Unlike the tissue/cell metabolome, the exometabolome does not need to be extracted from cells. However, blood contains many proteins which mean that the exometabolome have to be separated from the proteins and peptides. The main prerequisite for a protein removal method is that it must remove as much of the protein as possible while preserving as much of the exometabolome as possible. Therefore, the selection of an optimal protocol will rely on the exometabolome coverage and small amount of interfering proteins it gives at the end. The two most commonly used and recommended deproteinization methods in the literature were tested to find the more optimal approach. The criteria that the preferred method had to satisfy were that it

had to be quick and easy (high-throughput), repeatable, cost effective and able to remove as much proteins as possible without interfering with the exometabolome. The first method that was tested was ultra-filtration with a 3 kDa cut-off spin column. The second method was acetonitrile protein precipitation with a few variations. Daykin *et al.,* (2002) used a 1:1 (v:v) plasma to acetonitrile approach while most other researchers use a 1:2 or 1:3 (v:v) ratio (Williams *et al.,* 2006). Since there are some inconsistencies about the ratio of plasma to acetonitrile, it was decided to test these different options as well. Due to the concentrated nature of rodent bio-fluids, the method from Williams *et al.* (2006) was especially of interest as they too used rodent plasma.

B.2.4.1 Materials and Method

A single pooled plasma sample containing 45 µg/ml norleucine and acetaminophen was aliquot into several tubes so that each method had five tubes of 100 µl plasma (Fiehn & Kind, 2007). Ultra-filtration was performed with 3 kDa cut-off Nanosep® Omega™ centrifugal devices (Pall Corporation, cat # OD003C34) according to the manufacturer's specifications. Acetonitrile deproteinization was done by adding 100 µl, 200 µl and 300 µl ice cold acetonitrile to the tubes to test the 1:1, 1:2 and 1:3 (v:v) ratios, respectively. After the addition of acetonitrile, the samples were centrifuged at 5000 x *g*, 4 ℃ for 10 minutes and the supernatants transfer red to new tubes. The supernatants were dried to remove the acetonitrile and re-dissolved in water containing 30 µg/ml caffeine before analysis.

B.2.4.2 Results and Discussion

Virtually no difference was seen in the data plots of the 1:1, 1:2 and 1:3 (plasma:ACN) ratio samples (results not shown) despite the differences in precipitate and raw chromatograms. The 1:3 ratio samples contained relative more precipitate in comparison to the other tested samples. Furthermore, its chromatogram showed less protein/peptide peaks, distinguished by their numerous multi-charged peaks. Hence, it was concluded that the 1:3 (v:v) acetonitrile deproteinization method gave better protein removal in comparison to the other tested ratios without hindering the exometabolome coverage (as the data plots were comparable). In contrast to the above, there was a huge difference between ultracentrifugation and the selected 1:3 ratio ACN precipitation method as shown in Figure B.4A

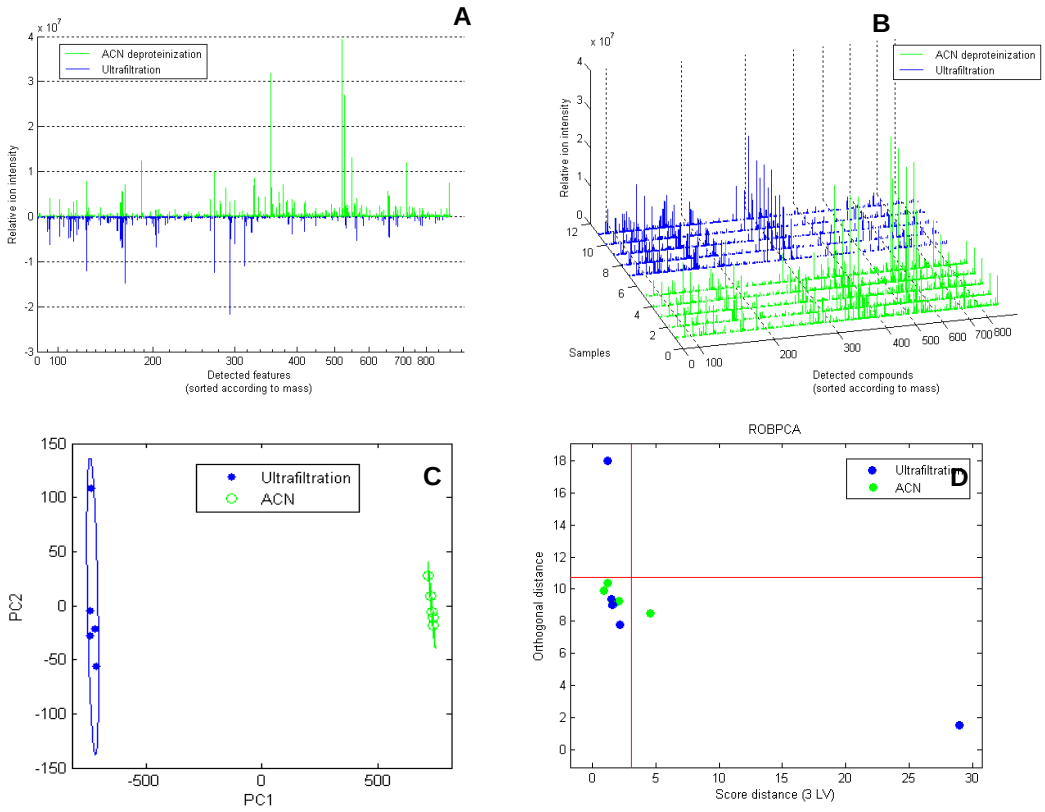


Figure B.4: Data and score plots of plasma samples deproteinized with acetonitrile and ultra-filtration. The mirror data plot (A) highlight the differences in metabolite recovery of both methods. Data plot B and the complementary scatter plots (C and D) show the repeatability of both methods. The PCA scores plot of both methods after VAST scaling (C) shows some intra-group variance on the second PC axis. A robust distance vs Mahalanobis distance plot (D) on three latent variables (three PCs) with cut-off lines shows outliers in the data. The detected compounds in the dataplots are sorted according to mass.

There is a clear difference in metabolome coverage between the acetonitrile method and ultra-filtration as illustrated in the data plots. These differences are better visualized in the mirror data plot (Fig B.4A). Most of the detected compounds from the ultra-filtration method had a mass lower than 300 Da while most of the detected compounds from the ACN deproteinization method are larger than 300 Da. Both methods showed good repeatability (Fig B.4B) with the ACN method having slightly less intra-group variance when looking at the PCA scores plot after VAST scaling (Fig B.4C). Since PC 1 explains the most variance in the data, it obviously shows the difference between both deproteinization methods (inter-group variance). However, PC 2 shows the more relevant information which is the variation between the samples in each group (intra-group variance). This slightly better repeatability of the ACN method is also seen in the distance-distance plot (Fig B.4D). Upon inspecting the chromatograms it was clear that ultra-filtration removed

virtually all proteins (data not shown). The ACN precipitation method did not give the same amount of protein and peptide removal but it seems that it gave a much wider preservation of the exometabolome. Hence, due to the lower exometabolome coverage, the higher costs of the spin-columns and their relatively lower repeatability, it was decided to use acetonitrile as the preferred method for protein precipitation. A complete description of the standardised procedure is given in Chapter 4.

B.2.5 PREPARATION OF URINE SAMPLES FOR METABOLIC FOOTPRINTING

B.2.5.1 Materials and Method

Two different urine preparation methods were compared to a urine sample which did not underwent any preparation. A single urine sample containing 45 µg/ml norleucine and acetaminophen was divided into several tubes so that each of the three methods had five replicates. For the first method, the five urine samples were only dried and re-dissolved in an equal amount of water containing 30 µg/ml caffeine. For the second method, the urine samples were added three volumes acetonitrile to precipitate large bio-molecules that might be present. After centrifugation at 5000 x *g*, 4 ℃ for 10 minutes, the supernatants were dried and re-dissolved in water containing 30 µg/ml caffeine. For the third method the urine samples were prepared in a similar way to that of the second method except that the precipitate was washed with methanol and water to ensure that no metabolites were left behind. The five replicates were dried and also re-dissolved in water containing caffeine. Exometabolome coverage, repeatability and speed were the criteria which were used to make a final decision.

B.2.5.2 Results and Discussion

Figure B.5 shows a mirror data plot of the detected compounds from the urine sample prepared with and without acetonitrile. Both data plots are exactly the same and true mirror images of each other. Except for speed, no other difference in repeatability (not shown) or metabolome coverage was observed between these methods as illustrated in Figure B.5. Even though a clear precipitate was visible in the ACN treated samples, these apparent proteins (or lipids) had no impact (gave no observable ionization problems) in the untreated sample as the metabolite levels are similar in both plots. Since all multi-charged species are excluded from the data extraction algorithm, the presence of proteins is not shown in Figure B.5 but only detected metabolites. In light of these results it seems that the use of urine without preparation is adequate and high-throughput. This method was, however, not selected for this study where mice samples were used. The ACN deproteinization method was selected as any small amount of proteins present in the mouse urine

(Bell, 1932; Pasikanti *et al.,* 2008) could easily block the guard column when a large batch of samples is analyzed.

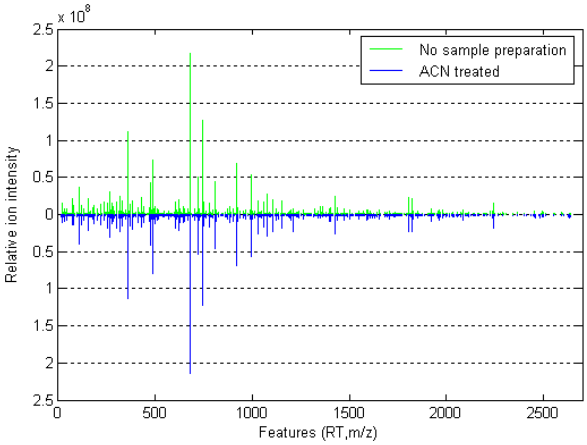


Figure B.5: Comparative (mirror) data plot of acetonitrile treated and untreated urine samples. The data plots only show detected metabolites and not proteins, as multiple charged species are excluded from the data extraction algorithm.

B.3 STANDARDISING THE EXTRACTION, TRIMETHYLSILYLATION AND GC-MS ANALYSIS OF PRIMARY METABOLITES

B.3.1 GC-MS ANALYSIS OF TRIMETHYLSILYL ESTERS

The instrumental setup used for silylation-GC-MS analyses of the derivatized samples is described in Chapter 4. These settings were used throughout this study when analyzing trimethylsilyl esters. Due to the untargeted nature of this approach, the temperature gradient was not further optimized for optimum separation of targeted compounds.

B.3.2 SELECTION OF INTERNAL AND EXTERNAL STANDARDS

As with the LC-MS approach, it was decided to use multiple standards per sample for this study despite the fact that many researchers still prefer a single standard to monitor extraction and derivatization (Boudonk *et al.,* 2009). This decision was supported by the fact that it allows non-linear chromatographic alignment (Sumner *et al.,* 2007) and advanced normalisation (Redestig *et al.,* 2009). It was also decided to use cheaper, more available standards for this purpose instead of stable isotopes. Since it was decided to perform a single extraction for all three the analytical platforms (due to limited sample volumes), the internal standards of choice had to be detectable. Unfortunately, it was *a priori* known that norleucine and acetaminophen would not make good

standards as it contained more than one active site that can be silylated in the form of nitrogen. Nitrogen-containing compounds often have problems in repeatable derivatization which meant that additional standards had to be selected for GC-MS analyses.

The important prerequisites for the standards were: (1) it should be pure, (2) contain one functional group with active hydrogen for derivatization, (3) should be easily derivatized, (4) should not be part of the metabolome being investigated and (5) should preferably not elute along other metabolites to make peak deconvolution and identification difficult. Again, it was also important that the standards of choice had a linear response to the instrument, regardless of the slope by which the signal dropped with decreasing concentrations. As with the LC-MS platform, it was also decided to use both internal and external standards. In contrast to the internal standards, the external standard of choice had to lack any derivatizable functional group, be thermally stable and volatile, and be easily detected. It also had to be pure, not be part of the metabolome being investigated and not co-elute with (many) other compounds.

B.3.2.1 Materials and Methods

The following compounds, each containing a single active hydrogen, were purchased and analyzed in order to find suitable standards for chromatographic alignment and data normalisation:

Table B.3: List of compounds tested for the selection of standards.

Compound	Supplier	Catalogue #	Purity
Internal standard candidates:			
3-phenylbutyric acid	Aldrich	11680-7	98%
Nonadecanoic acid (C_{19} fatty acid)	Sigma	N5252	≥99%
Heneicosanoic acid (C_{21} fatty acid)	Fluka	H5149	~99%
Heptacosanoic acid (C_{27} fatty acid).	Sigma	H6514	~99%
External standard candidates:			
Eicosane (C_{20})	Fluka	44818	≥99.8%
Heneicosane (C_{21})	Aldrich	286052	98%

These compounds were dissolved in hexane (Sigma-Aldrich, cat # 52766), methanol or chloroform (depending on their solubility) to attain 1 mM solutions. They were mixed in equal volumes, dried under nitrogen and then derivatized (Section B.3.5). Hexane was added to each sample after which it was analyzed for the first screening. Representative liver, muscle, plasma and urine samples from MT1+2KO mice were also prepared (as described in the subsequent sections) and analyzed separately. This was done to confirm the absence of the candidate standards in these

mice as well as any other compound with similar retention time and mass spectrum. In addition, these samples were also used to find favourable elution points where the chromatogram did not contain many (large) co-eluting compounds. Since both the polar and apolar phases of the extraction procedure are used for GC-MS analyses, it was not necessary to test the extraction efficiencies of these compounds. Moreover, it was expected that most of these compounds would actually prefer the apolar phase which is beneficial for the LC-MS work. Since only the polar phase was used for LC-MS, these compounds would be excluded from the run.

B.3.2.2 Results and Discussion

Figure B.6A shows the chromatogram of all the compounds tested. Although the same amount of each compound was prepared, the respective peaks differ in height and width. Peaks belonging to 3-phenylbutyric acid and nonadecanoic acid were slightly higher. As expected, the width of all the peaks was small (as none of them contained nitrogen, phosphates or sulphur groups). The two tested hydrocarbons, eicosane and heneicosane, eluted just before nonadecanoic acid. Figure B.6B shows the overlaid chromatograms of the representative mice samples. None of the tested compounds were found in the tissue samples indicating these compounds to be naturally absent in these mice.

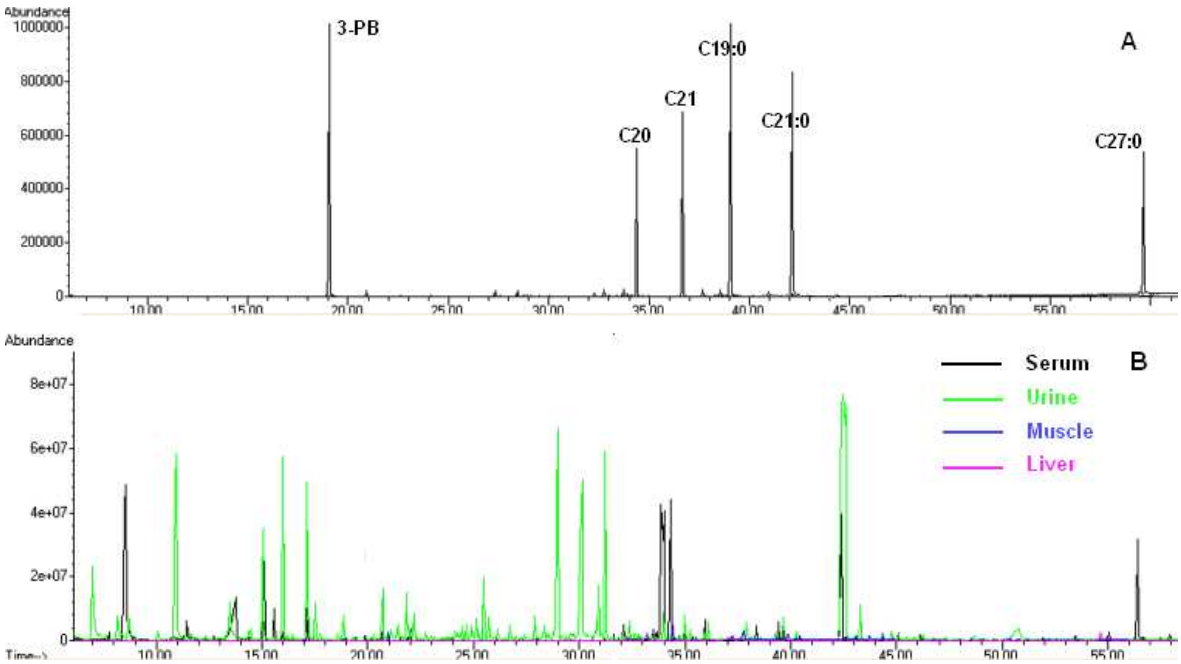


Figure B.6: Chromatograms of all the tested standards and mice tissue/bio-fluid samples. The peaks representing the standards are indicated in the top chromatogram while the various representative samples are shown in different colours in the lower chromatogram. 3-PB, 3-phenylbutyric acid; C19:0, nonadecanoic acid; C21:0, heneicosanoic acid; C27:0, heptacosanoic acid; C20, Eicosane; C21, heneicosane.

Comparing the peak intensities of the standards it is obvious that all candidates were derivatized more or less to the same extent. However, 3-phenylbutyric acid and nonadecanoic acid gave relatively higher peaks which made them the obvious choice for internal standards. Also, since 3-phenylbutyric acid is still commonly used as internal standard in organic acid analyses (Reinecke *et al.,* 2011), it was selected alongside nonadecanoic acid. Both standards also had favourable elution times ensuring that non-linear chromatographic alignment would be possible. Eicosane were selected as the external standard of choice as it eluted more favourable than heneicosane, which eluted too close to nonadecanoic acid.

B.3.3 METABOLITE EXTRACTION FROM TISSUE SAMPLES FOR METABOLIC FINGERPRINTING

The selected extraction procedure from Wu *et al.* (2008) was also considered for the GC-MS work. It is advantageous to have a single sample preparation procedure for both GC-MS and LC-MS as it saves time and sample (t'Kint *et al.,* 2009). To get sufficient coverage of the primary metabolome, it is necessary to use both the polar and apolar phases (Gullberg *et al.,* 2004). This overlap would ensure that compounds that are lipophilic and capable of derivatization would also be included with the readily derivatized hydrophilic compounds. For this reason, Gullberg *et al.* (2004) also used a single phase metabolite extraction method in which the ratios of methanol and chloroform were adapted. Hence, in order to validate the selected method for the GC-MS platforms, it was decided to compare the Wu *et al.* (2008) method to the organic solvent ratios used by Gullberg *et al.* (2004). For this comparison the metabolome coverage, in terms of detected compounds, was the main determinant followed by the repeatability of the method.

B.3.3.1 Materials and Methods

The metabolite extraction protocol from Wu *et al.* (2008) was used as described in Chapter 4. The ratios of the added methanol and chloroform were adapted to test the monophasic extraction as described by Gullberg *et al.* (2004). Figure B.7 summarizes the validation experiment that was performed. The polar (group 2 as indicated in the illustration) and apolar phase (group 4) from the biphasic extraction was used separately as controls. The sum of the components obtained from these controls was used as a guideline to show whether the single (group 1) and mixture of two phases (group 3) gave satisfactory coverage. All detected components were identified (using the NIST 08 library) to validate them as relevant metabolites. All peaks arising from contamination and derivatization artefacts were discarded. Since the influence of temperatures and alternative organic solvents were already tested with the LC-MS work, it was not tested again for this extraction process. This experiment was down with a single tissue sample which was divided into several

aliquots so that each test had five replicates. After metabolite extraction, all samples were dried under nitrogen and derivatized as described in Section B.3.6.

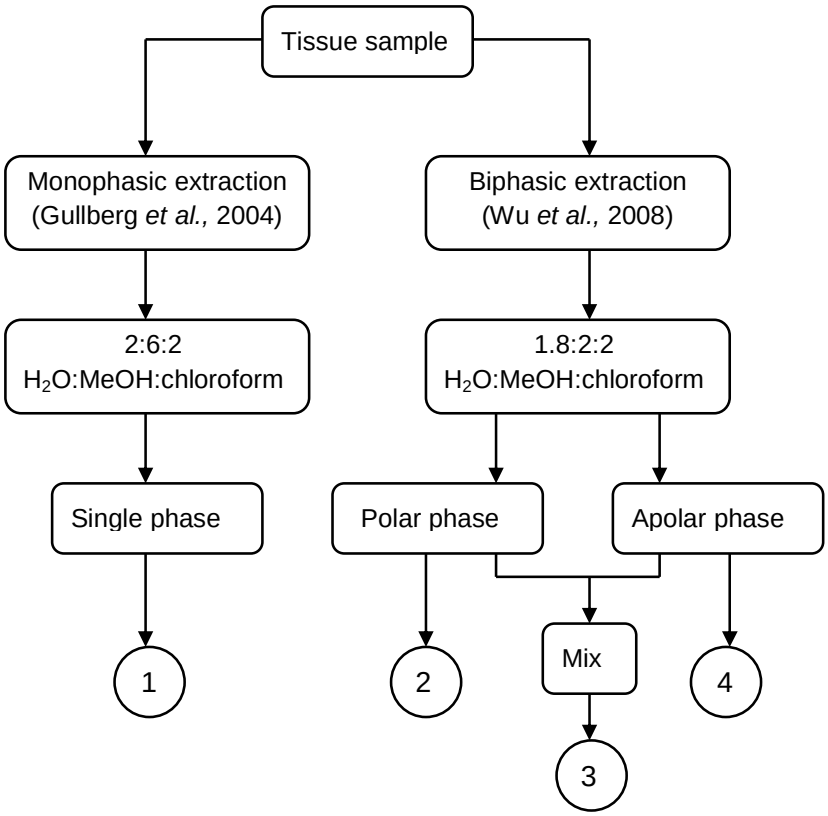


Figure B.7: Experimental strategy to study the difference in monophasic and biphasic metabolite extraction for silylation and GC-MS analyses.

B.3.3.2 Results and Discussion

Table B.4 summarizes the results obtained in terms of metabolome coverage. The average and standard deviation (SD) of all the detected metabolites is given per batch (which consisted of five replicates). The results indicate that the monophasic extraction method (on average) only gave a few components more in comparison to the biphasic extraction method. There was not a single compound that was consistently present in the monophasic method but absent in the biphasic method. Moreover, in some samples the biphasic extraction method produced more compounds in comparison to the monophasic extraction method, while the latter only gave more coverage on average (Table B.4). Upon comparing the repeatability of both methods, there was again no real difference (data not shown). Hence, the selected biphasic extraction method remained the primary choice. The sum of the compounds detected separately in the water and organic phases were clearly higher to that detected in the single phase or when both phases were mixed. The main reason for this might be due to the greater volume used to allow similar end volumes before

derivatization. Where the combined sample contained only 50 µl of the water and the organic phases, the separate samples of the respective phases contained 100 µl.

Table B.4: Metabolome coverage of the tested monophasic and biphasic metabolite extraction procedures.

Group	Extracted phase	Average number of components (mean ± SD), n = 5
1	Single phase containing both polar and apolar metabolites	114 ± 9.8
2	Water phase containing polar metabolites	81 ± 12.1
3	Both water phase and organic phase	109 ± 7.5
4	Organic phase containing apolar metabolites	105 ± 10.2

B.3.4 DEPROTEINIZATION OF PLASMA / SERUM SAMPLES FOR METABOLIC FOOTPRINTING

Due to the limited volume of mice plasma and serum available to this study, it was decided to use the same deproteinized plasma and serum used for LC-MS analyses in the silylation procedure (for GC-MS analyses). Hence, after deproteinization three aliquots of these samples were taken for each analytical platform. For GC-MS footprinting, the deproteinized sample was then dried under nitrogen, followed by silylation (Section B.3.6).

B.3.5 PREPARATION OF URINE SAMPLES FOR METABOLIC FOOTPRINTING

Organic acid analyses are one of the oldest and most used metabolic profiling methods for GC-MS. This method is still actively used at this institute for research and diagnostic purposes (Reinecke *et al.,* 2011). However, whether this method can be adopted in the field of (untargeted) metabolomics has raised a lot of debates recently as it is a semi-targeted approach. Due to the controversy over the scope of organic acid analyses in comparison to the analysis of merely deproteinized urine, it was necessary to compare these two methods. Except for the differences in speed, metabolome coverage and repeatability, one of the main differences between these two approaches is the amount of moisture that is left in the samples. It is well known that moisture influence derivatization negatively resulting in the incomplete and variable derivatization of compounds. It is believed that dried urine (and tissue samples for that matter) still contains some residual water which is the main reason why crude sample preparation methods are not commonly preferred and their validity also questioned. Therefore, the degree by which moisture influence the

derivatization process was also investigated (Annexure D) in order to validate the results obtained from crude preparation methods.

B.3.5.1 Materials and Methods

A pooled mice urine sample with known creatinine concentration was used for all tests. The creatinine concentration of the urine sample was determined with the QuantiChrom™ Creatinine Assay kit (DICT-500) from BioAssay according to the manufacturer's specifications. Figure B.8 summarizes the different approaches that were taken to compare the preparation methods, as well as to validate selected parameters. Five 100 µl aliquots of the urine sample with creatinine concentration of 35 mg% were prepared with each method. Internal standards were added in a 10 µg IS per mg% creatinine ratio. The volume of urine used depended on the amount that was available and not on the creatinine value as the original protocol stated (especially since the mice's urine showed ten times lower creatinine values in comparison to human urine).

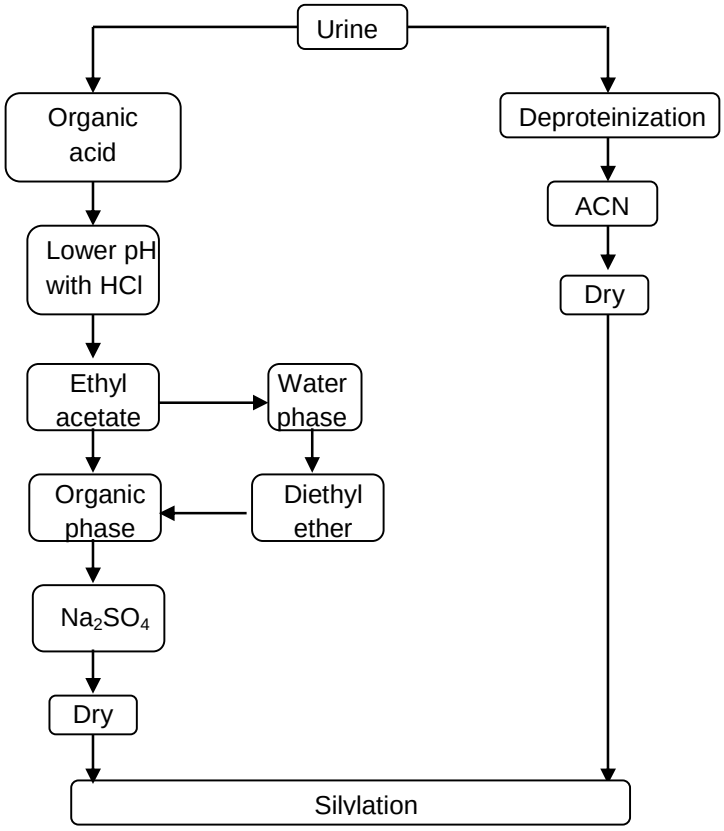


Figure B.8: Experimental strategy to investigate different urine preparation methods for silylation - GC-MS. The main focus is the difference between organic acid extraction and urine that was only deproteinated.

To extract organic acids from the urine, the pH was first lowered below 2 by the addition of a few drops of 5 N HCl (Saarchem, cat # 306 30 44). Six ml ethyl acetate (Merck, cat # 1.09623. 2500) was added and the samples mixed for 30 minutes. The mixtures were centrifuged for 10 minutes at 2000 x *g* (room temperature) after which the top organic phases were put in new tubes. The remaining water phases were again mixed with 3 ml diethyl ether (Merck, cat # 1. 00921.2500) for 15 minutes. The mixtures were again centrifuged and the top organic phase transferred to the tube containing the first organic phase. Sodium sulphate (Merck, cat # 1.06649.0500) was added to the organic phases to remove any residual water. The samples were centrifuged and the liquid transferred to another tube followed by drying under nitrogen. The dried samples were then derivatized and analyzed. The second, more untargeted method made use of only deproteinized urine. Similar to the LC-MS metabolic fingerprinting approach, the urine was also deproteinized with acetonitrile and then dried before silylation. Exometabolome coverage, repeatability and speed were the criteria which were used to evaluate the methods.

B.3.5.2 Results and Discussion

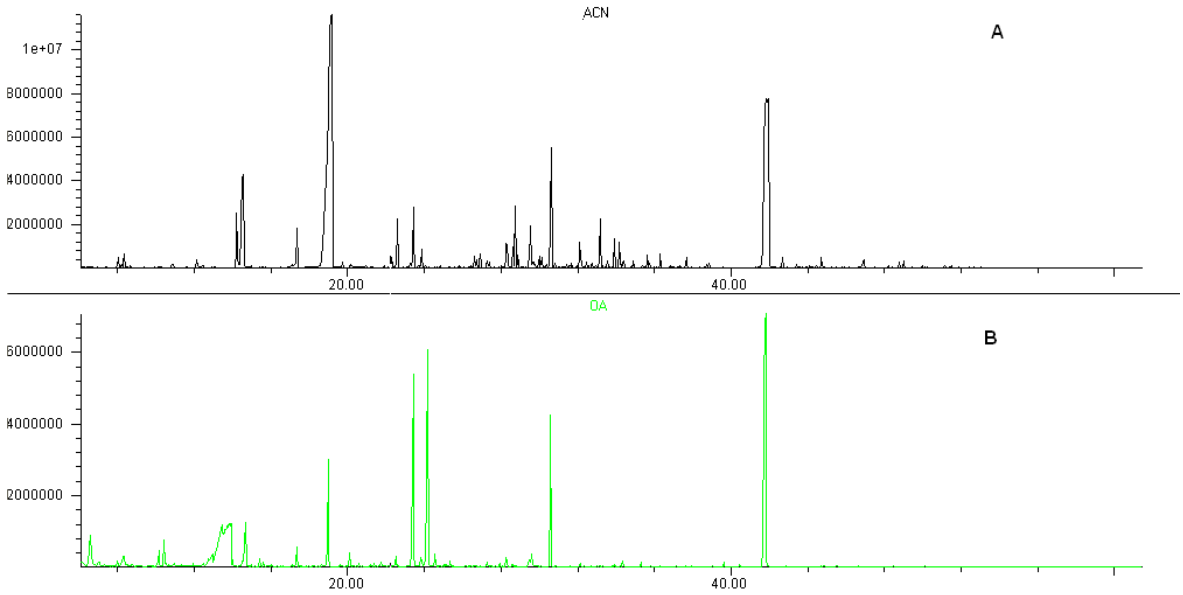


Figure B.9: Representative chromatograms of mice urine prepared with acetonitrile (A) and organic acid extraction (B). The chromatograms are scaled to the largest peak, therefore note the scale difference.

Figure B.9 shows two representative chromatograms of urine samples deproteinized with acetonitrile (A) and the other which was prepared by organic acid extraction (B). There are clear differences between these chromatograms. The most obvious difference is the lower peak intensities in the lower chromatogram. It is also obvious that the top chromatogram contain more

compound peaks. The repeatability of both methods were very similar (data not shown) but with the organic acid extraction method showing somewhat better repeatability.

One of the reasons why it is advantageous to have limited sample preparation in metabolomics experiments, apart from the bias toward certain compounds, is that the more steps there are in sample manipulation, the more sample loss can occur. This effect might explain the intensity differences of similar peaks in the two chromatograms, especially the internal standards. Although the reduction in internal standards compensate for the lower metabolite levels, some of the less abundant metabolites are totally absent (or miss calculated) when their concentrations are below the detection limit. Another obvious result visible in Figure B.9 is that more peaks were detected when urine was only deproteinized in comparison to organic acid extraction. There are perhaps two reasons for this phenomenon. Firstly, not all the metabolites in the urine were extracted with the organic acid extraction method, thereby limiting the scope, while the other method retained all compounds except for large lipids and proteins. The second reason for the fewer peaks might be due to the lower peak intensities.

In conclusion, it is evident that acetonitrile deproteinization of urine would be a better choice in terms of exometabolome coverage. However, it was ultimately decided to use organic acid extraction for metabolic profiling of the urine samples in this study. The main reason for this decision is because of the fact that the exometabolome does not consist of that many metabolites when compared to the intracellular-metabolome, especially since most compounds in urine are water soluble. Seeing that many of the polar compounds would also be detected by the complementary LC-MS platform, reasonable coverage of the exometabolome was expected despite the semi-targeted nature of this method. Lastly, the slightly better repeatability and this method's well recorded and widely accepted use gave credence to this decision.

B.3.6 SILYLATION AND OXIMATION OF METABOLITES TO GIVE TRIMETHYLSILYL ESTERS

Oximation and silylation protocols are well described for metabolomics investigations albeit most of those are used for plant-related studies. Since the amount and diversity of metabolites from plants and mice differ substantially, it is important to standardise and validate conditions specifically for mice samples. Silylation conditions had to compliment the work being done and had to be constant for all samples being investigated in order to limit technical variance.

Several oximation conditions have been described in the literature of which oximation for 17 hours at room temperature and 60 minutes at 60 ℃ are the most commonly used (Jonsson *et al.,* 2004).

These conditions were compared and one selected for this study. Seeing that the solubility of analytes is greatly influenced by the amount of methoxyamine (MOX) salt and pyridine (Gullberg *et al.,* 2004), it was decided not to evaluate the effect of different MOX concentrations. After a (quick) review of the literature, it is apparent that there are basically two types of silylation conditions commonly used. The first method requires silylation at room temperature for 60 minutes (Jonsson *et al.,* 2004) and the second silylation for 30 minutes at higher temperature (such as 37 ℃). These conditions were also tested and one selected for this study. After consulting the literature, the sample to BSTFA ratios was selected before testing so that BSTFA would always be in excess for "complete" derivatization. As the influence of moisture on derivatization was already investigated (Annexure D) the water released during the oximation step (Figure A.3) were seen as insignificant and no further testing was done on this matter.

Another aspect that needed some attention was whether the process of derivatization had to be stopped or not (Fiehn, 2008). One way of stopping the derivatization reaction is to add water which deteriorates the active derivatization agents in the mixture without hindering the derivatives (Li & Park, 2001). However, whether the derivatives are truly preserved or not is questionable (Koek *et al.,* 2010). Even though this is true or not, one thing that hinders this approach is that the added water has to be removed again before analysis. In the tested approach, the added water was removed from the derivatized samples by carefully drying them under nitrogen, followed by the addition of sodium sulphate and the re-extraction of the derivatives. Although this approach requires an additional drying step, one could argue that it would be better to just evaporate the derivatization agents instead of incorporating several laborious steps. Therefore, this option was also explored as shown in Figure B.10.

B.3.6.1 Materials and Methods

A single tissue sample was used for this experiment. Homogenization and metabolite extraction from this single sample was done as described earlier. The water and organic phases were divided so that each of the five tests had five replicates containing 50 µl (each) of the organic - and water phase. These samples were dried under nitrogen. As illustrated in Figure B.10, the main conditions selected for this study (the solid line, group 5) was validated by varying specific factors (conditions). Fifty µl of 20 mg/ml MOX (Aldrich, cat # 226904) in pyridine (Merck, cat # 1.09728.0100) was added to the dry samples which were then kept at room temperature for 17 hours. This was followed by the addition of 50 µl BSTFA (Supelco, cat # 3-3027) containing 1 % trimethylchlorosilane (TMCS, Fluka, cat # 92360). The samples were kept at room temperature for 60 minutes after which 50 µl hexane (containing 30 µg/ml eicosane) was added.

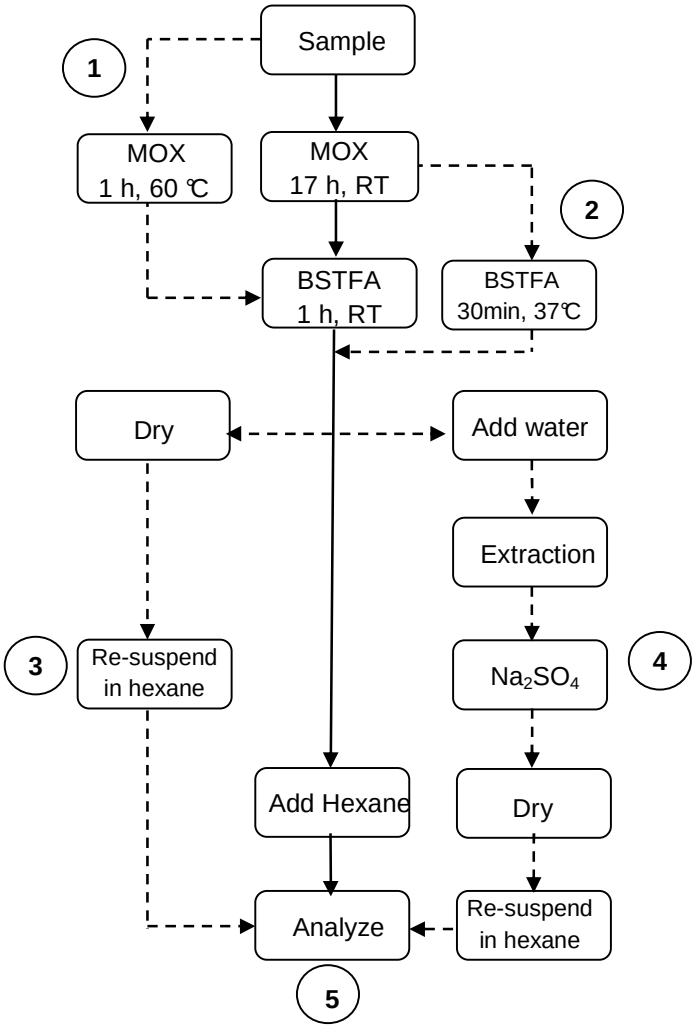


Figure B.10: Experimental procedure to validate selected oximation and silylation conditions. The solid lines (middle) indicate the method of choice (test group 5). Perforated lines indicate the factors that were changed during testing. The oximation conditions were changed in group1 while other silylation conditions were tested for group 2. After silylation group 3 were dried and re-dissolved in hexane. The silylation reaction in test group 4 was stopped with the addition of water followed by the re-extraction of derivatives. RT, room temperature.

For each test one factor was varied. The first group were oximated for 60 minutes at 60 ℃. For the second test group, silylation was performed at 37 °C for 30 minutes. The third group was dried after silylation, in order to stop derivatization, and re-dissolved in hexane containing the external standard. The silylated samples of the fourth group were given 50 µl water. After this the derivates were extracted with a methanol and chloroform mixture. Sodium sulphate was added to ensure that all residual water were removed followed by drying of the extract and re-dissolving it in hexane containing 30 µg/ml eicosane. For all tested procedures the metabolome coverage (or rather

preservation of it), repeatability, ease and speed of the method were the main points for consideration.

B.3.6.2 Results and Discussion

Figure B.11 shows five representative chromatograms from each test group. The bottom chromatogram shows the results of the selected conditions while the other chromatograms show the results when one factor was changed. Chromatogram 5 is also overlaid in the other chromatograms for easier comparison and is shown in green. There is truly no difference in the amount or intensities of detected compounds in chromatograms 1 and 5. Keto-acids and sugars such as pyruvate, glucose and galactose are unchanged in amount and abundance. The same similarities can be seen when comparing chromatogram 2 and 5. Except for some small differences in the detected compounds, there is virtually no other difference worth mentioning. The repeatability of both methods was also very similar (results not shown). However, chromatograms 3 and 4 stand out from the rest as they contain slightly less peaks. Most of the detected peaks have lower intensities indicating some sample loss. The repeatability of batch 3 and 5 were somewhat similar, while the repeatability of the fourth batch was worse (data not shown).

Oximation for 60 minutes at 60 ℃ showed no real difference in repeatability and metabolome coverage in comparison to oximation for 17 hours at room temperature. The same conclusion was made for the tested silylation conditions with no significant difference between silylation for 30 minutes at 37 ℃ or for 60 minutes at room temperature. Although the amount of detected compounds between these tested conditions were very similar, there were in fact some small differences in specific metabolite abundances and concentrations. Taking into account that the higher temperatures promote better solubility of certain compounds and rapid degradation of others, and *vice versa* for the lower temperature, it is again a case of compromises.

In contrast to the observations mentioned above, a huge difference between the profiles of the samples where derivatization was stopped and those that was not stopped was seen. The number of detected compounds was much lower in the batch that was dried before analysis. Although vigorous drying might be partly to blame, the main reason for this was because of the fact that many compounds are not again fully dissolved in hexane alone. The loss of compounds in the fourth batch was even worse and was probably a compounding result of the extra sample preparation steps and low solubility coefficients of the derivatives in hexane. In fact, it seems that most of the compounds were lost when water was added putting this method into question. Additionally, the repeatability of this method was worst of all the tested samples.

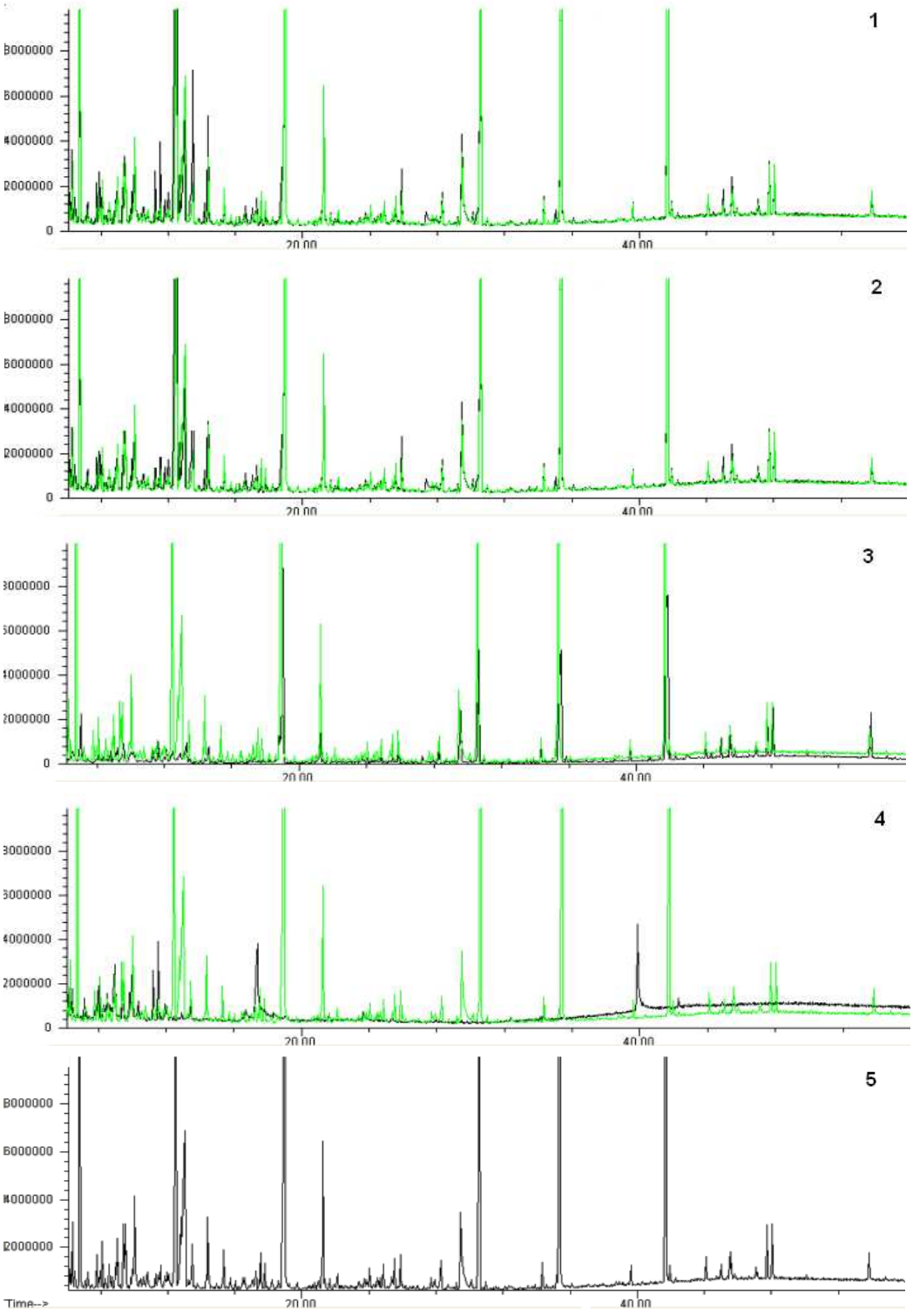


Figure B.11: Validation results of the selected oximation and silylation conditions. Chromatogram 5 (representing the results from condition 5 in previous figure) is overlaid in the other chromatograms for better comparison and is shown in green. The other conditions chromatograms are also shown in black.

Hence, it is obvious that the stopping of continuous silylation meant that metabolome loss will occur not to mention increased technical variance. In most cases, the selected method would give good results with wide metabolome coverage and also low technical variance. However, this approach might also be problematic with times as continuous silylation of specific compounds can incorporate great variance in the data. Fortunately, a few statistical approaches were explored which had the potential to correct for this time-dependent variance (Appendix C) using QC samples.

B.4 STANDARDISING THE EXTRACTION, METHYLATION AND GC-MS ANALYSIS OF LIPIDS AND FATTY ACIDS

The LC-MS and silylation–GC-MS strategies cover a large part of the polar metabolome, while apolar metabolites lack focus. Since MT1+2KO and MT3KO mice have a tendency toward obesity (Beattie *et al.,* 1998; Byun *et al.,* 2011), it was important to also include a process to monitor lipids. There were a number of approaches to consider but since lipidomics is a field on its own, it was not within the scope of this study to profile all the lipids. Instead, it was more sensible to rapidly fingerprint a reasonable part of the lipidome for comparison and differentiation of WT and MTKO mice. Although LC-MS are becoming the preferred analytical platform for lipidomic research, GC-MS still remains a powerful technique and the platform of choice for many researchers (Roberts *et al.,* 2008). Due to its lipidome coverage and ease of analyses it was decided to also use GC-MS for lipid profiling. Fatty acid methylation is the most commonly used derivatization method used to analyze fatty acids methyl esters (FAMEs) with GC-MS (Roberts *et al.,* 2008).

B.4.1 GC-MS ANALYSIS OF FAMES

The GC-MS system and setup that were used are described in Chapter 4. These settings were used throughout this study when analyzing FAMEs.

B.4.2 SELECTION OF INTERNAL AND EXTERNAL STANDARDS

Base-catalyzed methylation has the advantage that lipids do not have to be hydrolyzed to obtain the free fatty acids before preparing the esters. In fact, this method is incapable of esterifying free fatty acids and only methylate those fatty acids attached to glycerol (i.e. lipids). This “limitation” of the method is not a setback, as silylation covers the free fatty acids, but in light of internal standards this method would require a lipid standard to monitor methylation. This meant that the nonadecanoic acid added to the samples could not have been used as internal standard. However, due to this method’s good repeatability and somewhat targeted focus, it was decided to test and

use only an external standard for data normalisation after analysis. A few (odd-number) FAME standards were tested for this purpose. The choice of standard was mainly influenced by its elution time and ease of recognition.

B.4.2.1 Materials and Method

The following FAMEs were prepared and analyzed in order to find suitable standards for chromatographic alignment and data normalisation:

Table B.5: List of compounds tested for the selection of standards.

Compound	Supplier	Catalogue #	Purity
Methyl nonadecanoate (C_{19} FAME)	Fluka	74208	≥99.5%
Methyl heneicosanoate (C_{21} FAME)	Sigma	H3265	~99%
Methyl tricosanoate (C_{23} FAME)	Fluka	91478	≥99%
Methyl heptacosanoate (C_{27} FAME)	Sigma	H6639	≥99%

These compounds were dissolved in hexane to attain a single 1 mM solution containing all the standards. Representative liver, muscle, serum and adipose samples from MT1+2KO mice were also prepared (using the methods mentioned in subsequent sections) and analyzed separately. This was done to confirm the absence of the candidate standards in these mice as well as any other compound with similar mass spectrum and retention time. In addition, these samples were also used to find favourable elution points where the chromatogram does not contain many (large) co-eluting compounds.

B.4.2.2 Results and Discussion

Figure B.12 shows the chromatograms of the tested FAMEs and representative mice samples. When comparing both chromatograms (and library search reports) of the standard FAMEs and the representative tissue samples, the only odd-number FAME that were present in any of the mice samples was methyl-heneicosanoate. Methyl-nonadecanoate was absent in all the samples except the adipose tissue. Methyl-tricosanoate and methyl-heptadecanoate were absent in all tissue/bio-fluid samples. As mentioned the external standard of choice must be absent in the mice samples. In light of this, the choice was between methyl-tricosanoate and methyl-heptadecanoate. Of these two, methyl-tricosanoate eluted much earlier and was the better choice especially when used for chromatographic alignment. The use of only one standard meant that it was more advantageous that the standard be in the middle of the chromatogram if linear alignment (constant shift) is performed. Seeing that retention time shift augment with each subsequent eluting compound,

retention time correction with a late eluting standard will result in a total distortion (wrong alignment) of the early eluting compounds. Also, retention time alignment using an early eluting standard, will not give sufficient alignment for the late-eluting compounds.

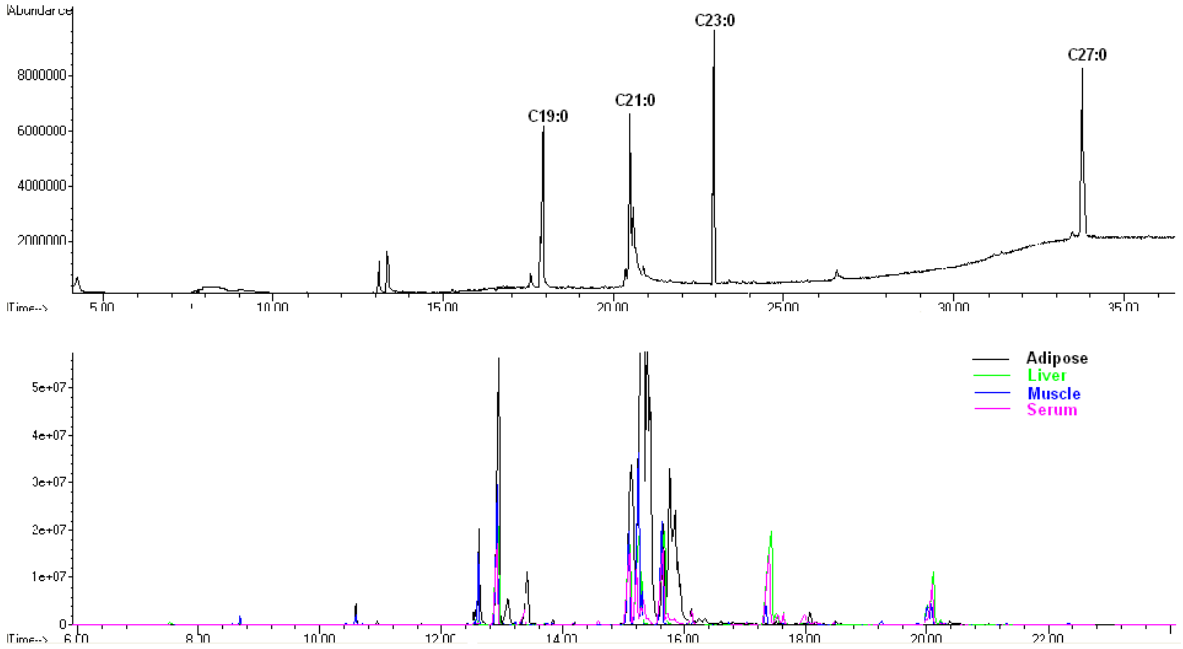


Figure B.12: Overlaid GC-MS chromatograms of all the tested standards (A) and representative mice samples (B). Peaks representing the standards are indicated while the chromatograms of the different samples are shown in different colours as shown in the legend. The time-axis of the chromatograms are not aligned. C19:0, methyl-nonadecanoate; C21:0, methyl-heneicosanoate; C23:0; methyl-tricosanoate; C27:0, methyl-heptacosanoate.

B.4.3 METABOLITE EXTRACTION FROM TISSUE SAMPLES FOR METABOLIC FINGERPRINTING

Although it is obvious that the lipid species would be extracted into the organic phase of the two phase extraction protocol, it was still necessary to test whether the use of both phases would give better results seeing that phospholipids or lipoproteins might be present in the water phase. Also, the use of the monophasic extraction protocol was also tested as was done in Section B.3.3.

B.4.3.1 Materials and Methods

The same materials and method was used as described in Section B.3.3. Methylation was performed according to the method described in Section B.4.6. Metabolome coverage was the

main determinant followed by the repeatability of the method. A single liver sample was used for both extraction protocols.

B.4.3.2 Results and Discussion

The results summarized in Table B.6 are similar to the silylation results, indicating no real difference between the monophasic and biphasic extraction methods. When covering the lipidome, one might expect that the monophasic extraction would perform worse than the biphasic method as it contains less chloroform, making it perhaps less soluble for lipophilic compounds. This was indeed the case with lipid-rich tissues that were prepared with the monophasic extraction method. One interesting observation is that the water phase also gave FAMEs, probably from phospholipids and similar species. Hence, the use of both phases of the biphasic extraction method was selected for this platform as well.

Table B.6 Lipidome coverage of the monophasic and biphasic metabolite extraction methods in terms of detected FAMEs after methylation.

Group	Extracted phase	Average number of FAMEs (mean ± SD)
1	Single phase containing both polar and apolar metabolites	35 ± 4.3
2	Water phase containing polar metabolites	5 ± 0.8
3	Combined water phase and organic phases	36 ± 2.6
4	Organic phase containing apolar metabolites	28 ± 1.7

B.4.4 DEPROTEINIZATION OF PLASMA / SERUM SAMPLES FOR METABOLIC FOOTPRINTING

Due to the deproteinization efficiency of acetonitrile as described earlier, this method was also employed before methylation. Hence, after drying of (50 – 100 µl) deproteinized serum samples, it was re-dissolved in 500 µl methanol and 500 µl chloroform, followed by methylation as described in Section B.4.6.

B.4.5 PREPARATION OF URINE SAMPLES FOR METABOLIC FOOTPRINTING

Urine does not normally contain lipid species unless some kind of failure (whether metabolic or kidney) or overload occurs (Streather *et al.,* 1993). Hence it is questionable whether the methylation of urine samples would give useful information. Due this and the limited volume urine available per mouse, it was decide not to perform methylation on the urine samples.

B.4.6 METHYLATION OF LIPIDS TO GIVE FATTY ACID METHYL ESTERS

A well known base-catalyzed methylation method (White & Ringelberg, 1998) previously used in studies at this institution was also selected for this work. As this was the first time that this method was used on mice samples, it was necessary to standardise and validate all procedures and conditions for repeatable analysis. The main aim was to ensure that the sample extracts contained more than enough derivatization reagent so that complete methylation of all compounds was achieved. Adipose tissue from these mice was used for this test as complete and repeatable methylation of this pooled sample with established volumes would certainly imply that the other tissue samples will give adequate results with the same volumes.

Furthermore, one disadvantage of the established procedure is that much strain is put on the GC columns due to the fact that many unmethylated fats, such as cholesterol and similar compounds, are injected into the system. This not only leads to carry-over and contamination of subsequent samples, but ultimately damage to the column. Even though the established method makes use of a post-methylation extraction step using hexane and water, it removes mainly polar compounds while unmethylated lipids is still extracted in the organic solvent phase. Hence, the protocol had to be adapted in order to limit the injection of these unmethylated compounds in the system. A previously published method by Wood *et al.* (1965), which was also explored by Prof. LJ Mienie at the Centre for Human Metabonomics, NWU, was tested to bridge this problem. This method involves the re-suspension of the dried methylated samples in BSFTA in order to derivatize all unmethylated compounds (Gutiérrez *et al.,* 2002). But since it is often believed that the more active silylation reagents can replace methyl esters with trimethylsilyl-esters, it was necessary to validate this method and theory with a standard mixture of FAMEs as well as a methylated PUFA (poly unsaturated fatty acids) mix obtained from pure marine cultures.

B.4.6.1 Materials and Methods

A pooled extract of mice adipose tissue (extracted with procedure described in Section B.4.3) were used to check the methylation volumes. The tissue extract (50 µl of the organic phase and 50 µl of the water phase) was dried under nitrogen and re-dissolved in 500 µl MeOH and 500 µl chloroform. One ml methanolic potassium hydroxide (KOH) was added and the mixture incubated for 60 minutes at 60 ℃. Methanolic KOH was prepare d by dissolving 280 mg KOH (Merck, cat # 8.14353.1000) in 25 ml MeOH. After methylation the samples were allowed to cool and reach room temperature before further work was done. This ensured limited loss of volatile FAMEs especially the shorter chains. The methylated and apolar compounds was then extracted by adding 2 ml hexane, 200 µl 1 N acetic acid (Sigma-Aldrich, cat # 27225) and 2 ml water. The samples were

vortexed and centrifuged for 5 minutes at 2000 x *g*, room temperature. The top organic phase was transferred to a clean tube, dried and re-suspended in 50 µl BSTFA (containing 1 % TMCS). The samples were incubated at 40 ℃ for 30 minutes before 50 µl hexane containing 30 µg/ml methyl-tricosanoate was added. In order to validate the existing conditions, all volumes of the protocol were lowered 2, 5 and 10 times.

To confirm that the included silylation step did not result in unwanted effects, a standard FAME and methylated PUFA mix were suspended in hexane and analyzed as controls. After analysis, this same suspension was dried under nitrogen and silylated for 30 minutes at 40 ℃ by the addition of 50 µl BSFTA (containing 1 % TMCS). After silylation, hexane was added to the samples so that the same start and end volumes were obtained. All chromatographic peaks were identified with an in-house created FAMEs library as well as the NIST 08 library in order to confirm the absence of two separate peaks per compound, i.e. a peak representing the methyl ester and trimethylsilyl ester.

B.4.6.2 Results and Discussion

Firstly, it was found that the established methylation protocol were satisfactory for the tissue extracts. No difference between the chromatograms, in terms of amount of FAMEs and intensities, could be seen even when the volumes of the established protocol were lowered twenty times (results not shown). The method showed great derivatization efficiency giving excellent repeatability regardless of the ratio methylation reagents and tissue extract. Hence the established protocol was used without any modifications (except for the extra silylation step at the end).

To test the effect of the additional silylation step at the end, a standard FAMEs and PUFA sample was analyzed before and after silylation. Figure B.13A shows the overlapped chromatogram of the standard FAMEs mix with and without BSTFA. There was no obvious difference in the peak areas of most of the FAMEs, nor were any other peaks visible in the silylated sample, representing TMS-derivatives. Some of the smaller peaks of the FAME mix were not visible in all the replicates in the BSTFA suspended FAME mix (such as 2-hydroxy-C14:0 methyl ester). Of all the large peaks, only the octanoic acid (C8:0) are significantly lower. Whether this is due to the BSTFA is questionable but is more likely due to sample loss during drying or re-suspension. It is known that careless evaporation of solvents can result in the loss of appreciable amounts of esters, up to C_{14}. In particular, an over vigorous use of nitrogen to evaporate solvents must be avoided as short chain fatty acids can also evaporate in subsequent sample handling due to their volatility (Christie, 2011). Figure B.13B shows the overlapped chromatogram of the methylated PUFA mix with and

without BSTFA. Again, no difference is seen between the chromatograms in terms of peak areas and also number of peaks.

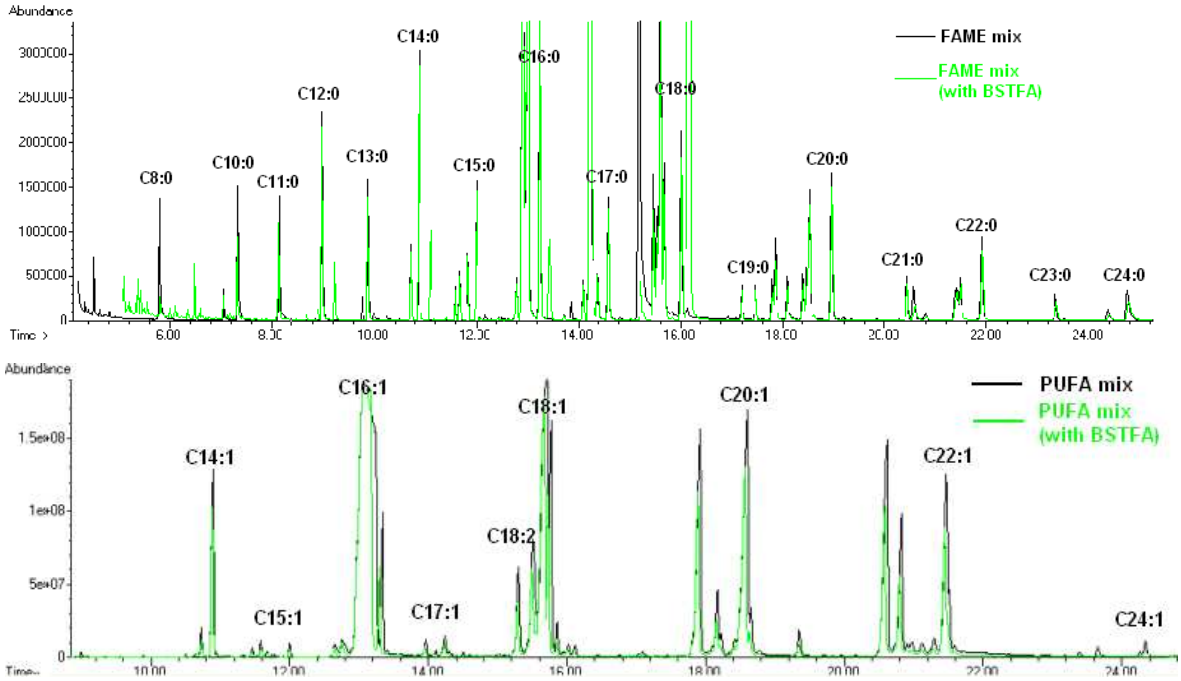


Figure B.13: Overlaid chromatogram of the FAME (A) and PUFA (B) mix alone and after silylation with BSTFA. The saturated fatty acid methyl esters are indicated in terms of the fatty acid chain length. The unsaturated fatty acid methyl esters are indicated in terms of the fatty acid chain length and number of unsaturated bonds. ":1" indicates one unsaturated bond in the molecule while ":2" indicates two, etc.

These results indicate that the methyl esters are stable enough to withstand silylation (which is often against common belief). In addition to the fact that the lipid-bound fatty acids are not hydrolyzed before hand, this two-step approach ensures that the relevant fatty acids are only detected as a methylated compound and never both methylated and silylated (unless the sample contained water). However, this does not disregard the fact that silylated fatty acids do exist with this approach. Since the free fatty acids are silylated, it is expected to be visible in the mice samples chromatograms, but this test confirmed that these are not from the lipid-bound fatty acids. The mouse adipose tissue that was used to test the methylation protocol conditions confirmed that unmethylated compounds such as cholesterol-type compounds (cholesterol, cholestan-15-one) were detected as TMS-derivatives. This means that there is less strain on the GC column and that less contamination of underivatized compounds will occur in subsequent samples. Therefore, in line with the results from Jover *et al.* (2005), this technique might be more laborious, but it is also the most rewarding.

Since the use of BSTFA with methylation results in the derivatization of certain compounds not readily methylated, one might expect that the oximation-silylation approach be unnecessary as more detail can be obtained from this methylation approach alone (best of both worlds effect). However, due to the nature of the established methylation protocol, which includes a post-methylation extraction step afterwards, all polar compounds are excluded from the silylation process. In fact, unless the pH is very low or the salt content very high, most of the organic acids, sugars, etc will be lost with this procedure which makes the oximation-silylation procedure still very valuable. In addition, since the response of methylated and silylated compounds is different, it is not technically correct to use both types of derivatives together in one data matrix, especially when normalising the detected peaks with similar internal standards.

B.5 ANNEXURE SUMMARY

In summary, all procedures and conditions of the three selected analytical platforms (reverse phase LC-MS, GC-MS with silylation and GC-MS with methylation) were validated and standardised for precision data generation. Suitable internals standards were found for retention time (and mass) alignment as well as data normalisation. The sample preparation methods were standardised in order to limit technical variance between samples. Derivatization conditions were also standardised for the same reasons and to ensure high coverage of the sample metabolome.

ANNEXURE C

METABOLOMICS DATA PROCESSING AND STATISTICAL ANALYSIS: APPROACH AND METHOD SELECTION FOR A HIGH-THROUGHPUT WORKFLOW

C.1 INTRODUCTION

Data processing and statistical analysis are key steps in the methodology of metabolomics investigations. Carefully planned and executed experiments can easily fail at giving good results when these important steps are neglected. Furthermore, there are several data processing and analysis pathways that could be taken. While there are various methods of data processing and analysis to consider, some methods might not compliment the experimental setup. For this reason it was decided to establish a standard workflow protocol (or series of processes) for this study. The steps involved in data extraction, processing and multivariate statistics are illustrated in Figure A.5, which will also be the outline of this annexure. The important step of data extraction and cleaning was firstly standardized for the LC-MS and GC-MS data, as well as validated. Several data pre-processing steps were standardized and their validity shown. The data pre-treatment and statistical methods and algorithms were selected and standardized for high-throughput analysis of the various datasets. The standardization of methods, as well as possible adaptations, was done with future (functional genomics) studies on the Metabolomics Platform of this institution in mind.

C.2 RAW DATA CLEANUP AND EXTRACTION

C.2.1 FIRST ORDER DATA CLEANING

Most of the vendor software packages include procedures and appropriate algortihms with recommendations for first order data cleaining. These recommended settings were used throughout this study and will not be discussed in detail. Seeing that over and under smoothing of very complex mass spectral data can cause alignment problems, the settings were in fact manually inspected to confirm their validity. This was especially done for the LC-MS platform. The signals from high resolution spectra are decimated into a more manageable m/z vector allowing one to compare and align different spectra under the same references and the same resolution. The default data smoothing and baseline filtering settings in MassHunter were adapted for the LC-MS data. A baseline reset > 5 if either edge < 100% were selected and 5 data points (scans) filtering to smooth the data. Four point sampling were selected to integrate the mass and chromatographic data. The baseline correction setting in AMDIS were used for GC-MS data.

C.2.2 DATA EXTRACTION

There are several data extraction software packages which differ in processing speed, algorithms and ease of use (user friendliness). Many of the available packages can also extract data inaccurately leading to incorrectly quantified peak areas, missing data and incorrect alignment, which ultimately add more unwanted variance to the data. It was therefore mandatory to compare some available software packages in order to find a data extraction package that would best fit the conditions of this study and give reliable metabolic data.

C.2.2.1 LC-MS data extraction

The downside of using open-access and freely available software, which is instrument independent, is that the raw LC-MS data files have to be converted into other universal data formats such as mzData, mzXML and netCDF (Fiehn, 2008). Given the size of the accurate mass data files this could be a daunting task, taking up valuable time and disc space. Moreover, many of the freely available software packages do not satisfy certain needs, like taking salt adduct and isotope patterns into account when extracting data. As a result, one compound are often listed more than once in a data matrix.

The combination of Agilent's MassHunter and Mass Profiler Professional (MPP) were therefore selected for LC-MS data extraction. This combination of software is relatively high-throughput and does not require the timely conversion of the data files into other formats. MassHunter is also user friendly and allows the user to process the raw data in numerous ways by changing specific parameters. One advantage of this software combination is the ability to filter the raw data in numerous ways. It is especially valuable to filter out small proteins and peptides that are often analyzed and extracted simultaneously with the metabolites when using other software packages. This simplifies statistical analysis and data interpretation. Another bonus is that it is able to deconvolute the data by adding the salt adducts and isotope traces of a compound together which many of the other software do not perform.

Two different data extraction methods were selected and used in this study. The first method extracted data in an untargeted fashion, returning all the detected features in the samples which were then used to re-extracted the data in a more targeted fashion (Kitagawa *et al.,* 2009). Despite the targeted recursive data extraction step, this method was recognized as an *untargeted* method as it extracted all the detected features. The second method were more *targeted* and extracted only known compounds listed in a given library. Both these methods will be discussed in the following section. The molecular feature extraction (MFE) algorithm in MassHunter was used to

extract data from numerous samples in an unbiased and untargeted fashion. However, due to this algorithm's speed of extracting data, it is known to miss some features during the extraction process giving the problem of "incomplete" data matrices. It also returns false positive features. As a result, the find by ion (FbI) function was introduced, which re-extracts the data in a more targeted fashion, finding the missing data and removing the false positives (Kitagawa *et al.,* 2009). For targeted data extraction, the find by formula (FbF) algorithm was used. This algorithm calculates the accurate mass and isotope pattern of a specified molecular formula and then searches this compound in all the data files.

C.2.2.1.1 MassHunter-MPP untargeted data extraction procedure

The MFE function was used to extract all the detected features in the data files. The following data filters were used in MassHunter for this first round feature extraction. The extraction algorithm was set to extract small molecules from chromatographic data with no restriction on mass or retention time windows. A peak height filter of ≥200 counts was used for all data sets. All positive ion species were allowed in the deconvolution algorithm which included +H, +Na, +K and $+NH_4$. The "salt dominated positive ions" checkbox was checked only when urine was analyzed, but was kept unchecked for all other samples as the salt content was expected to be low. The isotope grouping model was set to "common organic molecules" with default peak spacing tolerance of 0.0025 m/z plus 7.0 ppm. The charge states were limited to 2, which would ensure the inclusion of several lipid species. After extraction and deconvolution of the peaks, further filtering was done by only allowing deconvoluted compounds with an absolute height greater than 2000 counts to be shown. A mass filter was used to exclude the 922.0098 and 122.0905 masses which were from the reference mix infused throughout the run. Finally, the compound ion count threshold was set to allow only compounds containing two or more (isotope) ions to be extracted. This limits the extraction of noise peaks (which does not contain the C^{13} ion). The extracted features were identified (annotated) using the Metlin database and exported to a compound exchange file (*.cef).

These compound exchange files were imported into MPP where it were binned and aligned (using only the masses and retention times). Compound alignment was solely used when virtually no retention time drift was observed. When a clear drift was visible, retention time correction without using the standards was included in the alignment process. For compound alignment the RT window was set to 0.5 % and 0.5 minutes while the mass widow was set to 5 ppm and 2 mDa. The maximum allowed RT shift were set to 1 % + 0.5 minutes. The same compound filters that were used in the MFE algorithm, were also used in MPP. The compounds in the aligned data matrix were then used to create a compound list, which is basically a library of all the recorded features in

all the samples (including false positives). This list was then used to re-extract the data in MassHunter by using the more targeted FbI algorithm (Kitagawa *et al.,* 2009). A mass tolerance of ±7 ppm was allowed for recursive extraction of the data, while a retention time match tolerance of ±0.25 minutes was allowed. The rest of the FbI algorithm settings were similar to that of the MFE function. Default chromatographic integration settings were used as mentioned above.

The re-extracted data was then again exported to compound exchange files. For the final alignment of the re-extracted features into a data matrix, the same MPP settings were used as previously mentioned. Retention time alignment (using the standards) was only used when necessary. No normalisation or scaling of the data matrix was done in MPP. The data matrix was exported for further work in Excel, Matlab and R.

C.2.2.1.1 MassHunter-MPP targeted data extraction procedure

A molecular formula library of known metabolites was created from data available on the Human Metabolome Database (www.hmdb.ca). The library consisted of approximately 250 metabolites, which includes metabolites from all major metabolic pathways (that were expected to be detected with this platform). This library was used by the FbF function to search for these metabolites in the sample files. The same settings were used for the FbF and FbI algorithms. The extracted data was then exported to compound exchange files and aligned (using compound names) in MPP to give a data matrix. This targeted approach was only used to extract data from problematic batches with RT and mass shifts that cannot be corrected (i.e. batch effect). The compound intensities in these datasets were then corrected with the batch-normalisation method described in Section C.4.1.4.

C.2.2.2 GC-MS data extraction

AMDIS remains the work-horse when it comes to GC-MS data extraction (Likić, 2009). Unfortunately, it is not truly suited for *metabolomics* data extraction. Furthermore, experience has shown that it also give unwanted variance when creating a data matrix. The unwanted variance is not only a result of erroneous peak quantification (by assuming a homologous baseline across the entire chromatogram, Likić, 2009), but also from the presence of false positives in the data, over-deconvoluted peaks, erroneous alignment and missing values. Over-deconvoluted peaks refer to peaks that are in fact one compound peak, but recognized by AMDIS as two (Fiehn, 2008). Due to the limitations of AMDIS, a few other software packages (combinations) were evaluated for GC-MS data extraction, especially as the quality of the data put into statistics reflects the quality of results that will be obtained.

The software package of choice had to satisfy the following criteria:

i) The software (combination) had to be automated but flexible, meaning that most parameters can be modified by the user;
ii) It must be high-throughput and fast processing, without taking up too much computing resources;
iii) Be user friendly and fairly easy to understand;
iv) And most importantly, be able to extract the data correctly to provide a complete data matrix with little or no missing values or inadequate alignment.

To test the ability of the software packages to extract data correctly, five replicates of a tissue sample were analyzed and processed. This approach was taken to ensure that there was enough variance between the replicate samples, which the software packages could "amplify". Theoretically the results had to be identical as a single tissue sample was used, prepared and analyzed in a similar fashion. Obviously some small differences would occur such as small compound peaks that might be missing in one or two of the replicates, or perhaps there might be some RT shift causing the software to align certain peaks incorrectly. The objective was to investigate which software package would overlook (or correct for) these differences and which software would amplify these differences to incorporate extra variance in the data that do not exist.

C.2.2.2.1 Materials and Methods

A single urine sample was divided into five replicates and prepared as described in Section B.3.5. After organic acid extraction, the samples were derivatized (Section B.3.6) and analyzed (Section B.3.1). The raw GC-MS data files were converted to netCDF files in ChemStation (Agilent) and processed separately with the software packages listed in Table C.1.

Table C.1: List of GC-MS data extraction software packages investigated.

Software	**Functions**	**References**
AMDIS	Peak picking, deconvolution	Stein, 1999
Mass Profiler Professional	Component alignment, statistics	Agilent
MetaboliteDetector 1.5	Peak picking, deconvolution	Podstawka, 2010
SpectConnect	Peak alignment	Styczynski *et al.,* 2007
XCMS	Peak picking, alignment, statistics	Smith *et al.,* 2006
MZmine2	Peak picking, deconvolution, alignment	Pluskal *et al.,* 2010
MET-IDEA	Feature extraction, alignment, statistics	Broeckling *et al.,* 2006
MetaboAnalyst	Peak picking, alignment, statistics	Xia *et al.,* 2009

AMDIS can only export identified (annotated) components and not those that remain unidentified. Thus, data from the replicate urine samples were extracted with AMDIS by building a complete library of all the detected components and re-analyzing the samples. A report was generated after which the detected components were aligned with an in-house written function in Matlab. Those components that contained a '?' were manually checked and realigned if necessary. This was done in the following manner: firstly it was checked whether this compound were already identified in the sample. If not, the retention time of this semi-identified compound was compared to that of the accurately identified compound in the other samples. When the retention times were similar, the '?' was removed. This raw data matrix was then compared with raw data matrices obtained from the other software combinations. The other software packages were used according to the manufacturer's or author's specifications. MetaboliteDetector 1.5, XCMS, MetaboAnalyst and MZmine2 used the raw netCDF files for data extraction. MPP and SpectConnect used the *.elu and *.fin files that AMDIS created after deconvolution of the netCDF files. MET-IDEA uses the *.elu and *.fin files along with the netCDF data files to "re-extract" the data.

The raw data matrices obtained from the different software were normalised using the internal standards. The matrices were compared without further cleaning to rule out user-related bias. These matrices were also compared after identical data pre-processing steps were taken. The uncommon variables not detected in at least four of the five replicates were removed. Missing value imputation was performed using mean replacement. As it is difficult to measure the true variance in the data, it was decided to use the data obtained from AMDIS as the benchmark. Therefore, the results from the other software were compared to the AMDIS results. The coefficient of variance (CV) for all the detected compounds was determined as an indication of variance (Sysi-Aho *et al.,* 2007). The number of zeros in the data was also determined.

B.2.2.2.2 Results and Discussion

The results of this evaluation are shown in Figure C.1. Only 25 % of the components extracted by AMDIS had a CV lower than 50 %. In terms of quality control, this result would indicate that the experiment (or batch) is very unreliable for metabolomics work. A similar result was obtained after the deconvoluted components were aligned with MPP, with 34 % of the variables having a CV < 50 %. The raw data matrices obtained from MET-IDEA, SpectConnect, XCMS and Mzmine2 gave much better results in terms of variance. MET-IDEA outperformed all the tested software as 84 % of the extracted components had a CV < 50 %. Only when several of the other data matrices were thoroughly cleaned, did it give better results. After cleaning the SpectConnect matrix, only 9 % of the remaining components had a CV > 50%. No results for MetaboliteDetector 1.5 are shown in

Figure C.1. The version used was not suited for large batches as it could only process one sample at a time and was not able to export results to other formats.

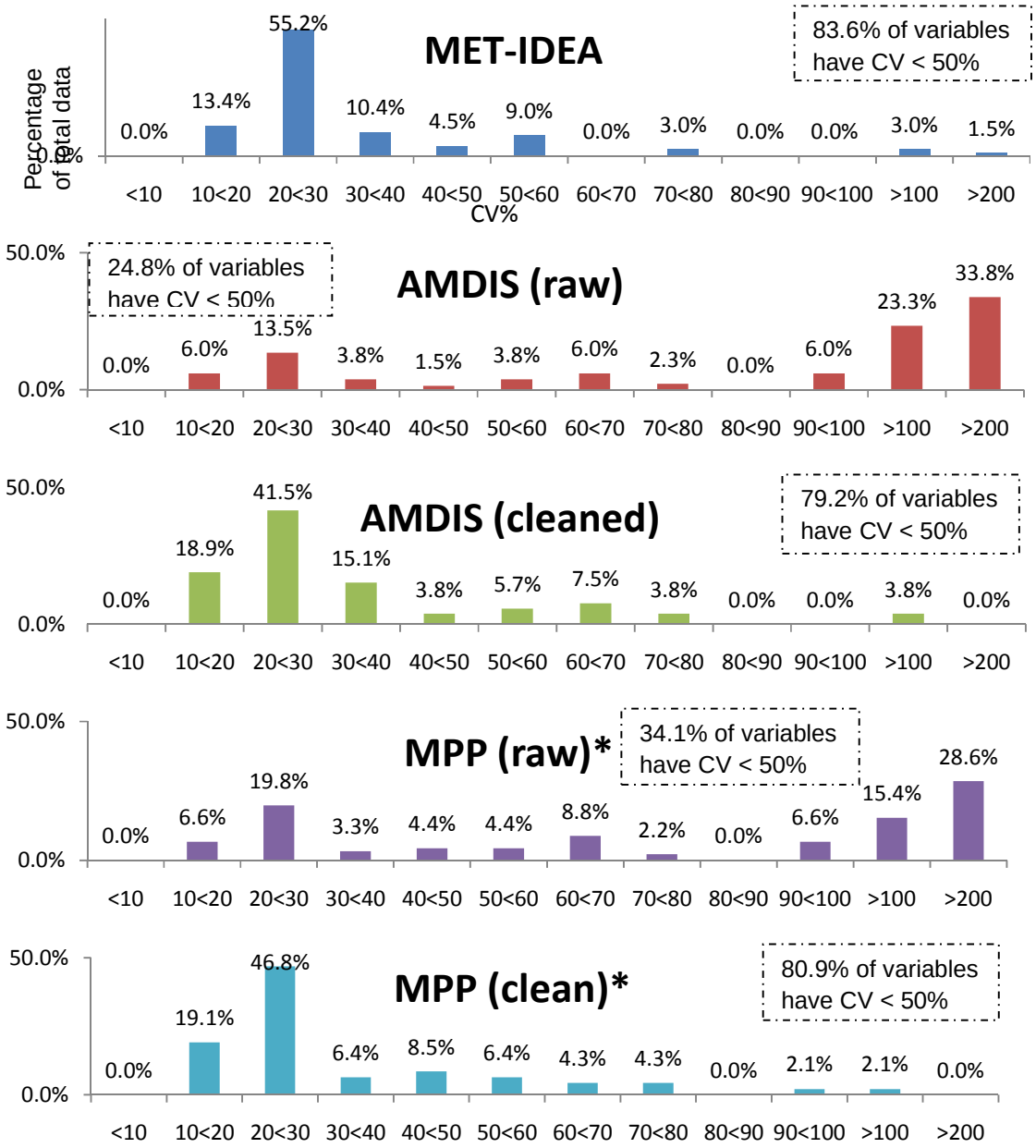
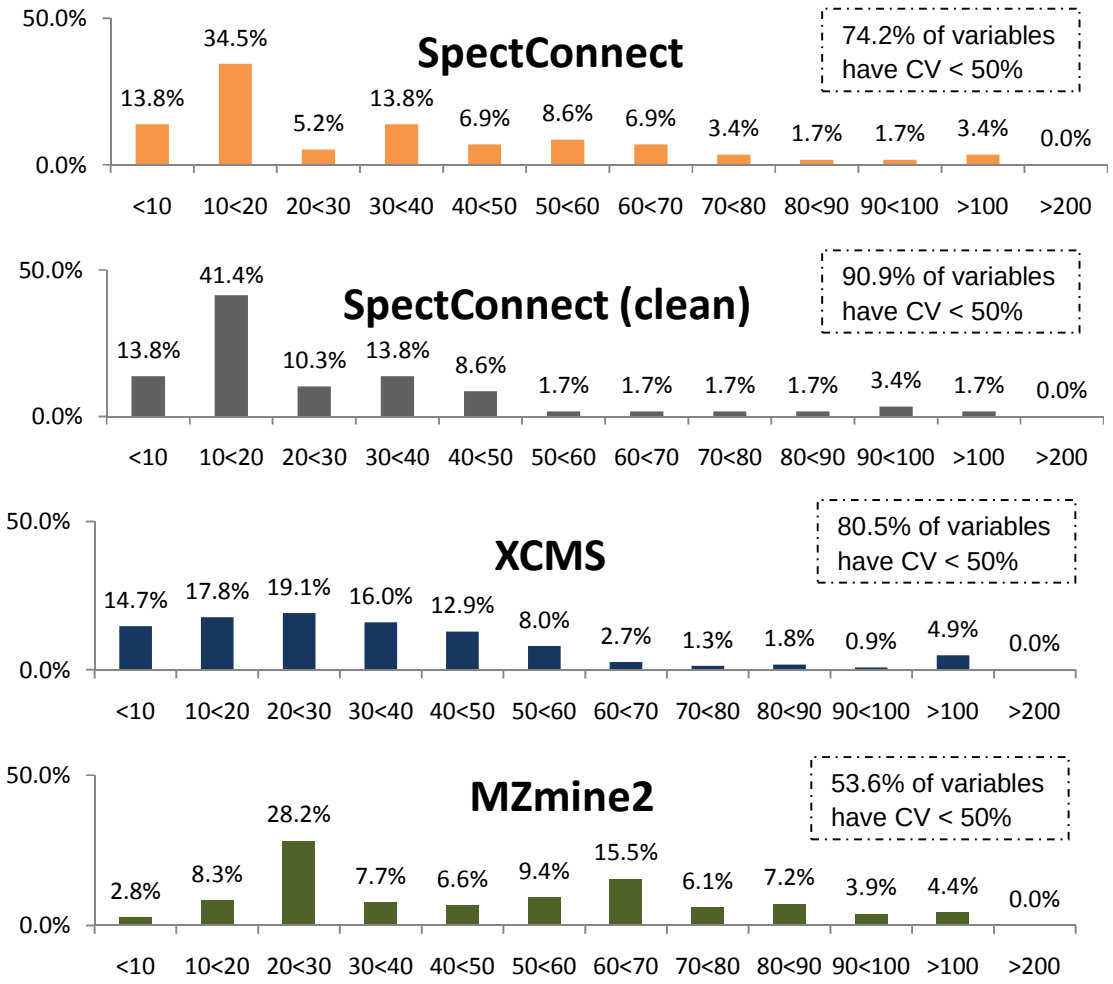


Figure C.1: Comparison of the variablitiy of the extracted data of five replicate samples with numerous software options. Each graph shows the percentage of data that falls within a given CV range and thereby the CV distribution (Sysi-Aho *et al.*, 2007). The percentage of measured data that have a CV smaller than 50 are also indicated in the boxes. “Raw” in brackets means that the matrix that were created by the software were used as it was created with minimum cleaning. “Clean” in brackets means that much manual labor went into cleaning the data matrix from uncommon variables and missing values.



* AMDIS .fin files containing library hits were not used for data extraction

Figure C.1 (continued): Comparison of the variablitiy of the extracted data of five replicate samples with numerous software options. Each graph shows the percentage of data that falls within a given CV range and thereby the CV distribution (Sysi-Aho *et al.,* 2007). The percentage of measured data that have a CV smaller than 50 are also indicated in the boxes. “Raw” in brackets means that the matrix that were created by the software were used as it was created with minimum cleaning. “Clean” in brackets means that much manual labor went into cleaning the data matrix from uncommon variables and missing values.

Table C.2 shows the amount of zeroes in the data matrices obtained from several of the different software packages. Matrices created by MET-IDEA, SpectConnect and MZmine2 gave the least missing values. After removing most of the uncommon variables, the number of variables in certain matrices dropped considerably. The original AMDIS created matrix contained 133 variables, which dropped to 53 after uncommon variables were removed. MPP with or without the AMDIS created *.fin files had the same effect. The SpectConnect, MET-IDEA and MZmine2 matrices did not

contain any uncommon variables. After data pre-processing, the MZmine2 matrix still contained 181 variables. However, after inspecting the data several replicate variables were found from over-deconvolution with the default settings.

Despite the fact that the true variance of the data was not determined (as it would require software processing), these results nonetheless clearly indicate the variance that different software can introduce. This experiment has therefore shown that data extraction is just as important as the experimental setup. Great differences in the amount of variables, overall variance and missing values were observed when five replicate urine samples were analyzed with different software. In terms of quality control, many of the variables containing a CV > 50 % would have been excluded from statistical analysis due to the "additional" variance given by the extraction process. Furthermore, since PCA and other chemometric methods compress all the variance in the data into a few latent components, the influence of software-related variance can greatly influence the results. While most of the unwanted variance is removed during data pre-processing, some additional (software-related) variance can still exist in the remaining data. Each software package perceives and calculates the apex and area of a peak in different ways (which also depends on the deconvolution algorithm). The selection of a suitable package therefore relies on this factor as well. In light of these results, the software combination that was best suited for this study was AMDIS and MET-IDEA. This combination allowed high-throughput data extraction, giving virtually no missing values. AMDIS was used for deconvolution and peak picking while MET-IDEA was used to re-extract (re-quantify peak areas) and align features into a data matrix. The process of data extraction will be discussed in the following paragraphs.

Table C.2: The amount of missing values (zeros) incorporated in the extracted data matrices from several tested software.

Software	Amount of zeros in raw data matrix (%)	Amount of variables	Amount of variables after cleaning (%)
AMDIS	42.4	133	53
MPP (with *.fin)	45.9	88	31
MPP	34.3	91	47
SpectConnect	3.4	58	58
MET-IDEA	0	69	69
MZmine2	6.5	181	181

C.2.2.2.3 AMDIS – MET-IDEA data extraction procedure

The Agilent GC-MS raw data files were first converted to netCDF files and put into a single folder. This conversion was performed in ChemStation (Agilent). All the QC samples were inspected in AMDIS to adapt the settings for optimal deconvolution and to find a good representative sample, which could be used as reference in MET-IDEA. Under the deconvolution tab in the "analysis settings" window, the component width was changed according to the average peak width – usually 20. The wider the average peak, the higher this value was set as it meant that more scans had to be used to pick and quantify the peak. As recommended, all other settings such as resolution, sensitivity and shape requirements were kept low. When these settings are on medium or high, up to 70 % false positive components could be obtained as AMDIS over-deconvolutes the peaks (Likić, 2009). The sensitivity was increased for some datasets when small but definite peaks were detected, but it also increased the noise in the data matrix. As recommended, TIC, 73 and 147 m/z were not used as model ions in the deconvolution process. This lowered the false positive rate. After optimal settings were found, all the raw data files were processed in batch mode to obtain *.elu and *.fin files for all the data files. These contain the detected components and matched library hits, respectively. The *.elu and *.fin files were used by MET-IDEA to extract and align the data. For the batch processing the NIST 08 library was used for component identification. Identification was based only on the component spectra data. All hits above 70 % were listed. The identity of the peaks in the reference QC sample was thoroughly checked. Peaks in this reference that, contained more than one hit, were inspected and the best hit used. In rare cases were the net match probability of several hits was similar, the best match was then taken based on the reverse search match percentage.

The reference QC sample (representing the whole batch) was then loaded in MET-IDEA to create an ion list. The ion list is basically a library of all the peaks in the sample (and batch). This ion list was used to extract similar components in the other samples in a similar targeted way as the FbI function in MassHunter does with the LC-MS data. After the ion list was created and the standards checked for RT and mass alignment, a calibration file was created which showed the RT and mass drift in the batch. This calibration file was then used to align the chromatograms (RT) as well as the spectra (masses), quantify the areas of the peaks and create a data matrix of all the samples. Importantly, since there was no noticeable RT and mass drift with the GC-MS instrument for samples analyzed within a week, it was not necessary to use linear regression for alignment, but rather "shift by constant" which was faster. The linear regression option was only used when samples were analyzed over a month's time.

C.3 DATA PRE-PROCESSING

Pre-processing steps (as well as the sequence of use) might differ between different data sets due to analytical instrument and data extraction software differences. For example, LC-MS data often contain salt-adducts due to under-deconvolution, while GC-MS data often contain replicate variables due to over-deconvolution. Both have to be corrected but the processes can differ. While it is impossible to establish a rigid, single sequence pre-processing procedure, it is possible to establish some kind of workflow (or checklist) for each dataset to make data processing more high-throughput. Taking the nature of each analytical platform into account the following pre-processing steps were performed for all datasets. Matlab functions were written to perform these pre-processing steps on the data matrices.

C.3.1 LC-MS DATA PRE-PROCESSING

After extraction of LC-MS data, the data matrices were inspected to ensure that similar compounds were aligned. It often happens that the RT shift is too large in a batch for MPP to correct it. When looking at the accurate masses, a clear trend (constant shift) was then visible. When a constant shift was observed, it was corrected “manually” using in-house written Matlab functions. However, when the RT shift was not constant, the alignment settings in MPP were adapted and the features re-aligned. In the rare case where this did not result in better aligned data, the data was then re-extracted with the targeted approach previously described and the data aligned using the compound names.

Since not all the detected compounds were identified before statistical analysis, it was expected that non-biological compounds, contaminants and artefacts be present in the data matrix. Their inclusion can complicate data interpretation greatly and is thus best to remove them. For this reason a blank sample containing only the standards were prepared with all the other samples in the batch in the assumption that it would capture most of the contaminants in the solvents or artefacts formed during the sample preparation procedures. The average normalised peak area of a possible contaminant in the samples was divided by the normalised area of this compound in the blank sample. All contaminants which gave a value lower than 5 were removed from the matrix. It is also common practice to analyze a blank sample periodically to monitor column carry-over (compounds that do not completely elute from the column and that leak into the subsequent samples). This approach was however not used as such compounds would theoretically contain a large CV and would be eliminated from the matrix in the subsequent pre-processing steps. The elimination of these compounds was in fact confirmed by an approach described by Burton *et al.*

(2008) and Mohamed *et al.* (2009). Two dummy groups were created of the first and last eight samples in a randomized batch and analyzed with PCA. When separate grouping was obtained, the compounds that resulted in this separate grouping were removed.

The next step involved the removal of uncommon variables not present in all but one of the samples of an experimental group. This approach was taken to ensure the use of mean replacement of the missing values. The created Matlab function first counted the amount of zeros (or ones when using LC-MS data) for every variable in an experimental group. If the data set contained three experimental groups, then each variable would return three values. The maximum value (of the three) was then determined, after which all variables with a maximum of >1 were removed. Missing value imputation was performed after normalisation by replacing the missing values with the mean of the experimental group (nearest neighbours; Steinfath *et al.,* 2008; Steuer *et al.,* 2007). It was shown in a poster presentation (Hrydziuszko & Viant, 2010) at the Metabolomics conference in Amsterdam (2010) that mean replacement remains superior to other more advanced methods.

The ROBpca function in the LIBRA (Library for Robust Analysis) toolbox was used to screen for outliers in the data (Verboven & Hubert, 2005). This screening was done for each experimental group separately, focusing on the intra-group variance. As shown before, this function visualizes the sample scores in a (Mahalanobis and orthogonal) distance-distance plot to identify pertinent outliers. Identified outlier(s) that were the least similar to the other samples in the group were flagged, but not immediately removed. Since there is no accurate statistical measure to use when confronted with the choice of removing or keeping possible outliers, all identified outliers were kept for analysis. The flagged samples were only removed from the dataset after the first round of statistical analysis (PCA) confirmed this (t'Kindt *et al.,* 2009; Maggio *et al.,* 2008; Urbanczyk-Wochniak *et al.,* 2007).

The final cleanup step involved the removal of unstable and unreliable variables, which had a CV above 50 % in the QC samples. Since the QCs are identical samples, prepared and analyzed throughout the experiment, identical peak intensities are expected. This means that any variable with a relatively high CV in the QC samples have high technical variance due to unreliable extraction, derivatization, analysis or due to degradation. Regardless of the cause for high variance, the data for these variables were not considered as accurate and were removed.

The use of these data pre-processing steps, to improve statistical analysis, was validated (against using none). A single tissue sample was divided into ten replicate pieces of equal weight. The replicates were also divided into two groups so that five samples would be prepared with one method, while the other five was prepared by another. The use of two sample preparation methods ensures inter-group variance, while intra-group variance comes from the five replicates in each "experimental" group. The aim of this experimental was to highlight the inter-group variance, while limiting the intra-group variance as a means to mine relevant information.

C.3.1.1 Materials and Methods

A MT1+2KO mouse liver sample was divided into ten replicates of equal weight. The first five replicate samples were prepared with the established sample preparation protocol (Section B.2.3). The remaining samples were prepared with the same method except that chloroform was replaced with acetonitrile to incorporate inter-group variation. The samples was analyzed (Section B.2.1) and normalised with the added standards (Section C.4). To illustrate and validate the effectiveness and necessity of the pre-processing steps, the attained data matrix was analysed with PCA before and after data cleanup. It must be noted that since a single liver sample was analyzed in replicate, it was not necessary to include a QC sample. The replicates of each experimental group were used as QCs to remove unreliable compounds from the data. Compounds that had a CV > 50% in both groups were removed.

C.3.1.2 Results and Discussion

Figure C.2 shows the data and score plots of the data before and after data cleanup. There is a clear difference between the data plots as most of the compounds larger than 500 Da were removed with the pre-processing. The PCA score plot of the "raw" matrix is very similar to the score plot of the "clean" data matrix except for the intra-group variance on PC1. It is not always possible to comment on the intra-group variance when looking at two different PCA score plots. However, the difference between the intra- and intergroup variance in this case can clearly be observed.

As mentioned in Annexure A, there are mainly three types of variation in the data. These are relevant biological-, irrelevant biological-, and technical variance. The key to metabolomics experiments is to reveal the relevant biological variation above all the other variation in the data. In order to reveal the relevant biological variation, it is necessary to remove irrelevant variation such as unreliable (unstable) variables (Burton *et al.,* 2008; Shurubor *et al.,* 2005). Also, missing values in the data and uncommon variables only add more unwanted variance which does not give

relevant information but rather masking it and making data interpretation more difficult. This effect was demonstrated as shown in Figure C.2. Since multivariate statistical methods, such as PCA, consider all the variance in the data, it is therefore obvious that the irrelevant variance should be limited where possible. After data cleanup, group separation (and relevant information) was "unmasked". This means that the loadings would also be more reliable and informative.

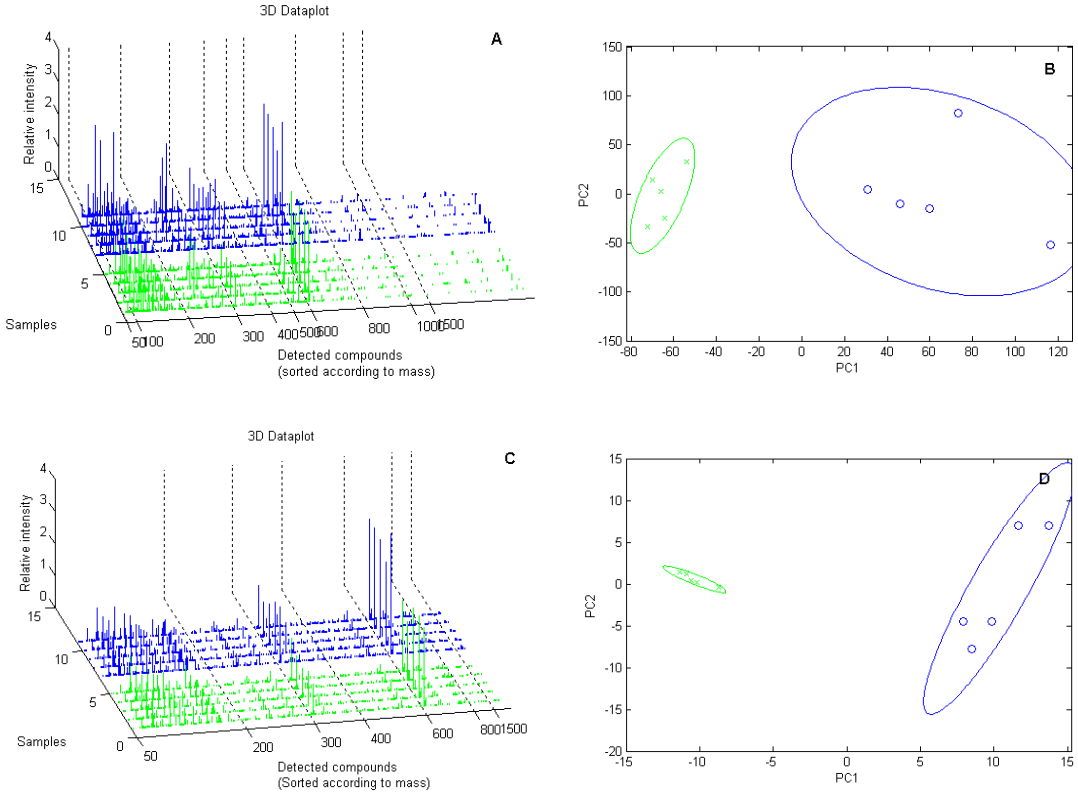


Figure C.2: Data and score plots of LC-MS data before and after pre-processing. The top plots (A and B) were created from the raw data after data extraction. After data pre-processing the bottom two plots (C and D) were created. The detected compounds are sorted according to mass in the data plots.

C.3.2 GC-MS DATA PRE-PROCESSING

As mentioned previously, a standard protocol for data pre-processing had to be developed to ensure high-throughput data analysis. Therefore, all datasets were treated in a similar fashion with some small differences between the LC-MS and GC-MS data. As with the LC-MS data, the first step in data pre-processing entailed inspecting the raw data matrix for any alignment errors. The combination of AMDIS and MET-IDEA seldom gave alignment problems, which meant that re-alignment was often not necessary. However, MET-IDEA can in certain cases align samples incorrectly, resulting in incorrectly quantified peak areas. Therefore, the alignment was always

inspected in MET-IDEA by focusing on the internal standard peaks. When the quantification window under the internal standard peaks of a sample(s) was skewed, the parameters of the calibration file were modified and the samples then re-aligned. The data matrices were also inspected for deconvolution errors such as over-deconvolution. The correlation analysis function in MET-IDEA was used to identify variables that correlate linearly across all the samples. The logic behind this is that over-deconvolution of a compound peak force MET-IDEA to select different (and in some cases the same) model ions for this single compound leading to somewhat similar results, which correlate across the entire batch. These compounds were then inspected manually and the replicate variables removed. As a rule, the higher number model ion was retained as it is more compound-specific in comparison to low value ions. The data matrix was then exported to Microsoft Excel.

Following the formation and export of the GC-MS data matrix, irrelevant variables belonging to contaminants or artefacts was removed. Since the reference sample was analyzed with the NIST 08 library, most of the detected compounds were confidently identified. All the non-biological compounds were then removed from the matrix. However, many compounds remained unidentified. This meant that among the unknown compounds possible contaminants or artefacts could exist. A second approach was used to monitor unidentified contaminants similar to the one used for the LC-MS platform. A blank sample, containing only the relevant internal standards, was put through the same sample preparation and derivatization steps to ‘capture’ the contaminants (such as hydrocarbons) and artefacts from the silylation process. The same approach was then used to remove these supposed contaminants and those from carry-over.

Similar to LC-MS data, the following pre-processing step was to perform variable selection or reduction. However, unlike LC-MS data, the GC-MS data given by MET-IDEA seldom contain any missing data in terms of zeros in the data matrix. Therefore, the step of removing uncommon variables was omitted. Furthermore, it was also seldom necessary to replace missing values. However, mean replacement was also used to replace the few zeros that might exist. Outliers and unreliable variables were screened and treated as described earlier.

C.3.2.1 Materials and methods

The same experimental setup as described in Section C.3.1.1 was used to validate the selected data pre-processing steps on GC-MS data. After metabolite extraction, the samples were dried, derivatized and analyzed as described in Section B.3.6 and B.3.1, respectively.

C.3.2.2 Results and discussion

Figure C.3 shows the data and score plots of the data before and after data cleanup. Small differences are visible in the data plots, while larger differences can be seen in the PCA scores plots. The main observable difference between the data plots is the amount of variables. Before data cleanup there was approximately 200 variables. After data cleanup this number dropped to just under 70. Grouping in the score plot of the raw data was poor. Also, PC1 indicates that the highest amount of variance in the data is the variance between the samples in the groups (intra-group variance). The relevant variance which indicates the group differences is given by PC2, which indicates the second highest variance in the data. The score plot of the clean data show better grouping and also less intra-group variance. The highest amount of variance in the data (as indicated by PC1) is the differences between the two groups. The intra-group variance is given by PC2 suggesting it to be inferior to the relevant variance after data cleanup.

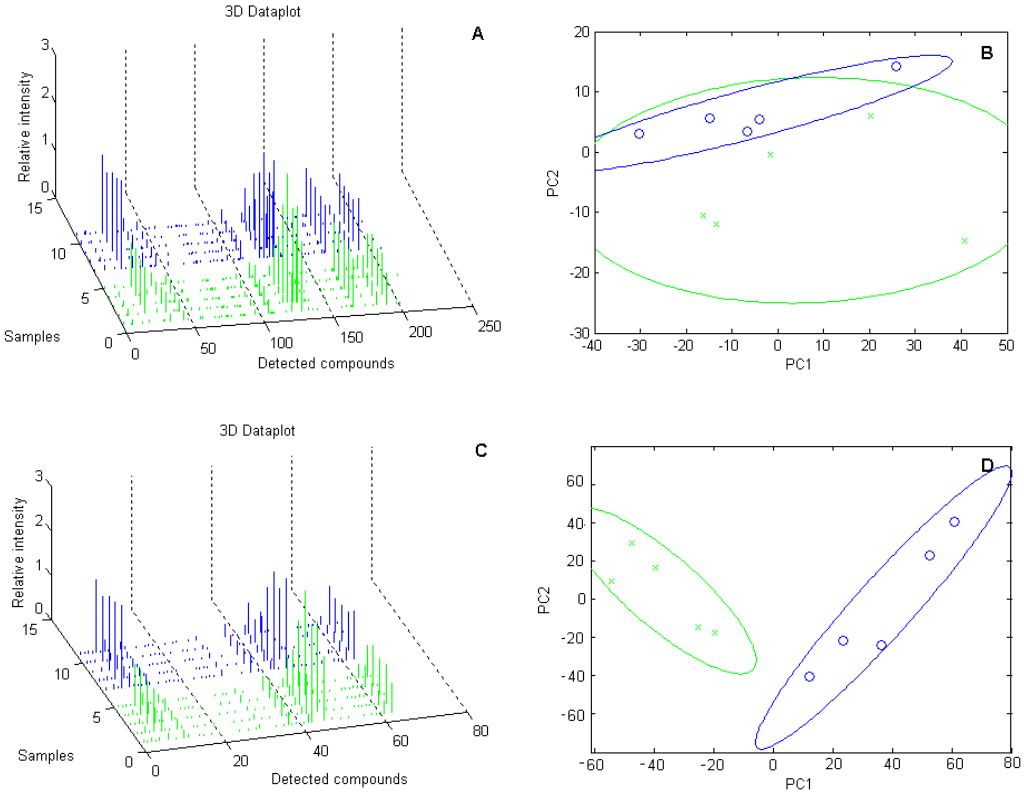


Figure C.3: Data and score plots of GC-MS data before and after pre-processing. The top plots (A and B) were created from the raw data after data extraction. After data pre-processing the bottom plots (C and D) were created. The x-axis of the data plots indicates the number of variables.

The main cleanup step performed with this data set was to remove unreliable variables. This gave less intra-group variance and better separation of the groups as seen in the score plot. Since the

separation between the groups shifted from PC2 to PC1 after data cleanup, it was concluded that the relevant biological variation in the data was unmasked by the elimination of irrelevant variance. While grouping was visible in the first score plot (Figure C.3B), the loadings from this result would not be quite as reliable as the loadings from the second plot (Figure C.3D).

The results of both the LC-MS and GC-MS data before and after data cleanup signify the importance of data pre-processing. Even though there is no 100 % (in)correct pre-processing protocol, careless use of these pre-processing steps might cause information loss when too much or too little 'irrelevant' variables is removed. Nonetheless, the sequence of events as mentioned above was standardized to allow high-throughput processing of data sets.

C.4 DATA NORMALISATION AND PRE-TREATMENT

C.4.1 DATA NORMALISATION

Unwanted variation in the form of experimental factors, such as different sample volumes, instrument responses and sample loss, can be corrected during data normalisation. There are numerous methods that can be used to normalise data (Redestig *et al.,* 2009). The best known normalisation method is where the peak areas of all the detected compounds are divided by the peak area of an internal standard. Other methods are described in Annexure A. These normalisation methods function in different ways with every method having special preferences. Every method also has some advantages and disadvantages which have been documented (Sysi-Aho *et al.,* 2007; Redestig *et al.,* 2009; Warrack *et al.,* 2009). Instead of testing and validating the different methods, it was decided to incorporate three different normalisation methods into the data processing protocol, to be used as required. The three selected methods would then cover all normalisation needs during sample processing. For example, when problems occurred with the internal standards of a dataset, the use of an alternative normalisation method, using the total signal, could be used.

C.4.1.1 Normalisation to sample weight

In this study all of the test samples were normalised to their respective weights (weight, creatinine or protein levels) before sample preparation. This means that this type of normalisation was not required afterwards. While it could save time to do this normalisation after analysis, it is also advantageous to do it before sample preparation as it ensures that the ratio of solvent and cells/metabolites is similar in all samples. Consequently, all samples then have the same extraction efficiency coefficient. This is especially important when working with tissue samples

because of the probability that more metabolites can be extracted (per weight) when the tissue sample is small. The larger tissue sample might give limited extraction as the solvent gets saturated more quickly. In such case, if one assumes that both sample solvents were equally saturated with metabolites, it is then obvious that the concentration per sample weight of these would drastically differ giving the illusion that the second, larger sample contained less metabolite per tissue weight. The pre-analysis normalisation approach also ensures that no bio-fluid sample is too diluted as this would result that most of the compounds are below the detection limit of the instrument and thus 'absent'. Despite the use of pre-analysis normalisation it was still decided to incorporate this method into the data processing protocol to allow more accurate normalisation of specific samples (such as urine). The creatinine values detected in the urine samples after LC-MS analysis could be used (along with the internal standards) to normalise the urine samples more accurately.

C.4.1.2 Normalisation to internal standard(s)

All of the samples analyzed with the LC-MS and silylation-GC-MS platforms contained multiple internal standards. The CCMN method from Redestig *et al.,* (2009) was therefore primarily used throughout this study to normalise the data. However, only one external standard was used with the methylation-GC-MS platform. These data sets were normalised by dividing all the peak areas of all the detected compounds with that of the single standard. The external standard was used to correct for possible instrument changes over time (in the batch). One limitation of this approach is that variation introduced in the sample preparation steps cannot be corrected. The great repeatability of the methylation method and the semi-targeted nature of this approach mean that this problem could be ignored. However, certain tissue samples are more problematic when it comes to sample preparation and is it mandatory to remove/correct for experimental factors to make them more comparable. In this case the method described in the following paragraph was used.

C.4.1.3 Normalisation to total signal

This normalisation method was employed for certain problematic data sets where problems occurred with the internal standards or where the standard could not correct for specific experimental factors (as mentioned in the previous section). MSTUS makes use of the total useable signal for normalisation (Warrack *et al.,* 2009) and is well suited for clean data sets where uncommon variables, contaminants and xenobiotics have been removed. As demonstrated by Warrack *et al.* (2009) this approach has some advantages and allows for accurate normalisation of

the data especially since the ratio of all the (useable) compounds is used instead of the ratio with a single compound (internal standard).

C.4.1.4 Batch normalisation and signal correction

The phenomenon referred to as *batch-effect* is a great threat in metabolomics and have been a hot topic at a recent Metabolomics Conference in Amsterdam (2010). While it is possible to correct for many experimental factors using internal and external standards, it takes more effort and sophisticated tools to correct a batch-effect, to make the samples in different batches (inter-batch) or same batch (intra-batch) more comparable. Numerous methods have been published up till now which deals with batch normalisation or signal correction. Most of these methods make use of (pooled) QC samples which span the batches. Because the QC samples are in fact identical samples, these are normally used to correct any batch-effect (which includes within-batch or between-batch effects). Most of the reported methods correct the abundances and/or distribution of each variable irrespective of the other variables (Draisma *et al.,* 2010; van der Kloet *et al.,* 2009; Wagner *et al.,* 2007). However, no report exists in the field of metabolomics (according to our knowledge) that deals with the problem of batch-effect in a multivariate way. Keeping in mind that metabolite covariance is important for multivariate metabolomics studies, the use of univariate methods to correct batch-effect could influence the results in a negative way. This problem can be solved by using multivariate approaches (such as principal component regression) to correct the variance of all the detected metabolites.

Batch-effect can be corrected in a multivariate way using the principal workflow of the univariate methods. The method of van der Kloet *et al.* (2009) uses linear regression to determine the regression line (and error factor or residual) for each variable in the QCs. This model is then used to correct the variables in the samples. But instead of performing this method on each variable, it could be applied to the first PCA score, which account for most of the variance in the data. Another method closely related to principal component regression is PLSR (partial least squares regression), which first reduces the data into fewer latent variables before regression modelling is performed. However, the method selected for batch normalisation in this study is basically an extension of the direct orthogonal signal correction (dOSC) method (Westerhuis *et al.,* 2001). PLS was used to find the residual matrix when the QCs were projected onto a response vector. PCA was performed on the residual matrix to find the largest variance that does not correlate with the response vector and which thus explains the largest error. This error was then removed from the sample matrix. This method was only used when a clear batch-effect was visible.

C.4.2 DATA PRE-TREATMENT

A few relevant data pre-treatment methods were described in Annexure A. These pre-treatment methods focus on different variance (or variables) in the data leading to different results. For example, VAST scaling give more weight to the stable variables which have a low coefficient of variance. Power transformation, on the other hand, keeps the dimension of the data intact while reducing the impact of the more abundant compounds slightly. When an important metabolite (or biomarker) in a dataset is for example a low, but stable (low CV) compound present in the urine, the use of power transformation alone might result in insufficient grouping of the experimental groups and thus be unable to identify this biomarker. Since VAST scaling reduces the impact of unstable variables containing a high CV, this pre-treatment method stands a better chance of identifying this biomarker. In light of this, and to keep with the trend of screening for information in an untargeted way, it was decided to use more than one pre-treatment method in this study. The loadings obtained with each method are then combined and ranked to give a more complete picture of the possible biological variance. The ranking system is described in the following section (C.5.2.1).

The four selected pre-treatment methods are VAST scaling, power-, fourth root- and log transformation. As mentioned, power transformation keep the structure of the date mostly intact and does not make the data dimensionless as when autoscaling is used. It removes heteroscedasticity and moderately reduces the impact of the highly abundant compounds. With this pre-treatment method, the focus still remains on the more abundant compounds, while giving some weight to the lower intensity compounds. Measurement errors in the data remain insignificant and do not distort the data. High abundant compounds that are important in grouping are easily identified when this pre-treatment method is used. Fourth root transformation which is closely related to power transformation was selected as the second method (Idborg *et al*., 2005b). It reduced the impact of the highly abundant peaks slightly more than power transformation but not as much as log transformation. Log transformation also removes the heteroscedasticity, while greatly reducing the impact of the highly abundant compounds making the data basically dimensionless. Every variable in the data set is virtually given an equal weight. However, while all variables have equal say, the focus of this method really shifts to the lower abundant compounds as it is their variance that are amplified to dominate the analysis. Hence, the advantage of this is that relevant biological variance in low abundant metabolites can be detected. Unfortunately, the measurement errors are also inflated which can lead to highly distorted results if these are not prior removed.

Despite the 'focus' of each method one common trait in all of them is that the variables with the most variance (inter- and intra-group variance) have the largest weight in the outcomes. This influence linearly decreases to those variables having the smallest intra-group variance, which seems inappropriate for many metabolomics data sets. One would expect that this trait be in the opposite direction with the more stable variables having the most weight, linearly decreasing to those variables that are very unstable (such as measurement errors). Fortunately, Keun *et al.* (2003) developed a data pre-treatment method that ranks the variables according to their stability. Hence, the more stable the variable is (per experimental group), the more weight it has in the statistical analysis. This method was also selected as it brings the focus to the more stable variables which are especially useful for metabolomics data.

C.5 STATISTICAL ANALYSIS

C.5.1 UNIVARIATE STATISTICS

The use of multivariate statistical methods is becoming more and more popular in metabolomics as it focuses on the covariance of metabolites, thereby finding patterns and biomarkers not readily found by the more classical univariate methods. However, this advantage of multivariate analysis is in fact also a flaw. Not all metabolites co-vary in a biological system which means that information could potentially be lost with the solitary use of multivariate statistical methods. The inclusion of univariate methods, which analyze each metabolite separately, is thus advantageous especially in an explorative, functional genomics study like this.

The commonly used Student's T-test (in case of two groups) and ANOVA (in case of three and more groups) was incorporated in the statistical analysis workflow. The MetaboAnalyst web service (Xia *et al.,* 2009; www.metabolanalyst.ca) was used to perform these tests as it allows the high-throughput analysis of the numerous data sets (multiple univariate analysis). The data was log transformed before these tests were performed to ensure a more normal distribution of the metabolite values (Boudonck *et al.,* 2009; Jonsson *et al.,* 2004; Steuer *et al.,* 2007). The effect size of the metabolites that differed *significantly* between the groups were determined from the log transformed data to indicate whether the difference were *practical* (Ellis & Steyn, 2003; Reinecke *et al.,* 2011). Equation C.1 was used to determine the effect size d, where $\tilde{y}_1$ and $\tilde{y}_2$ are the respective group means and SD_{max} the maximum standard deviation of the two groups. A d-value above 0.8 indicates an effect size that is highly practically significant. Only when a metabolite had a $p < 0.05$ and $d > 0.8$ were it seen as important.

$$d = \left| \tilde{y}_1 - \tilde{y}_2 \right| / SD_{max}$$

Equation C.1: Equation for determining effect size between two groups.

C.5.2 MULTIVARIATE STATISTICS

C.5.2.1 Principal Component Analysis (PCA)

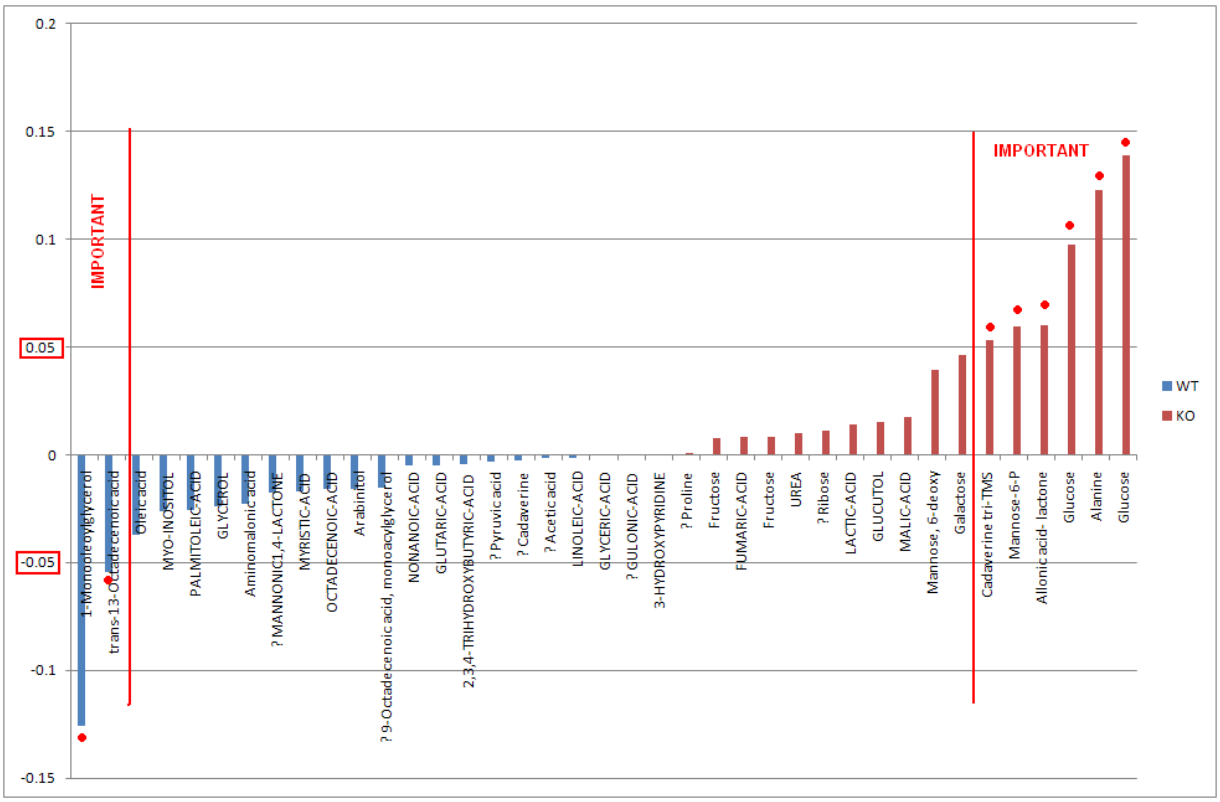


Figure C.4: An S-type bar plot of the average Euclidean distances determined from the loadings matrix. Cut-off lines for the selection of IMs were visually determined and are indicated by the vertical lines in this example.

PCA is perhaps one of the best known and most used chemometrics techniques (Sumner *et al.,* 2007) and was the preferred multivariate statistical method in this study. With the addition of PCA in the screening protocol, the unsupervised covariance of metabolites are analyzed which could add important information regarding the biological system. PCA of all pre-treated data sets was done in Matlab using the singular value decomposition algorithm provided. Scatter plots of the scores was then plotted to show grouping of experimental groups. Loadings of the score plots that showed adequate grouping was used to determine the influence of each variable to the scores

grouping. A ranking system was required to summarize the influence of a variable in several score plots seeing that four data pre-treatment methods was used per data set. The ranking system worked in the following way: loadings corresponding to score plots that gave noticeable grouping were exported to identify the most prominent IMs (Steuer *et al.,* 2007). The Euclidean distance of each loading was determined for each score plot. The average Euclidean distance of each loading was then used to rank the importance of the variables. Positive and negative signs were used to indicate loading direction. An S-type bar plot was created for these Euclidean distances to facilitate the identification of the most prominent variables (as shown in Figure C.4). Due to the nature of multivariate statistics, it does not contain rigid cut-off values like ANOVA and T-tests and is the selection of IMs commonly determined by visual selection. The most pertinent variables are thus visually identified as indicated in the example by selecting a 'visual cut-off'.

C.5.2.2 Partial least squares discriminant analysis (PLS-DA)

The inclusion of a supervised multivariate test is often advantageous as it could 'overlook' technical variance and find a discriminatory plane between the experimental groups, thereby giving more relevant information. PLS-DA was therefore included in the workflow to find important co-varying metabolites which PCA was not able to identify. This analysis was also performed with the use of the MetaboAnalyst web service (Xia *et al.,* 2009). Only log transformed data was used for PLS-DA. The VIP scoring results given by this service was not used as it ignores the grouping pattern of the PLS-DA. The Euclidean distance of the loadings was therefore also used in a similar way a described above.

C.6 ANNEXURE SUMMARY

In summary, this annexure described the establishment of a data processing protocol (workflow) for high-throughput data extraction, pre-processing, normalisation, pre-treatment and statistical analysis. An evaluation of these processes was required as they were not performed or documented at the platform where this study was conducted. This annexure also contains extra details for several of the steps taken and methods used which was not given in the previous chapters. Several GC-MS data extraction software was tested of which the combination of AMDIS and MET-IDEA was selected. Several common data pre-processing steps were incorporated into the data processing protocol and their validity to remove unwanted variance was confirmed. The inclusion of several data normalisation methods was motivated and the method of batch correction described. The use of univariate and multivariate statistics was motivated and the method by which loadings from different score plots were rank are also described.

ANNEXURE	RE-EVALUATING THE INFLUENCE OF MOISTURE IN DERIVATIZATION: DOES RESIDUAL WATER IN DRIED URINE AND TISSUE SAMPLES MATTER?
D	

D.1 INTRODUCTION

In addition to its comparatively good repeatability, organic acid extraction prior to GC-MS analysis is commonly preferred over other less-laborious methods mainly because all residual water is removed with the use of this method. The addition of sodium sulphate (Na_2SO_4) in the final steps of this method ensures the removal of any moisture. It is well known that moisture influences derivatization negatively, resulting in the incomplete and variable derivatization of compounds. Not only does it inactivate the derivatization agents, like BSTFA, it also results in the generation of polysiloxanes and highly toxic and corrosive hydrogen fluoride which can produce hydrogen gas when it comes into contact with metal (Dettmer *et al.,* 2007; Fiehn & Kind, 2007; Morgenthal *et al.,* 2007). It is believed that dried urine and especially dried tissue samples still contain some residual water. It is for this reason that tissue samples and crude sample preparation methods are not commonly preferred and their validity also questioned. Therefore, it was necessary to re-evaluate this concern and examine the degree by which moisture influence the derivatization process.

D.2 MATERIALS AND METHOD

The experimental procedures are shown in Figure D.1. A single urine sample was used to investigate the influence of moisture on derivatization along with an additional moisture removal step for crude sample preparation methods. For the crude, untargeted method, 100 µl urine was deproteinized with three volumes (300 µl) acetonitrile. This was done in triplicate. The samples were centrifuged for 5 minutes at 2000 x *g*, 4 ℃ and the precipitate discarded. The samples were dried, re-dissolved in 50 µl methanol and 50 µl chloroform and treated with a small spatula tip of sodium sulphate. The samples were mixed, the supernatant removed and dried before silylation.

The rest of the urine was used to extract organic acids as described in Section B.3.5. After the final drying step, water was added in increasing volumes so that the total percentage was in relation to the amount of silylation reagent to be added. The selected amounts of moisture (as indicated in the figure below) were expected to be well over the probable percentages in dried samples where no moisture removing step is included. Silylation was performed by adding 50 µl BSTFA. The samples

were incubated at 40 ℃ for an hour before GC-MS an alysis. All tests were done in triplicate. The analyses of these samples were done before the periodic GC column change to prevent possible damage of expensive new columns.

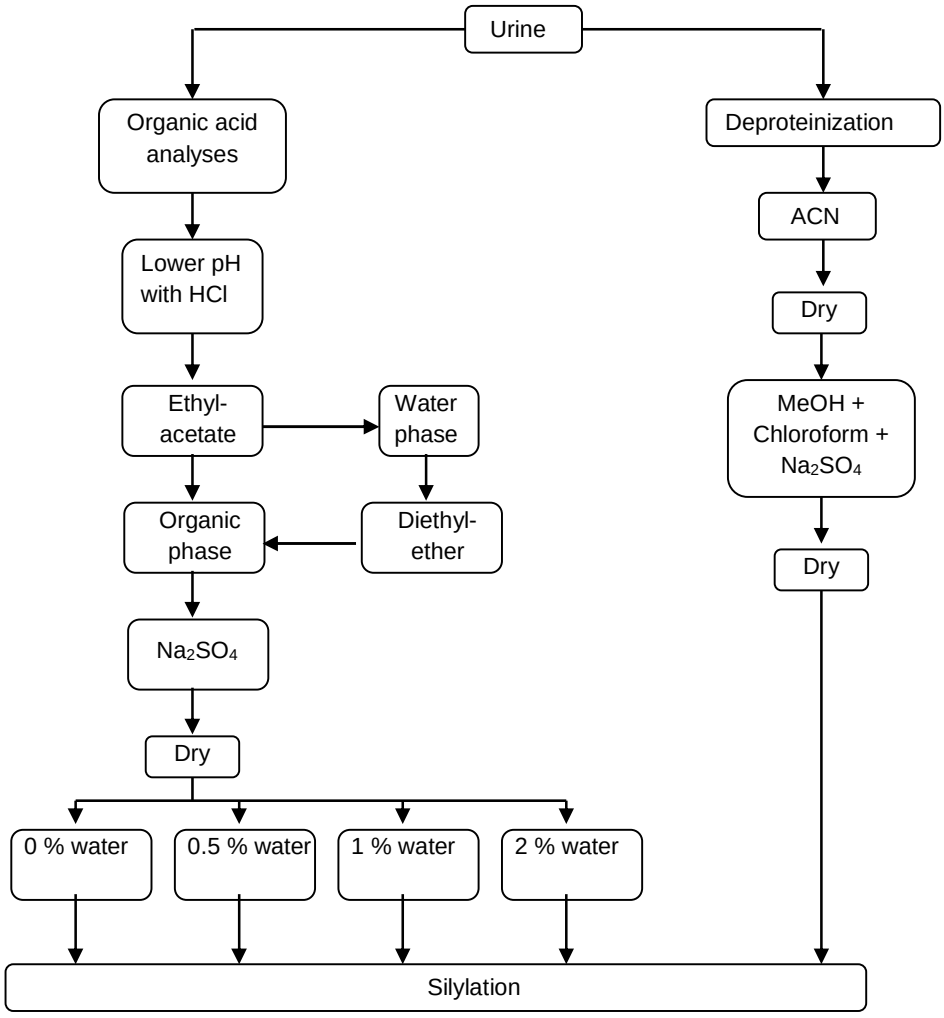


Figure D.1: Experimental strategy to study the influence of moisture on silylation. The inclusion of an additional moisture removing step in the crude, untargeted sample preparation protocol was tested as indicated on the right. The percentage of moisture to silylation agent was also tested to find a suitable cut-off and confirm the validity of crude sample preparation protocols in GC-MS analysis (steps on the left).

D.3 RESULTS AND DISCUSSION

The addition of a moisture removing step to the untargeted metabolomics protocol has the advantage of ensuring better silylation, but it prolongs the preparation time. Moreover, technical variance and sample loss is unavoidable in every additional sample preparation step. The latter

was especially seen when the results from the untargeted protocol with and without this additional step was compared (results not shown). Methanol and chloroform were used to re-dissolve the polar and apolar compounds, respectively. However, most of the sample loss could take place during this re-dissolving step, due to poor solubility of certain compounds, and when transferred. As a result, it is difficult to conclude whether moisture or sample loss is a greater threat to the experimental setup. This also applies to technical variance introduced in every additional preparation step versus variance from imperfect silylation. Because of this, the actual influence of moisture is better seen in the targeted protocol results.

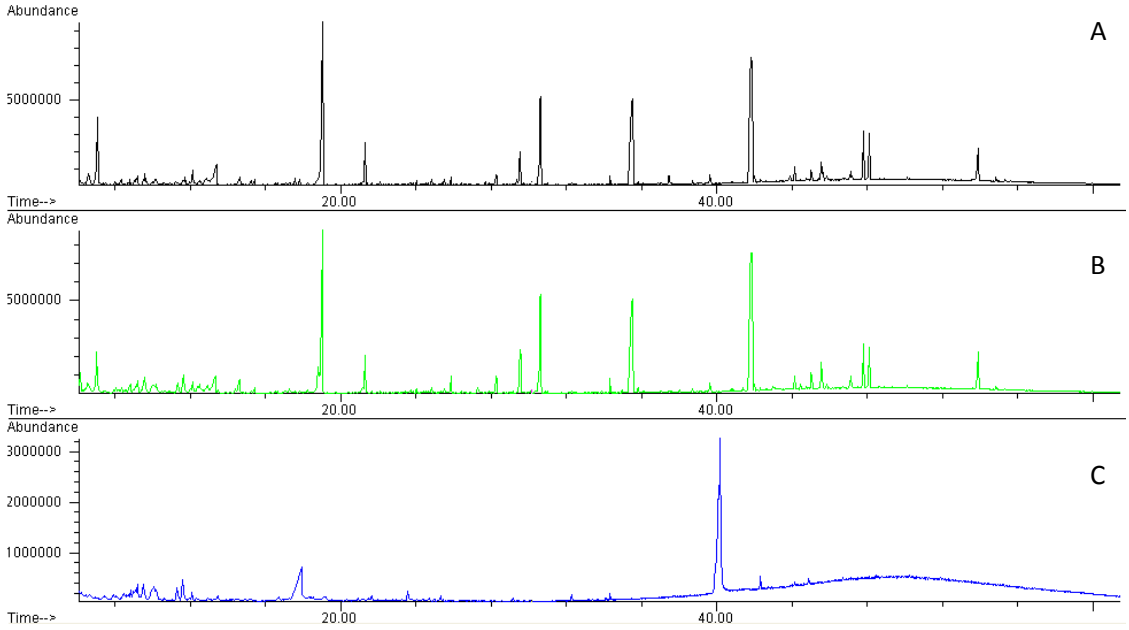


Figure D.2: Chromatograms showing the influence of moisture on silylation. Top chromatogram (A) was obtained from the sample containing no water. The lower two chromatograms, B and C, were obtained from the post-extracted samples containing 1 and 2 % (v/v) water, respectively, prior to derivatization and analysis.

Organic acid extraction is widely used at this institution (Centre for Human Metabonomics) not only for research but also for service purposes in the screening of inborn errors of metabolism (Reinecke *et al.,* 2011). The quality of results obtained by this method is known in terms of variance allowed and exometabolome coverage. Hence, this method was used to compare the influence of moisture (alone) on the results and exclude variance from sample handling. By assuming that the results are comparable after the extraction procedure, the variance thereafter

would then explain the influence of the added water. The chromatographic results of this experiment are shown in the Figure D.2.

Representative chromatograms of samples that contained 0, 1 and 2 % water, respectively, are shown in Figure D.2. The top two chromatograms are identical in terms of metabolome coverage and the intensity of the peaks. The lower chromatogram differs considerably from the others. It contains fewer peaks of which the intensities are also lower. These results therefore indicate that up to 1 % (v/v) water (1 µl water: 50 µl BSTFA) had no real effect on derivatization. The variance between the triplicates was also comparable to that of the sample without moisture and confirms that up to 1 % moisture has no real negative effect on silylation. However, derivatization was greatly hampered when the water content increased to > 1 %, giving incomplete and variable derivatization. Samples analyzed after the 2 % moisture-containing samples, were also normal which confirms that the effect seen is not due to column wear.

In conclusion it was confirmed that moisture influence derivatization negatively by giving incomplete and variable silylation which should be avoided. However, as it is very unlikely that any dried urine or tissue sample would even contain as much as 1 % (v/v) moisture, the need for a moisture-removing step is questionable (especially considering the sample loss and variance that can often be introduced). The reliability of simple, untargeted methods is comparable to that of established methods and was in fact even slightly better in this experiment in terms of variance.

ANNEXURE	BODY WEIGHT GAIN OF WT, MT1+2KO AND MT3KO MICE ON NORMAL AND HIGH-FAT DIET
E	

The following images are supplementary to the study of the high-fat diet induced differences between the WT, MT1+2KO and MT3KO mice. These figures were kindly provided by Prof. Juan Hidalgo (Institute of Neurosciences, Department of Cellular Biology, Physiology and Immunology, Faculty of Biosciences, Autonomous University of Barcelona, Spain. Email: Juan.Hidalgo@uab.es).

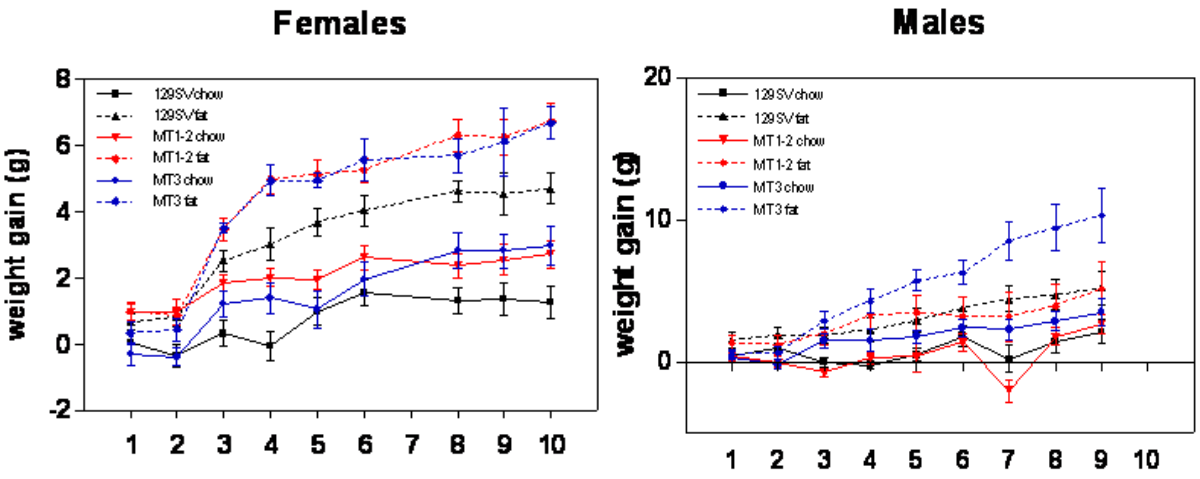


Figure E.1: Body weight gain of the WT, MT1+2KO and MT3KO mice that was fed a high-fat diet or a control (chow) diet. Measurements were taken every week (x-axis).

The body weight gain of the male and female WT, MT1+2KO and MT3KO mice are indicated in Figure E.1. The solid lines represent the mice that were fed a normal chow diet. The perforated lines represent the mice that were fed a high-fat diet. The body weight gain patterns of the control group indicate that the knockout mice (MT1+2KO and MT3KO) gained more weight in comparison to the WT except for the MT1+2KO male mice which were very similar to the WT. This is in agreement with reports of Beattie *et al.* (1998) and Byun *et al.* (2011) which described the tendency of these mice to become obese. The body weight gain of all three strains that were fed the high-fat diet was clearly greater than the control, as would be expected. The male MT3KO

mice and the female MT1+2KO and MT3KO mice clearly gained more body weight than other mice on the high-fat diet.

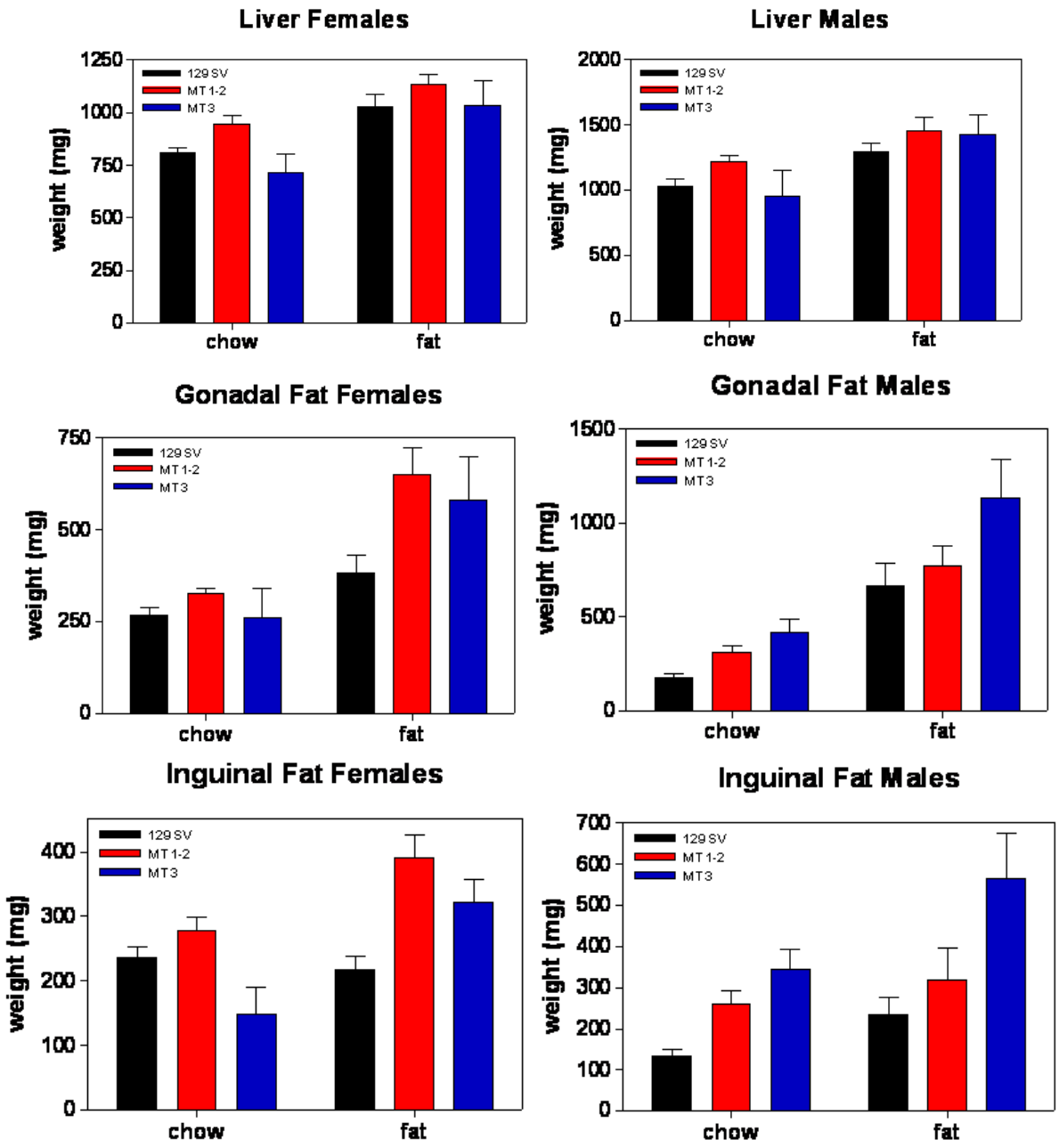


Figure E.2: Absolute weights of liver, gonadal and inguinal fat depots from the WT, MT1+2KO and MT3KO mice. The x-axis signifies the diet.

The absolute liver weights of the male and female mice that received the normal chow diet were very similar with the livers of the MT1+2KO mice having slightly more weight. This pattern in the absolute weights stayed the same in the mice that received the high-fat diet. The MT1+2KO female mice and MT3KO male mice had more gonadal and inguinal fat deposits with the normal chow and high-fat diet.

ANNEXURE F

A COLLABORATIVE INVESTIGATION OF THE ROLE OF MTs IN MITOCHONDRIAL FUNCTION AND DISEASE: THE ROTENONE STUDY

F.1 DESCRIPTION OF THE COLLABORATIVE STUDY

A large collaborative study was designed to comprehensively study the involvement of MTs in mitochondrial disease using a complex I inhibited-MT1+2KO mouse model. A wide range of assays were used to study the involvement of MTs in every functional molecular level i.e. DNA-, protein (enzyme)- and metabolite level. A schematic presentation of this collaborative investigation is shown in Figure F.1. It is not within the scope of this annexure to discuss all the assays in detail but merely to list them for referral to the results obtained from these tests. The focus of the metabolomics and oxidative stress investigations (shaded area in Figure F.1) were to detect differences in the small molecules (metabolites and ROS). The activity of all the OXPHOS enzymes including several antioxidant enzymes (such as glutathione peroxidase, superoxide dismutase, etc) was studied. Moreover, the native OXPHOS enzyme structures and ratios were studied with blue native polyacrylamide gel electrophoresis (BN-PAGE) and western blotting (Pretorius, 2011). The mtDNA copy numbers were measured and the DNA repair capabilities of these mice compared (Steenkamp, 2010). The degree of DNA methylation was also measured for comparison (unpublished).

F.2 SUPPLEMENTARY DATA SHOWING COMPLEX I INHIBITION

The effectiveness of the rotenone treatment and the selected model was confirmed with the analysis of complex I activity in several organs. Details regarding the materials and methods used and discussion of the results can be found in the dissertation of Pretorius, (2011). The liver and brain did not show significantly reduced activity of complex I. However, complex I was significantly inhibited in the heart and skeletal muscle as indicated in Figure F.2 and F.3. The heart muscle of both WT and MT1+2KO mice showed significant inhibition in comparison to the vehicle control group with p-values of 0.000001 and 0.02, respectively. Complex I activity was reduced by ~24 % in the WT mice and ~8 % in the MT1+2KO mice. Furthermore, the skeletal muscle of both WT and MT1+2KO mice also showed significant inhibition in comparison to the vehicle control group with p-values of 0.01 and 0.02, respectively. Complex I activity was reduced by ~34 % in the WT mice and ~20 % in the MT1+2KO mice.

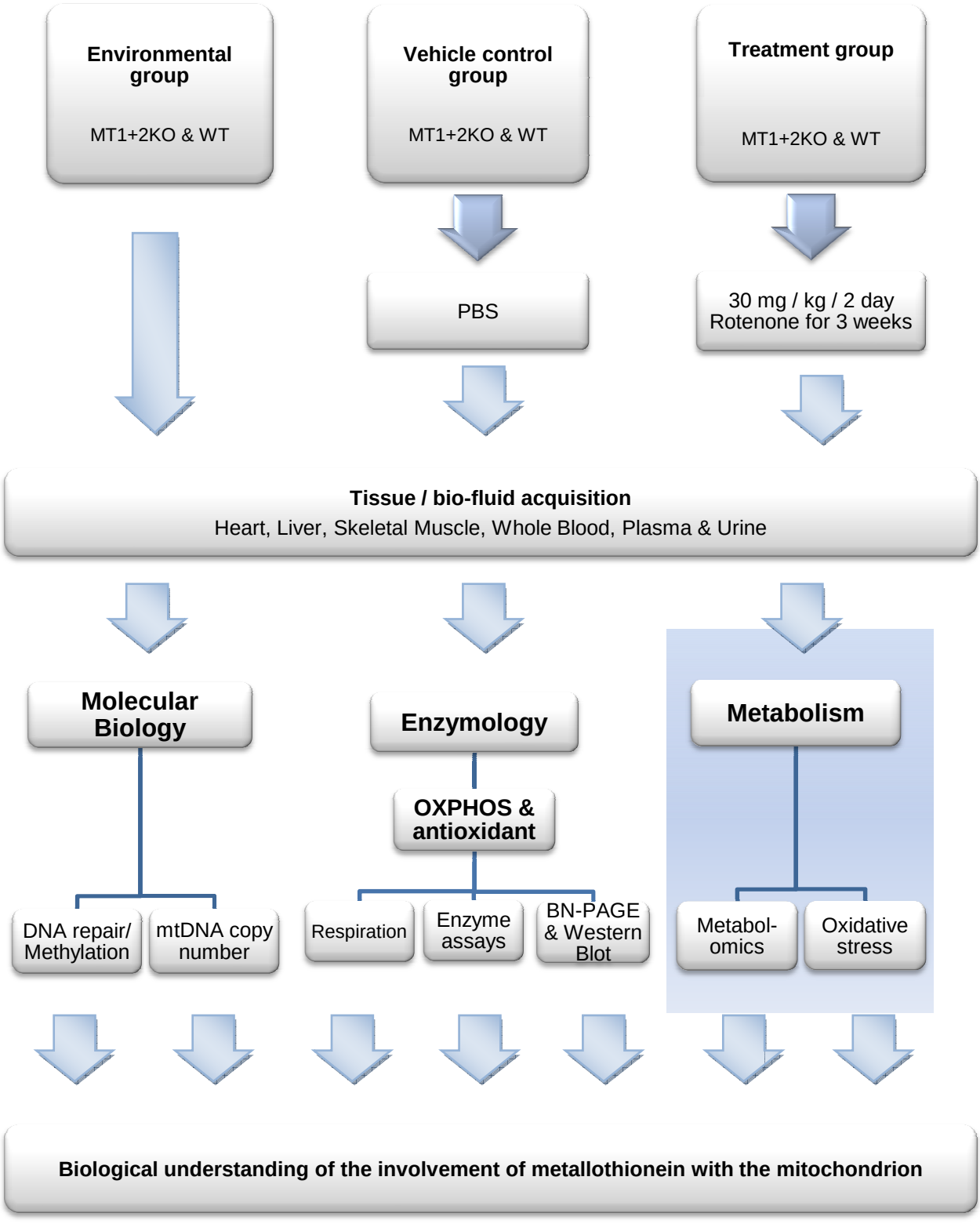


Figure F.1: Schematic presentation of the strategy designed to comprehensively investigate the role of MTs in mitochondrial disease (complex I deficiency).

From the results obtained it is clear that the skeletal muscle and heart was most sensitive to the rotenone treatment while liver and the brain did not give satisfactory inhibition. The absorption, transport and metabolism of the administered rotenone might play a role in this selective

inhibition of organs, especially when considering that the liver is the main degradation site of rotenone with the use of the cytochrome p450 enzymes system. The vehicle used can also play a role. In previous studies, much lower dosages of rotenone suspended in peanut oil were used to treat the mice (van Zweel, 2010). However, these results indicated that the vehicle had a compounding effect as rotenone suspended in PBS did not give the same toxicity. Taking into account that numerous other researchers use dimethyl sulphoxide (DMSO) and other oils as vehicle, it is fair to conclude that some of the effects recorded might be due to a compounding effect of the rotenone and this vehicle. Moreover, the inhibition of complex I has down-stream effects such as the up- or down regulation of specific genes to cope with the reduced respiration activity. One seemingly possible result of the three week rotenone treatment is the over-production of complex I to counter the lower activity (Figure F.4). The higher amount of complex I would therefore also explain the relatively small inhibition that was found in the liver and brain.

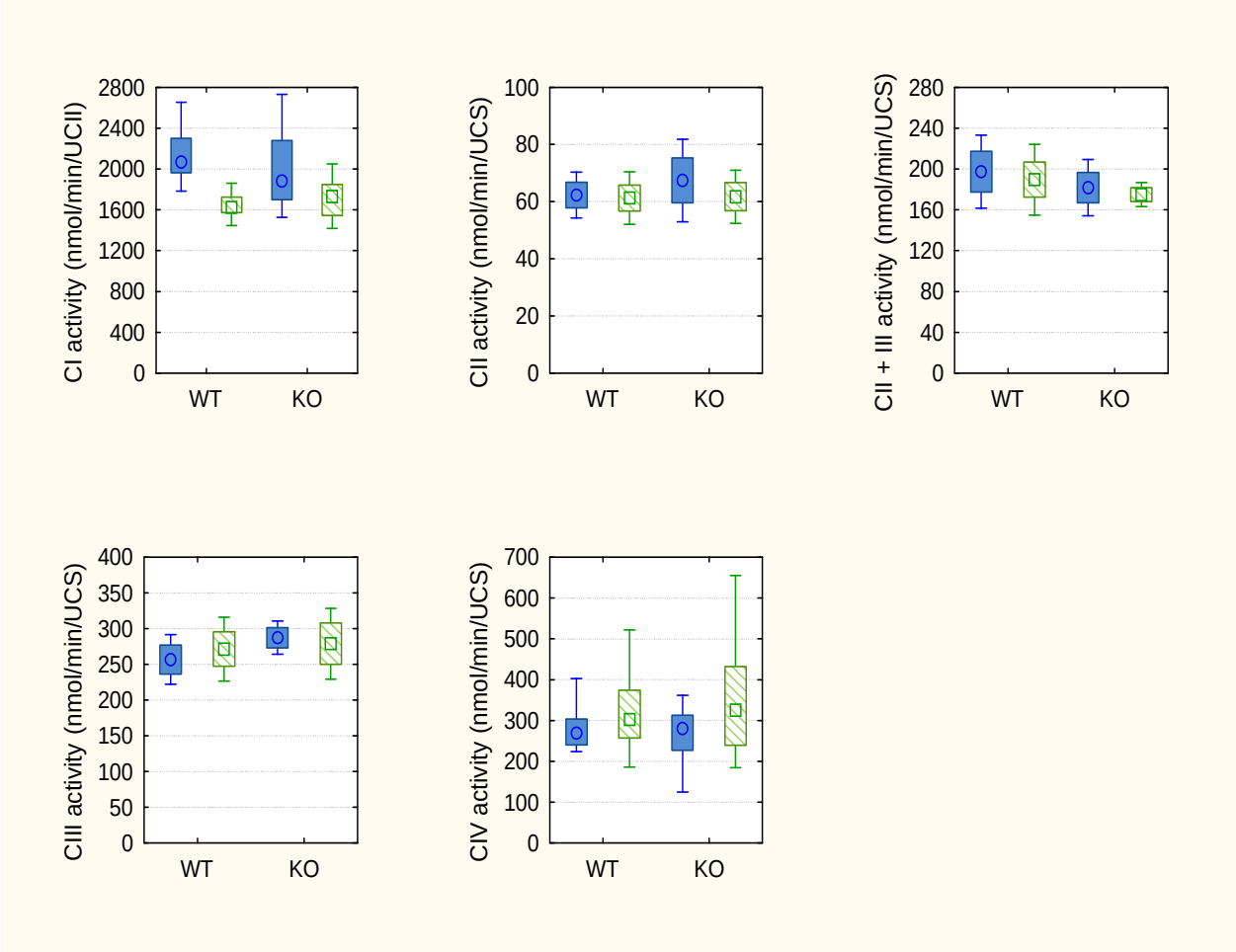


Figure F.2: Box and whiskers plots showing the different OXPHOS enzyme activities in the heart. The top left plot shows complex I activity in the WT and MT1+2KO mice treated with PBS (shaded box) or rotenone (clear box).

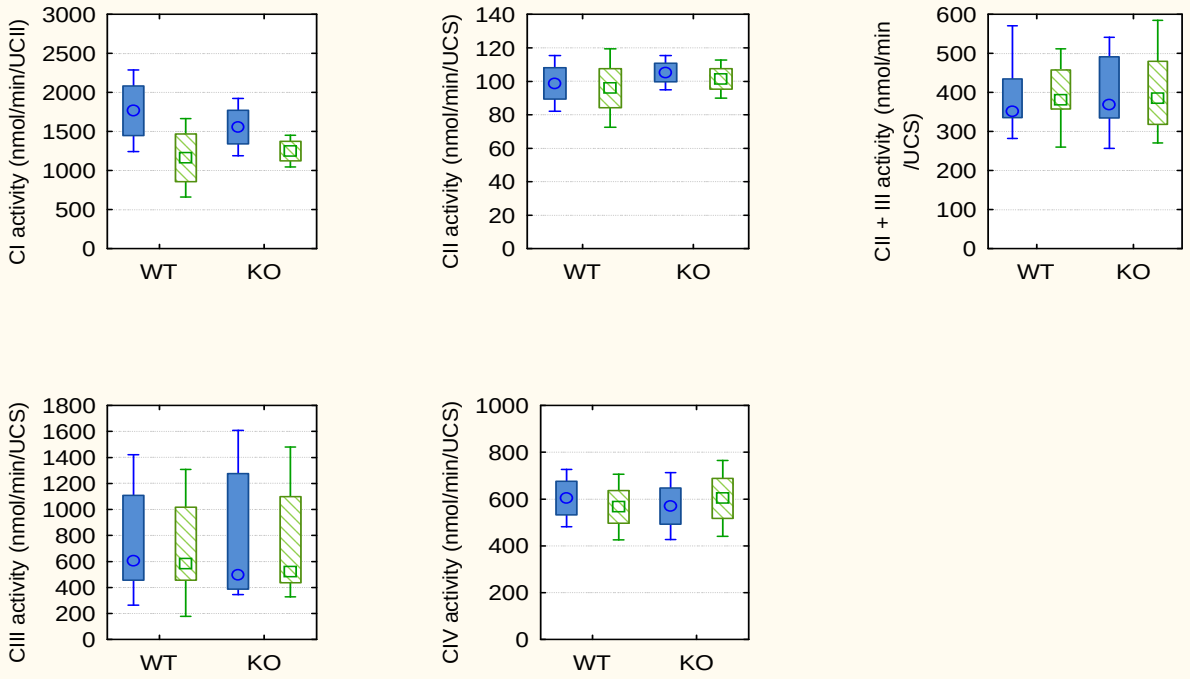


Figure F.3: Box and whiskers plots showing the different OXPHOS enzyme activities in the skeletal muscle. The top left plot shows complex I activity in the WT and MT1+2KO mice treated with PBS (shaded box) or rotenone (clear box).

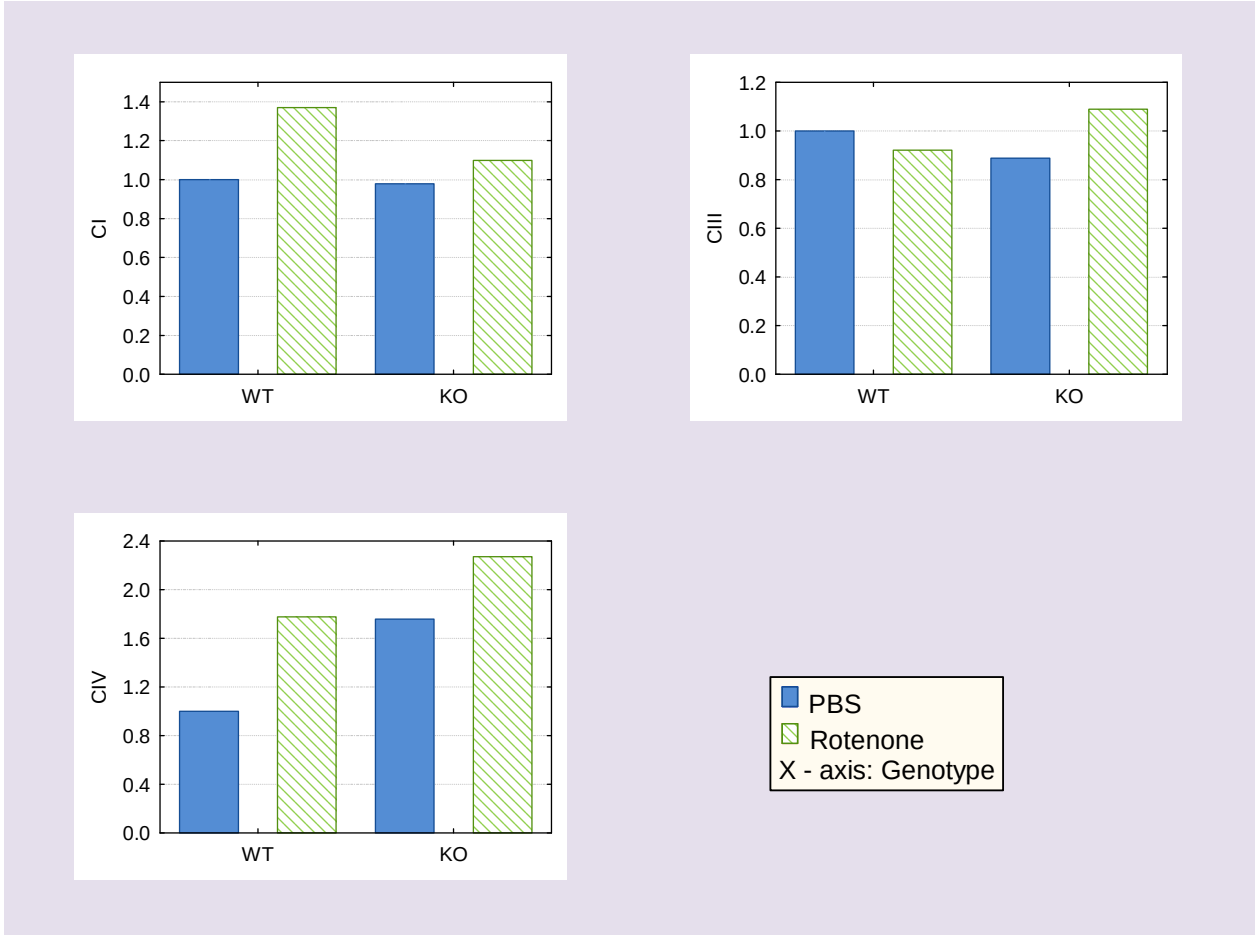


Figure F.4: Bar plot presentation of the enzyme quantities as determined using BN-PAGE and western blot analyses of heart mitochondrial fractions.

ANNEXURE	
G	**SUPPLEMENTARY MATERIAL: DVD**

ANNEXURE	
H	**PUBLICATIONS AND CONFERENCE PROCEEDINGS**

The findings or methodologies developed during this study (or parts thereof) were included as part of the following publications and international conferences:

Lindeque JZ, Levanets O, Louw R, van der Westhuizen FH. 2010. The involvement of metallothioneins in mitochondrial function and disease. Current Protein & Peptide Science 11, 292-309 (Impact factor 3.01).

Reinecke CJ, Koekemoer G, van der Westhuizen FH, Louw R, Lindeque JZ, Mienie LJ, Smuts I. Metabolomics of urinary organic acids in respiratory chain deficiencies. Metabolomics (Impact factor 3.87). In print, DOI 10.1007/s11306-011-0309-0.

FH van der Westhuizen, JZ Lindeque, ZAO van Zweel, J Smeitink, J Hidalgo, R Louw. Metallothionein involvement in the heart pathology of rotenone-induced complex I deficient mice. 7th Euromit Conference, Stockholm, Sweden, June 11 – 15, 2008. Poster presentation.

Pretorius M, Lindeque JZ, Steenkamp A, Dreyer W, Bothma K, Crous A, Pretorius PJ, Louw R, van der Westhuizen FH. The investigation of metallothionein involvement in mitochondrial function and disease. 7th Annual Meeting of ASMRM and 10th J-mit in Fukuoka, Japan, Dec 15-18, 2010. Poster presentation.

Louw R, Lindeque JZ, Pretorius M, Dreyer W, Bothma K, Crous A, van der Westhuizen FH. The involvement of metallothioneins in mitochondrial function and disease. 8th Euromit Conference, Zaragoza, Spain, 20-23 June, 2011. Oral Presentation.

van der Westhuizen FH, Lindeque JZ, Hidalgo J, Pretorius M, Louw R. The involvement of metallothioneins in mitochondrial function and disease. 8th Conference on Mitochondrial Physiology (MiP2011), Bordeaux, France, 5-8 September, 2011. Oral Presentation.

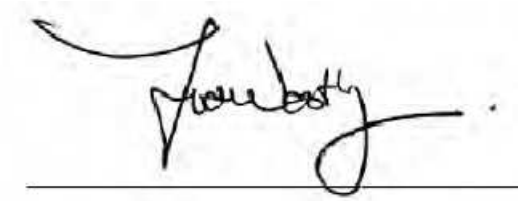
ANNEXURE	CO-AUTHOR CONSENT
I	

To whom it may concern,

We, the co-authors of the review paper titled "**The involvement of metallothioneins in mitochondrial function and disease (Current Protein and Peptide Science, vol. 11, 2010)**" hereby give permission that this paper, as presented in Chapter 2, may be included as part of the thesis submitted for the degree *Philosophiae Doctor*.

Dr. Oksana Levanets,
Centre for Human Metabonomics,
North-West University,
Potchefstroom,
South Africa

Dr. Roan Louw,
Centre for Human Metabonomics,
North-West University,
Potchefstroom,
South Africa



Prof. Francois van der Westhuizen,
Centre for Human Metabonomics,
North-West University,
Potchefstroom,
South Africa