The identification of genetic variation in the acyl-CoA synthetase genes (ACSM2A and ACSM2B)

T van der Westhuizen

22053298

Dissertation submitted in partial fulfillment of the requirements for the degree *Magister Scientiae* in *Biochemistry* at the Potchefstroom Campus of the North-West University

Supervisor: Ms R van der Sluis

Co-supervisor: Dr E van Dyk

May 2016

“Don’t say you don’t have enough time. You have exactly the same number of hours per day that were given to Helen Keller, Pasteur, Michelangelo, Mother Teresa, Leonardo da Vinci, Thomas Jefferson, and Albert Einstein.”

– H. Jackson Brown Jr.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to the following persons and institutions for their contributions and support towards the completion of this study:

- Firstly, I would like to express my sincere gratitude to my supervisor, Dr. Rencia Van Der Sluis and co-supervisor, Dr. Etresia Van Dyk for the continuous support of my M.Sc. study and related research, for their patience, motivation, and immense knowledge.

- I would like to thank Prof Albie van Dijk, for her guidance through this process.

- I thank my fellow labmates and friends for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last two years.

- My parents, and family for all their love, moral support and motivation.

- National Research Foundation (NRF) for their financial support during this study (Thuthuka grant reference number: TTK20110803000023154).

- I am grateful to God for the good health and wellbeing that were necessary to complete this dissertation.

ABSTRACT

Investigations into the role of the glycine conjugation pathway and specifically the functioning and significant importance of the mitochondrial medium-chain ligases have seriously been neglected over the last 30 years. The metabolism of drugs and benzoate to acyl coenzyme A (CoA) intermediates in humans has increased in recent times as the exposure to environmental factors, nutrition and the chronic use of medication rises. Seeing that no defect of the glycine conjugation pathway has been reported thus far, it can be assumed that this pathway is essential for survival. In this study the question was raised on whether the human acyl CoA synthetase medium-chain family member 2B (*ACSM2B)* open reading frame (ORF) is also as highly conserved as previously reported for the Glycine N-acyltransferase (*GLYAT*) ORF. It was hypothesised that genetic variation in the ORF of the *ACSM2B* gene should be low. However, focus was also on the human acyl CoA synthetase medium-chain family member 2A (*ACSM2A)* gene as there is some confusion in current literature regarding the distinction of these two highly similar genes. The hypothesis was investigated by analysing the genetic variation data across 6 different population groups (AFR; EUR; EAS; AMR; AA; EA) of the *ACSM2A* and *ACSM2B* ORF available on public databases, along with the coding region of a small cohort of South African Afrikaner Caucasian individuals that was sequenced. The *ACSM2A* and *ACSM2B* ORF of 8537 individuals (consisting of data aquired from the 1000 Genomes Project, NHLBI ESP and South African Afrikaner Caucasian) analysed, identified genetic variants at a low frequency (%) and mostly occurring only as heterozygotes and in a single population group. Of the 47 (1000 Genomes Project), 15 (National Heart, Lung and Blood Institute Exome Sequencing Project, NHLBI ESP), and 4 (South African Afrikaner Caucasian Population, SA) non-synonymous SNPs identified within the coding region of the *ACSM2A* gene, the L64P variant had the highest homozygous SNP genotype frequency (29.0%), followed by the N463D (12.5%), and the R5Q variant (5.5%). All other variants were found at frequencies <5%. Of the 43 (1000 Genomes Project), 15 (National Heart, Lung and Blood Institute Exome Sequencing Project, NHLBI ESP), and 1 (South African Afrikaner Caucasian Population, SA) non-synonymous SNPs identified within the coding region of the *ACSM2B* gene, the T278A variant had the highest homozygous SNP genotype frequency (4.0%), followed by the I305V variant (0.7%), and the D322N variant (0.1%). The results of this study indicated that the acyl CoA synthetase gene *ACSM2B* ORF is not as highly conserved as the *GLYAT* ORF, as *ACSM2B* is not part of the evolutionary path of polyphenol biotransformation. However, it is evident from this study that very low genotype frequencies exist for the SNPs identified within the coding region of the *ACSM2B* gene (T278A: 4.0%, I305V: 0.7%, D322N: 0.1%)

compared to genotype frequencies identified for *GLYAT* (N156S: 90%; S17T: 4.6%; R131H: 0.1%) from a study conducted by Van der Sluis *et al*., (2015). The results of this study indicated that the acyl CoA synthetase genes (*ACSM2A* and *ACSM2B*) ORF is relatively conserved and that the current reference sequence used in the present study for the *ACSM2A* and *ACSM2B* genes should probably be considered as the wild-type. With increased levels of benzoic acid exposure in humans, the HXM-A protein (encoded by *ACSM2B*) might not be able to effectively detoxify such large amounts. Thus, findings underline the importance of future investigations into the *ACSM2A* and *ACSM2B* genes, and their proteins to better understand the effect of SNPs on protein function. This study also contributed significantly to a better understanding of the nomenclature regarding the acyl CoA synthetase genes, especially confusion surrounding the *ACSM2A* and *ACSM2B* genes as several discrepancies in the literature were pointed out.

Key Words:

Acyl CoA synthetase genes; acyl CoA synthetase medium-chain family member 2A, acyl CoA synthetase medium-chain family member 2B; conserved open reading frame; detoxification; xenobiotics; benzoate; Afrikaner Caucasian; Single nucleotide polymorphism

TABLE OF CONTENTS:

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS:

°C	degrees Celsius
1000G	1000 Genomes project
AA	African American
Abs	absorbance
ACS	acyl CoA synthetases
ACSM	acyl CoA synthetase medium-chain family member
ACSM2A	acyl CoA synthetase medium- chain family member 2A
ACSM2B	acyl CoA synthetase medium- chain family member 2B
ACSVL	very long-chain fatty acid
AFR	African
Ala	alanine
AMP	adenosine monophosphate
AMR	American
Arg	arginine
Asp	aspartic acid
ATP	adenosine triphosphate
BLAST	Basic Local Alignment Search Tool algorithm
bp	base paire
$C_9H_9NO_3$	hippuric acid
CAF	Central Analyitical Facility
CASTOR	Coenzyme A sequestration, toxicity or redistribution
cDNA	complementary DNA
CEC	capillary electrochromatography
CoA	coenzyme A

$-COO_2H$	carboxyl group
Cys	cysteine
CZE	capillary zone electrophoresis
dbSNP	SNP database
DME	drug metabolising enzymes
DNA	deoxyribonucleic acid
e^-	electrons
e.g.	for example
EA	East Asian
EC	Enzyme Commission
ER	endoplasmic reticulum
ESP	Exome Sequencing Project
ETC	electron transport chain
EU	European American
EUR	European
g	gram
GC	gas chromatography
gDNA	genomic DNA
Gln	glutamine
Glu	glutamic acid
Gly	glycine
GLYAT	glycine N-acyltransferase
GST	glutathione-S-transferase
H^+	hydrogen ion; proton
H_2O	water
HGVS	Human Genome Variation Society

His	histidine
Hmz	homozygote
HPLC	high performance liquid chromatography
HPLC-MS	high performance liquid chromatography mass spectrometry
Htz	heterozygote
HUGO	Human Genome Organization
HXM-A	xenobiotic/medium chain fatty acid: CoA ligase A
HXM-B	xenobiotic/medium chain fatty acid: CoA ligase B
ID	identification
Ile	isoleucine
kDa	kilodalton
Leu	leucine
Lys	lysine
Mb	unit of length for DNA fragment of 1 million nucleotides
Met	methionine
mMg	milligram
min	minute
ml	millilitre
mM	millimolar
NAD^+	nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NAT	N-acyltransferase
NCBI	National Center for Biotechnology Information
NGS	next-generation sequencing
$-NH_2$	amino group

NHLBI ESP	National Heart, Lung and Blood Institute Exome Sequencing Project
NWU	North-West University
OMIM	Online Mendelian Inheritance in Man
ORF	open reading frame
CYP450	cytochrome P-450s
PGM	Personal Genome Machine
P-gp	permeability glycoprotein
Phe	phenylalanine
PPi	pyrophosphate
Pro	proline
RefSNP	reference SNP
RNA	ribonucleic acid
rs	reference SNP cluster identification number
s	seconds
SA	South African Afrikaner Caucasian Population
Ser	serine
-SH	sulfhydryl group
SNP	single nucleotide polymorphism
SNV	single nucleotide variation
SPE	solid phase extraction
SULT	sulfotransferase
TCA	tricarboxylic acid
Thr	threonine
Trp	tryptophan
UGT	UDP-glucuronosyltransferase

Val	valine
VCF	variant caller file
VEP	variant effect predictor
XM-ligases	xenobiotic medium-chain ligases
α	alpha
β	beta
µg	microgram
µl	microliter
µM	micromolar

CHAPTER 1: INTRODUCTION

1.1 Introduction to the present study

It is evident from the literature to date that there is a direct correlation between an impaired detoxification system's inability to excrete xenobiotics (foreign, exogenous compounds), and multi-factorial and complex diseases. A variety of diseases associated with an impaired detoxification system have been identified and include diseases such as cancer, systemic lupus erythematosus, Parkinson's disease, fibromyalgia, and chronic fatigue/immune dysfunction syndrome (Kawajiri *et al*., 1990, Bandmann *et al*., 1997; Gronau *et al*., 2003; Liska *et al*., 2006; Badenhorst *et al*., 2014). As humans, we are exposed to a great amount of toxins throughout our lifetime. It is imperative that detoxification systems exist to minimize the potential damage from xenobiotics. Detoxification reactions primarily consists of two phases, Phase I (the functionalization phase), and Phase II (the conjugation phase). Phase II drug metabolising enzymes (DMEs) contribute significantly to the metabolism of large amounts of foreign compounds. DMEs are responsible for the regulation of metabolism and the disposition of various endogenous substances and also contribute significantly to maintaining homeostasis in the human body (Jančová and Siller, 2012).

This study will focus mainly on Phase II DMEs, specifically the mitochondrial medium-chain acid: CoA ligases (ACSM). ACSM ligases are primarily associated with metabolism via amino acid conjugation, related to the activation of substrates such as benzoic acid as well as salicylic acid (Knights, 1998, Van der Sluis *et al*., 2015). Amino acid conjugation of various substrates occurs through two steps i) the formation of acyl CoA as an obligatory step in the metabolism of a wide range of endogenous substrates, as well as fatty acids. The first step is catalysed by ATP-dependent acid:CoA ligase, HXM-A (EC 6.2.1.1-2.1.3; AMP forming; encoded by the *ACSM2B* gene). The second step involves the linkage of the activated acyl group via an acyl CoA: amino acid N-acyltransferase (GLYAT) (EC 2.3.1.13) to the amino group of the acceptor amino acid.

To date, very little information and in-depth understanding exist around the influence of genetic variation on the glycine conjugation pathway, specifically variations in mitochondrial medium-chain acid: CoA ligases (Knights *et al*., 2007). Conjugation reactions (such as the glycine conjugation of benzoic acid to yield hippuric acid) play an essential role in the toxicity of many chemicals, because the formation of a xenobiotic acyl CoA thioester during the first step of amino acid conjugation is an obligatory step (Jančová and Siller, 2012). Hippuric

acid, the glycine conjugate of benzoic acid, is the third most abundant metabolite found in urine after creatinine and urea (Lees *et al*., 2013). Benzoic acid is widely used in pharmaceuticals as well as preservatives in food, and can be as high as 280% of the daily recommended allowance (Piper, 1999, Nair, 2001, Tfouni and Toledo, 2002, Lees *et al*., 2013, Van der Sluis *et al*., 2015). Hippuric acid is a hydrophilic metabolite excreted in the urine of humans and other mammals, and exists as a product of the metabolic interaction between a mammalian host and the microorganisms inhibiting its gastrointestinal tract (Wikoff *et al*., 2009; Williams *et al*., 2010; Lees *et al*., 2013; Badenhorst *et al*., 2014). Interindividual variation in the glycine conjugation pathway contributes greatly to the metabolic formation of toxic metabolites such as reactive electrophiles (Jančová and Siller, 2012, Knights *et al*., 2007).

A previous study by Van der Sluis *et al*., (2015) on the conservation of the coding regions of the *GLYAT* gene, aimed to characterise the genetic variation in the *GLYAT* ORF. Results indicated that the ORF of *GLYAT* is highly conserved, suggesting that the glycine conjugation pathway is an essential detoxification pathway as no defects to date have been described on this pathway (Van der Sluis *et al*., 2015)

This study was formulated in order to identify whether the ORF of the *ACSM2B* gene is also as highly conserved as previously reported for the *GLYAT* ORF (Badenhorst *et al*., 2014; Van der Sluis *et al*., 2015). It was hypothesised that genetic variation in the ORF of the *ACSM2B* gene should be low. This was accomplished by analysing not only the genetic variation data of the *ACSM2B* gene, but also the *ACSM2A* gene, available on the 1000 Genomes Project, and the NHLBI ESP public databases. The exomes of a small cohort of South African Afrikaner Caucasian individuals was also sequenced by a high throughput target enrichment sequence strategy to identify known or possible novel *ACSM2A* and *ACSM2B*-associated gene variants, since genetic variation data of a particular gene within the Afrikaner Caucasian population are not available on the previously mentioned databases (Kruse *et al*., 2009; Xu *et al*., 2009; Xu *et al*., 2011; Xu *et al*., 2012; Heathfield *et al*., 2013, Rodriguez-Murillo *et al*., 2014; Van der Sluis *et al*., 2015).

The data obtained from the non-synonymous SNPs identified within the coding region of the *ACSM2A* and *ACSM2B* genes were then compared and analysed in all population data available in this study (1000 Genomes Project data, NHLBI ESP data, and SA data) to determine the level of genetic variation in these genes. The main outcomes of this study contributes to the identification of genetic variations in the acyl CoA synthetase genes (*ACSM2A* and *ACSM2B)* and ultimately provides insight into the role of the glycine

conjugation pathway and improves future investigations and general understanding regarding these genes (*ACSM2A* and *ACSM2B*).

CHAPTER 2: LITERATURE OVERVIEW

2.1 Introduction

In this chapter an overview of the different detoxification systems present in humans will be discussed. The focus will be on Phase II detoxification systems, specifically the mechanisms and importance of the glycine conjugation pathway. Included in Chapter 2 will be a discussion on acyl CoA metabolism and toxicity, and its central role in the glycine conjugation pathway, together with a thorough investigation on the acyl CoA synthetase family and the nomenclature generating such confusion in current literature. Chapter 2 will focus especially on the acyl CoA synthetase medium-chain family members, *ACSM2A* and *ACSM2B*. This overview will form the basis and motivation behind the study, including the methods used to accomplish this study. The problem statement, aim, objectives and experimental strategy are also presented in Chapter 2.

2.2 Introduction to the detoxification system

Throughout one's lifetime, the human body is constantly and unavoidably exposed to a number of compounds, natural or man-made (Parkinson, 2001; Liska *et al*., 2006; Jančová and Siller, 2012; Gonzales *et a*l., 2006). Due to the foreign nature of these compounds to metabolism they are defined as xenobiotic compounds. Xenobiotics comprise a wide variety of major categories such as: drugs, active ingredients in pesticides, pharmaceutical products, food constituents, food additives (with nutritional value/ preservatives), agrochemicals, industrial chemicals, secondary plant metabolites and pollutants (Parkinson, 2001; Ioannides, 2002; Liska *et al*., 2006; Knights *et al*., 2007; Murphy, 2008; Jančová and Siller, 2012; Gonzales *et al*., 2006). Lipophilicity (affinity for lipids) is a physical property of most xenobiotic compounds enabling them to penetrate the lipid membranes of cells, to be transported by lipoproteins in the blood, and to be rapidly absorbed by a target organ, consequently complicating their elimination (Parkinson, 2001). As a result, the elimination of xenobiotics from the body depends on their conversion to more polar, hydrophilic, readily excretable products (Parkinson, 2001; Jančová and Siller, 2012). A human's ability to

metabolize and clear drugs and other foreign compounds is a natural process (Gonzales *et al*., 2006). Humans have a well-developed xenobiotic biotransformation system, which transforms and removes toxic and foreign compounds from the body as hydrophilic metabolites eliminated through the urine or bile (Parkinson, 2001; Liska *et al*., 2006; Jančova and Siller, 2012; Gonzales *et al*., 2006). Xenobiotic biotransformation is a crucial and complex organisation of reactions that consists of the same enzymatic pathways and transport systems that are utilized for normal metabolism of dietary constituents (Parkinson, 2001; Liska *et al*., 2006; Gonzales *et al*., 2006).

Table 2.1: Comparison between Phase 0, I, II and III biotransformation reactions

Phases of biotransformation	**Phase 0** **(Absorption)**	**Phase I** **(Bioactivation or inactivation)**	**Phase II** **(Inactivation)**	**Phase III** **(Elimination)**
Types of reactions	Membrane transporters (Example: ATP binding cassette, (ABC) transporters)	Hydrolysis, Oxidation, Reduction (Example: cytochrome P450 system)	Conjugation reactions (Example: amino acid conjugation)	Membrane transporters (Example: ATP binding cassette, (ABC) transporters)
Mechanism	Transport across cell membrane	Exposes the functional group. An atom of oxygen is incorporated into the chemical	Adds polar compound(s) to the functional group	Transport across cell membrane
Outcome	Prevent accumulation of harmful toxicants inside the cells	May result in metabolic activation/ prepare the compound for Phase II metabolism/ may be equally or more active than the parent compound	Facilitates excretion/ highly polar compound, excretable in urine/ usually results in an inactive compound	Play crucial roles in drug distribution, and final elimination of the compound/ metabolites from the body

Biotransformation can take place in most tissues in the body, yet the liver is the primary site. The biotransformation of xenobiotics and endogenous compounds are mainly divided into four phases: Pase 0, Phase I, Phase II and Phase III, and are primarily metabolized by four different types of reactions: oxidation, reduction, hydrolysis, and conjugation, however, Phase 0 and Phase III requires the use of transporter proteins as opposed to detoxification enzymes (Table 2.1) (Figure 2.1) (Parkinson, 2001; Xu *et al*., 2004; Liska *et al*., 2006,

Parkinson and Ogilvie, 2008; Jančová and Siller, 2012; Penner *et al*., 2012). Detoxification enzymes exist in the smooth endoplasmic reticulum (ER), cytosol (intracellular fluid) and to a lesser degree in the membranes of the mitochondria, nuclei and lysosomes (small spherical organelles) of the liver's hepatocyte

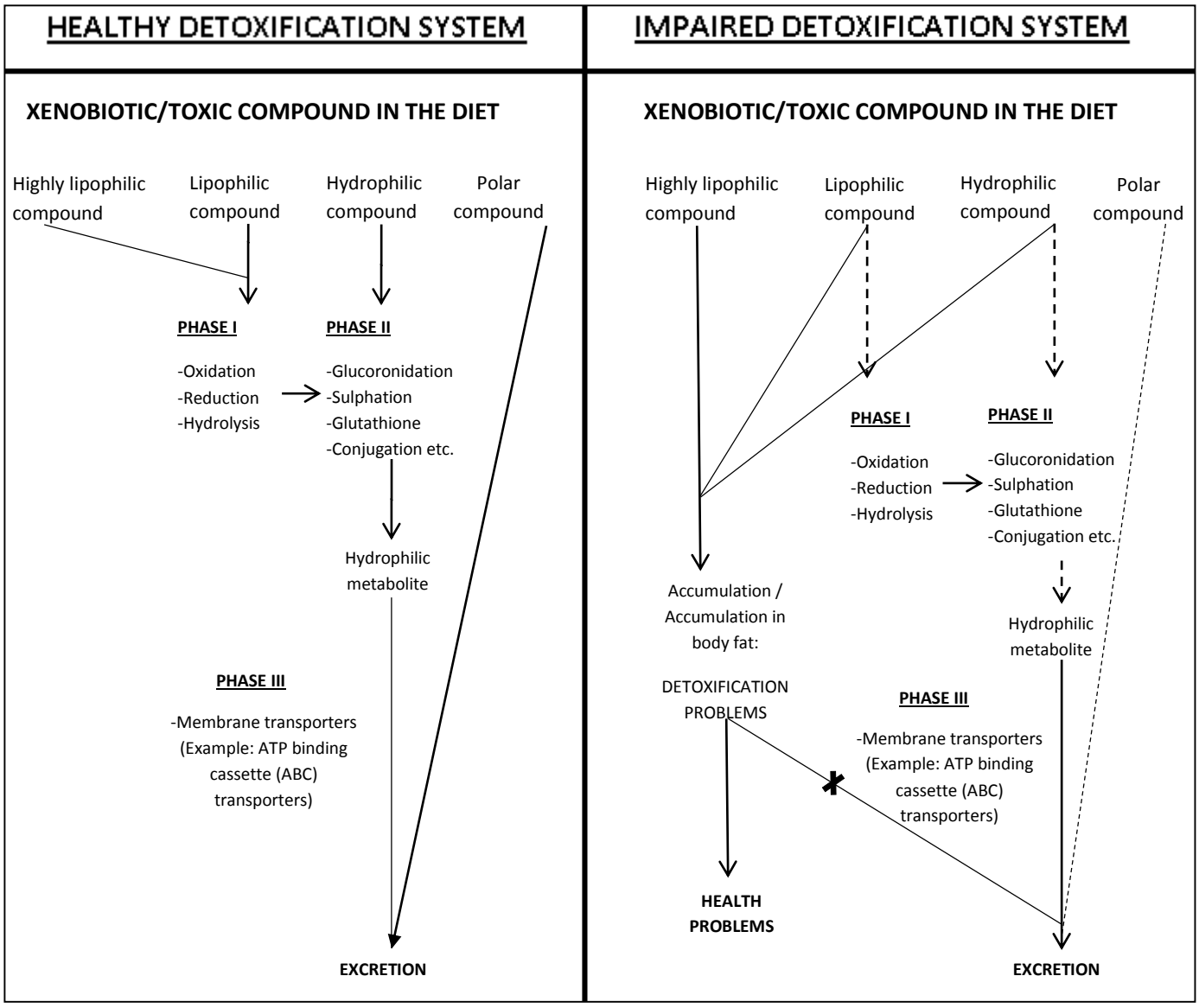


Figure 2.1: Indication of the processes of xenobiotic metabolism and the importance of detoxification systems (Adapted from Liska *et al*., 2006).

Phase 0, a newly recognised phase of biotransformation, consist of the action of ABC transporters (belonging to subfamily ABCB (P-gps/permeability glycoprotein)). These proteins function to transport toxicants out of cells, before enzymatic modifications of the compounds by Phase I and II occur, thus preventing the accumulation of harmful toxicants inside the cells (Epis *et al*., 2014). Phase I, as indicated by Table 2.1, is often referred to as

the functionalization phase consisting of oxidation, reduction and hydrolysis reactions (Figure 2.1) (Liska *et al*., 2006; Gonzales *et al*., 2006). Generally, these reactions lead to the inactivation or bioactivation of a compound by introducing or unmasking a functional group (-OH, $-CO_2H$, $-NH_2$, or -SH), in so doing creating an intermediate or metabolite of the original chemical compound which has been bioactivated into a more chemically reactive, toxic compound. This reactive compound will remain in the body potentially causing damage, especially to the liver where it was formed as it is not fully biotransformed. Although there are several Phase I DMEs, the most abundant and important are the cytochrome P450s (P450s). Phase II biotransformation reactions result in the conjugation of the modified xenobiotic compound with another substance by adding large water-soluble charged (polar) compound(s) to the functional group of the xenobiotic compound (Jančová and Siller, 2012). Phase III is the process of removing the water soluble Phase II conjugated compound from the cell and involves the role of membrane transporters that function to shuttle drugs and other xenobiotics across cellular membranes as indicated in Table 2.1 (Xu *et al*., 2004, Omiecinski, 2010; Penner *et al*., 2012). As a result, the molecular structure of a compound, influencing its solubility and/or toxicity, will be affected (Parkinson, 2001). The compound being excreted from the body will be far less toxic than its parent compound. Thus, the pharmacological and toxicological activity of a xenobiotic compound is in many ways the consequence of its metabolism (Parkinson, 2001; Fura *et al*., 2004; Gonzales *et al*., 2006; Penner *et al*., 2012). Despite the fact that there are numerous biotransformation enzymes, each enzyme has an affinity for a certain molecular compound or substrate.

Xenobiotics can exert adverse effects on human health by disrupting or interacting with multiple cellular communication pathways and disrupting essential biological structures needed for the body to function such as DNA, cellular membranes, and protein (Parkinson, 2001; Liska *et al*., 2006). The human body is thoroughly equipped to handle the presence of foreign chemicals by elimination and detoxification, thus contributing to the homeostatic response of the body (Figure 2.1) (Xu *et al*., 2004; Parkinson, 2001; Jančová and Siller, 2012).

Most drugs and other environmental toxins are so lipid-soluble that they would remain in the body for an indefinite period of time, even years (Nebert, 1981; Xu *et al*., 2004, Parkinson, 2001; Jančová and Siller, 2012). Were it not for the biotransformation of these compounds, lipophilic xenobiotics would be excreted from the body so slowly that they would eventually overwhelm and kill an organism (Nebert, 1981; Xu *et al*., 2004; Parkinson, 2001; Jančová and Siller, 2012). Toxins enter the liver as either water- or fat-soluble molecules. Water-soluble toxins are rather easily metabolized and excreted into the urine. In contrast, fat-soluble toxins can be stored in adipose (fat) tissue where they are protected from the body's

detoxification systems. With the knowledge of the extent to which drugs and xenobiotics may cause harm to living systems, it is understandable to say that detoxification is a crucial and important element of biotransformation in living organisms. In order to regulate metabolism and provide safe and secure medication, a comprehensive understanding of drug metabolism is needed (Gibson *et al*., 2001; Parkinson, 2001; Lin and Lu, 1997; Jančová and Siller, 2012).

2.2.1 Biotransformation versus metabolism

When applied to drugs and other foreign or toxic compounds, the terms biotransformation and metabolism are often used synonymously. The term metabolism is an essential pharmacokinetic process and often suggests the total fate of a xenobiotic; this consists of the absorption (Phase 0), biotransformation (Phase I/Phase II), distribution and metabolite excretion (Phase III) of a specific compound (Epis *et al*., 2014). Since products of drug biotransformation are recognized as metabolites and the metabolism of a drug or toxin in the body is often referred to as an example of a biotransformation reaction, it is common to see why these terms are often used synonymously to describe the same set of reactions in living systems (Parkinson, 2001). Metabolism is a necessary and basic process in living organisms that limits the time of a substance in the body. Biotransformation on the other hand is a specific term used to describe a series of enzyme-catalysed processes mandatory for the chemical transformation of xenobiotic compounds and toxins in the body of living organisms in order to convert lipid-soluble, non-polar, non-extractable forms of chemicals to hydrophilic metabolites that are extractable in bile and urine (Murphy, 2008; Jančová and Siller, 2012; Parkinson, 2001; Xu *et al*., 2004).

2.2.2 The history of drug metabolism: The discovery of the glycine conjugation pathway

The age-old question of what happens to the body when toxins accumulate and how the body deals with the presence of foreign compounds has enhanced the research in drug metabolism over the years (Liska *et al*., 2006; Murphy, 2008; Omiecinski *et al*., 2011). Drug discovery is a process involving numerous disciplines and interests exploring the cause and effect relationship toxins and foreign compounds have on the metabolism of living organisms

(Murphy *et al*., 2008). In order to understand the metabolic working of certain drugs and xenobiotics entirely, it is necessary to comprehend the transformation these compounds undergo in the body (Conti and Bickle, 1997). Detoxification mechanisms and the toxic effects of drugs and other xenobiotics have been studied extensively in various mammalian species since the 18th and 19th century, when the urine of animals and humans were analysed after the administration of particular compounds (Conti and Bickle, 1997; Liska *et al*., 2006). To date, the challenge to understand detoxification mechanisms continues (Liska *et al*., 2006).

The study of the fate of benzoic acid in the body resulted in the discovery of the glycine conjugation pathway. This was the first biotransformation reaction to be discovered that, in turn initiated the study of drug metabolism; see Table 2.2 indicating the major discoveries made on the subject of glycine conjugation (Badenhorst *et al*., 2013; Conti and Bickel, 1997; Steventon & Hutt, 2001; Lees *et al*., 2013). Benzoic acid is generally considered to be the substrate for the study of glycine conjugation; therefore, this will be used as an example to shed light on the history of the discovery of amino acid conjugation (Liska *et al*., 2006).

In 1773, Roulle was the first to mention the possible presence of benzoic acid in cow's and later in camel's urine (Table 2.2) (Conti and Bickel, 1997; Liska *et al*., 2006). In addition to findings by Roulle, Fourcroy and Vauquelin also observed the possible presence of benzoic acid in the urine of herbivorous animals in 1799 (Table 2.2) (Conti and Bickel, 1997). In 1801, it was discovered that urine contained a similar compound to benzoic acid, rather than benzoic acid itself after ingestion (Table 2.2) (Keller, 1842; Conti and Bickel, 1997). Woehler was the first scientist to experimentally investigate the fate of a foreign compound in 1824 (Table 2.2) (Conti and Bickel, 1997; Liska *et al*., 2006; Murphy, 2008). After administering benzoic acid to a dog, what Woehler discovered was that benzoic acid absorbed in the gastrointestinal tract reached the urine unchanged (Conti and Bickel, 1997).

Table 2.2: Table summary containing key dates regarding the discovery of the glycine conjugation pathway (Adapted from Conti and Bickel 1977).

THE GLYCINE CONJUGATION PATHWAY	
KEY DATES:	**DISCRIPTION**
1773	Rouelle was the first scientist to detect benzoic acid in the urine of cows.
1799	Likewise, Fourcroy and Vauquelin, also detected an acid in the urine of herbivorous animals which they believed to be benzoic acid.
1801	Scheele and Proust found that the acid in urine was similar but not identical to benzoic acid.
1824	Woehler was the first researcher to experiment and investigate the fate of a foreign compound. He found that benzoic acid in dogs was excreted as unchanged benzoic acid.
1829	Liebig discovered the compound hippuric acid from equine urine containing nitrogen, thus it was similar, but not the same as benzoic acid.
1830	Woehler speculated about the biotransformation of benzoic acid to hippuric acid.
1841	Ure succeeded in a biotransformation experiment in humans that showed benzoic acid to be excreted as hippuric acid.
1842	Keller confirmed the findings of Ure.
1844	Liebig found that hippuric acid was also a normal constituent of human urine.
1845	Dessaignes identified benzoylglycine and the synthesis of hippuric acid.
1857	Kuhne and Hallwachs carried out the first investigations focused upon the localization of glycine conjugation, indicating the liver vessels as the site of hippuric acid formation.
1863/66	Meissner and Shepard indicated the kidney as the site of hippuric acid formation.
1875	Spengel demonstrated that dogs are able to use exogenous glycine for the conjugation of benzoic acid and formation of hippuric acid.
1876	Bunge and Schmiedeberg confirmed that hippuric acid formation is localized in the kidneys.

The compound known as hippuric acid was only discovered in 1829 in the urine of horses after administration of benzoic acid by Liebig (Table 2.2) (Conti and Bickel, 1997; Lees *et al*., 2013). The name “hippuric” was assigned by Liebig form the Greek word for horse, hippos, because the acid was first isolated from the urine of horses (Lees *et al*., 2013). Unlike benzoic acid, hippuric acid contains nitrogen (Figure 2.2) (Conti and Bickel, 1997,)

Figure 2.2: The formation of hippuric acid after ingestion of benzoic acid (Adapted from Conti and Bickel, 1997).

But it was not until 1845 that Dessaignes reported on the discovery of the structure of hippuric acid. Hippuric acid has the chemical formula $C_9H_9NO_3$ (Figure 2.2) (Dessaignes *et al*., 1845, Dessaignes, 1857; Conti and Bickel, 1997; Lees *et al*., 2013). H and C NMR spectroscopy, capillary zone electrophoresis (CZE), capillary electrochromatography (CEC) combined with NMR spectroscopy, high performance liquid chromatography (HPLC), HPLC coupled with mass spectrometry (HPLC-MS), gas chromatography (GC), GC-MS, solid phase extraction (SPE), colorimetric reaction, immunochromatographic analysis and microfluidic chip-based electrochemical immunoassay are just some of the techniques that can be used to measure urinary hippuric acid concentration (Tomokuni and Ogata 1972; Buchet and Lauwerys 1973; Kira 1977; Sakai *et al*., 1983; Bales *et al*., 1984; Kongtip *et al*., 2000; Akira *et al*., 2001; Williams *et al*., 2005; Kawai *et al*., 2008; Park *et al*., 2007; Sarkissian *et al*., 2007; Lees *et al*., 2013). The biosynthesis of hippuric acid occurs within the mitochondrial matrix and requires two reactions, but this will be discussed in more detail in Section 2.3.

The first successful study of human metabolism was performed in 1841 by Alexander Ure (Table 2.2) (Ure, 1841; Conti and Bickle, 1997; Penner *et al*., 2012). By administering benzoic acid to himself and volunteers, Ure noticed the conversion of benzoic acid to hippuric acid after the ingestion of benzoic acid (Figure 2.2) (Badenhorst *et al*., 2013; Keller, 1842; Conti and Bickle, 1997). This conversion was later confirmed in 1842 by Keller, a student of Ure who felt challenged by Ure's results (Table 2.2). With an experiment on himself, Keller was able to confirm that benzoic acid was in actual fact metabolized to

hippuric acid (Ure, 1841; Keller, 1842; Conti and Bickle, 1997). In 1844 Liebig stated that hippuric acid was also a normal constituent of human urine consuming a mixed diet and that the excretion of hippuric acid was not diminished by the administration of benzoic acid (Table 2.2) (Conti and Bickle, 1997). In 1857, Kuhne and Hallwachs carried out the first investigations focused upon the localization of glycine conjugation, indicating the liver vessels as the site of hippuric acid formation (Table 2.2) (Conti and Bickle, 1997). In 1866, Meissner and Shepard also indicated the kidney as the site of hippuric acid formation (Table 2.2) (Meissner and Shepard, 1866; Quick, 1931).

The pursuit in understanding drug metabolism is far from over. Even though most of the major pathways of drug metabolism had already been discovered by the end of the 19th century, the impact of drugs on metabolism ensures that biotransformation reactions are still the focus of many studies in medicine and pharmacology today (Murphy, 2001; Liska *et al*., 2006).

2.3 Introduction to Phase 0, Phase I, Phase II and Phase III biotransformation reactions

Biotransformation reactions enzymatically disassemble unwanted chemicals and significantly promote the excretion of foreign compounds through complex systems of detoxification enzymes (Liska, 1998; Parkinson, 2001; Jančová *et al*., 2010). Since the 18th century research continued in the field of detoxification reactions. R.T. Williams defined the field of detoxification, and in 1947 he described two phases with regard to the biotransformation of non-reactive compounds; functionalization and conjugation (Williams, 1947; Liska, 1998; Liska *et al*., 2006; Murphy, 2001; Omiecinski, 2010). Functionalization incorporates oxygen to form a reactive site, while conjugation results in the addition of a water-soluble group to the reactive site (Table 2.1) (Liska, 1998). Phase I and Phase II, respectively, later replaced the original terms: functionalization and conjugation (Liska, 1998; Williams, 1947; Liska *et al*., 2006; Jančová *et al*., 2010). Biotransformation reactions occur throughout the body, with the liver being the predominant detoxifying organ because of its rich capacity of DMEs (Parkinson, 2001; Rushmore, 2002; Jančová and Siller, 2012). Certain compounds may undergo biotransformation in other secondary tissues as well, containing lower levels of DMEs, such as the kidney, lungs, and the skin (Parkinson, 2001; Rushmore, 2002; Xu *et al*., 2004). DMEs are distributed throughout the body and are triggered by the exposure to xenobiotic compounds such as those found in the environment, the diet, and cigarette

smoke (Parkinson, 2001; Rushmore, 2002). With that being said, drugs and foreign compounds as chemical entities can be substrates, inhibitors, or inducers of DMEs (Parkinson, 2001, Rushmore, 2002). Phase I activities are membrane associated, reactions of Phase I usually occur in the endoplasmic reticulum (microsomes) (ER), whereas Phase II metabolism involves soluble enzymes located in the cellular cytosol fraction (Liska, 1998, Knights, 2000).

2.3.1 Phase 0 of the detoxification process: Absorption

In addition to the three well-known phases of biotransformation (Phase I, Phase II and Phase III), another phase has been added recently (Phase 0). Phase 0, along with Phase III, involve the action of efflux pumps (located in the cellular membrane), belonging to the ATP-binding cassette (ABC) transporter family. During Phase 0, ABC transporters belonging to subfamilies ABCB (P-gps), transport toxicants out of the cells before enzymatic modifications by Phase I and/or Phase II of the compounds can occur, and thereby prevent the accumulation of harmful toxicants inside cells (Sarkadi *et al*., 2006; Bernaudin *et al*., 2009; Epis *et al*., 2014).

2.3.2 Phase I of the detoxification process: The functionalization phase

Phase I DMEs in the detoxification process, responsible for the first line of defence, are collectively known as the cytochrome P450 supergene family of enzymes (P450), located in the endoplasmic reticulum membrane (Figure 2.3) (Penner *et al*., 2012). P450 either directly neutralizes xenobiotic compounds, or modifies the compound by adding or exposing a functional group (-OH, NH_2, -SH, -COOH) (Liska, 1998; Parkinson, 2001; Jančová *et al*., 2010; Rushmore, 2002; Gonzales *et al*., 2006). During Phase I, the activity of a compound can be altered in one of several ways, similar or different to the parent compound. The compound can be transformed from an inactive to an active compound, known as bioactivation (Liska *et al*., 2006). Bioactivation results in chemically reactive intermediates, identified as prodrugs, products that are more toxic than the parent compound (Liska *et al*., 2006; Gonzales *et al*., 2006). If bioactivation does not occur, the compound can be metabolised to a different compound, or transformed from an active to an inactive compound to facilitate excretion. During detoxification, P450s perform two functions: 1) they make toxins more water-soluble, and 2) they convert the toxin into a molecule usually less toxic

and, therefore, less reactive towards DNA, proteins, etc. (Liska *et al*., 2006). The result is a more water-soluble, less toxic molecule easily transported in blood, through our kidneys, and out into the urine for elimination.

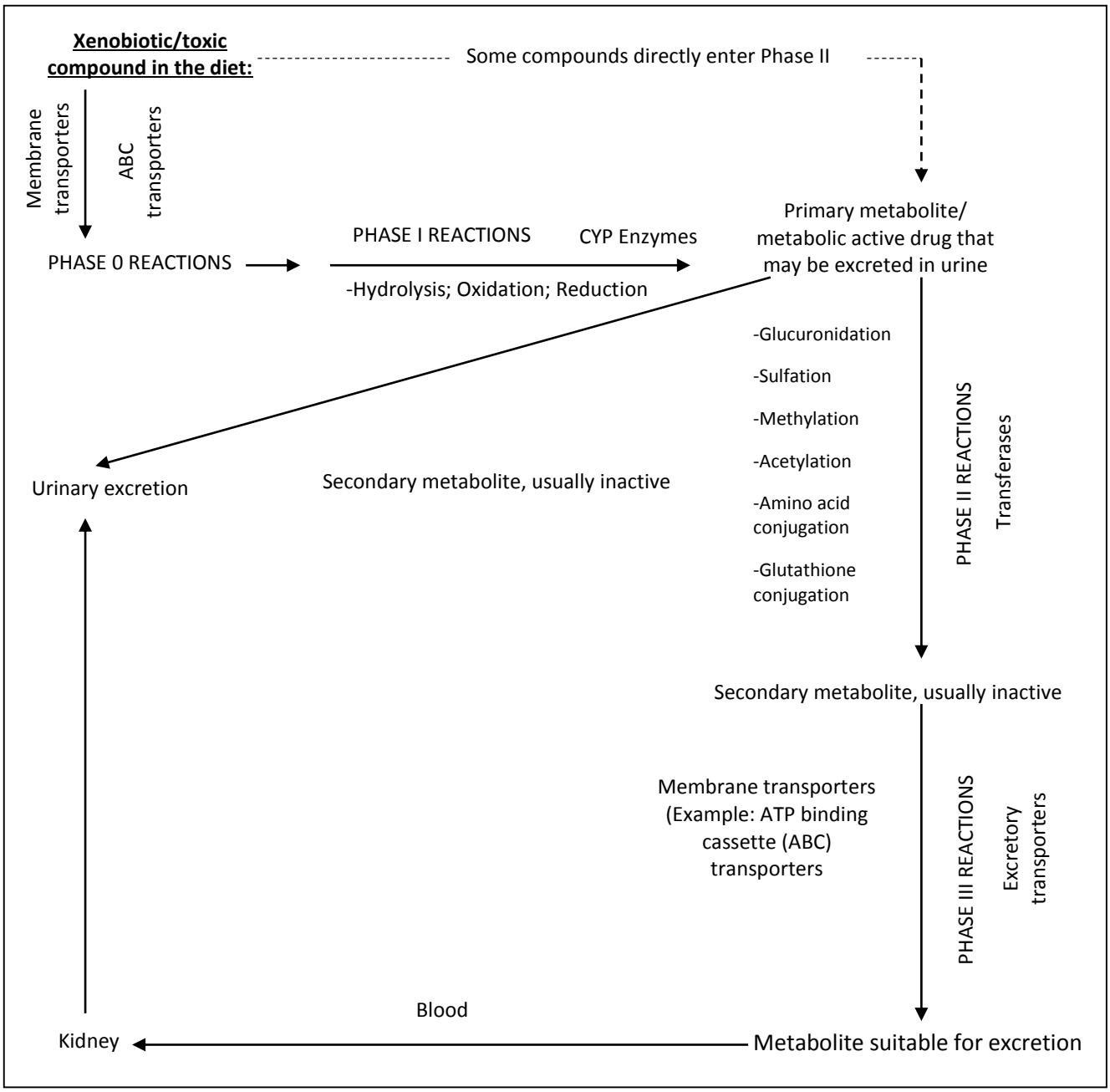


Figure 2.3: Xenobiotic biotransformation by means of Phase 0, Phase I, Phase II and Phase III biotransformation reactions (Adapted from Liska *et al*., 2006).

2.3.3 Phase II of the detoxification process: The conjugation phase

Further research revealed that biotransformed intermediates modified by Phase I enzyme systems undergo further biotransformation in the liver by a second series of enzymes called conjugases (Figure 2.3) (Liska *et al*., 2006). The conjugases are enzymes that attach molecules such as glucuronic acid, sulfate, glutathione, glycine, taurine, or methyl groups to the biotransformed intermediates (Figure 2.3) (Liska *et al*., 2006; Gonzales *et al*., 2006).

In order for reactions of Phase II biotransformation to take place, Phase I reactions are often called for, although it is not a requirement and the compound may be eliminated directly after the Phase I reaction (Figure 2.3). Some compounds, already in possession of a functional group, only enter Phase II metabolism, see figure 2.3 (Parkinson, 2001; Liska *et al*., 2006). During Phase II of the detoxification process, water-soluble substances are added to the xenobiotic compound(usually at the reactive site formed during Phase I) to increase its solubility, and consequently produce a conjugate. Phase II metabolism consists of four primary enzymes, UDP-glucuronosyltransferase (UGT), sulfotransferase (SULT), glutathione-S-transferase (GST) and N-acyltransferase (NAT) (Figure 2.3) (Parkinson, 2001; Rushmore and Kong, 2002; Jančová *et al*., 2010; Omiecinski, 2010; Gonzales *et al*., 2006). After Phase II modifications, the body is able to eliminate the transformed toxins in the urine. Phase II consists of the following reactions: glucuronidation, sulphation, acethylation, methylation, amino acid conjugation and glutathione conjugation reactions (Parkinson, 2001; Jančová *et al*., 2010). One of the most important Phase II detoxifying enzymes is known as glutathione-S-transferase (GST). GST is a family of enzymes catalysing the formation of thioester conjugates between the endogenous tripeptide glutathione and xenobiotic compounds, in so doing, aid in rendering the toxin more water-soluble and less toxic to the body. Furthermore, GST represents an important role in cellular protection against oxidative stress, reactive and toxic electrophiles such as reactive oxygen species (superoxide radical and hydrogen peroxide), produced through normal metabolic processes (Jančová *et al*., 2010). Besides GST, the body uses several other molecules to bind to the toxin and increase its solubility including sulphates, amino acids, and glucuronic acid.

2.3.4 Phase III of the detoxification process: The elimination phase

Much speculation surrounded the existence of a third phase of metabolism (Liska *et al*., 2006). In recognition of membrane transporters that function to shuttle drugs and other

xenobiotics such as products of Phase I and II reactions across cellular membranes into the bloodstream for elimination, Phase III biotransformation is a newly formulated descriptor of this action (Omiecinski, 2010). In contrast to the well-known history of discovery in Phase I and Phase II metabolism dating back to the 18th century, Phase III metabolism only dates back to 1976 with the discovery of the ATP-binding cassette (ABC) family of drug transporters (Juliano, 1976; Omiecinski, 2010). P-glycoprotein (permeability glycoprotein) (P-gp) was the initial member of what are currently referred to as the ATP-binding cassette (ABC) family of drug transporters (Benet, 1997; Brinkmann & Roots, 2001; Xu *et al*., 2005; Liska *et al*., 2006). The predominant role of these transporters is that of regulation of bile formation and the excretion of xenobiotics (Figure 2.3) (Chin *et al*., 1993; Liska *et al*., 2006).

2.4 Biotransformation Phase II (conjugation) reactions: a deeper look into amino acid conjugation

Research in amino acid conjugation (enzyme multiplicity, protein structure, and xenobiotic substrate selectivity), although gaining some momentum now, has not been studied as much over the years compared to the extensive knowledge on most xenobiotic-metabolizing enzymes such as the broad substrate specificity of the cytochrome P450 systems (Phase I reactions) (Knights *et al*., 2007; Beyoğlu *et al*., 2012; Knights and Miners, 2012). The detoxification of xenobiotic carboxylic acid (-COOH), including arylacetic, aryloxyacetic, aromatic acids (such as benzoic acid), and heteroaromatic acids, as well as endogenous acids is achieved through amino acid conjugation (Ioannides, 2002; Jančová *et al*., 2010; Knights *et al*., 2007; Knights and Miners, 2012; Beyoğlu *et al*., 2012). During amino acid conjugation, conjugates recognised as enzymes, attach molecules such as glucuronic acid, sulphate, glutathione, glycine, taurine, or methyl groups to the carboxylic group of an organic compound following Phase I reactions, glycine being the foremost amino acid utilized in humans (Ioannides, 2002; Knights *et al*., 2007; Liska *et al*., 2006; Beyoğlu *et al*., 2012; Knights and Miners, 2012). The selection of amino acids added to aromatic acids (such as benzoic acid), mainly depends on the chemical class of aromatic acid, as well as the species in question. Considering benzoic acid, the only amino acid utilized for conjugation in mammals is glycine (Ioannides, 2002; Beyoğlu *et al*., 2012). Reactions with endogenous substrates yield highly hydrophilic intermediates, the addition of endogenous substrates such as glycine to aromatic acids such as benzoic acid, has long been considered a process of detoxification (Ioannides, 2002). Amino acid conjugation of exogenous carboxylic acids is a two-step enzymatic process localised in the mitochondria (Ioannides, 2002; Knights *et al*.,

2007; Reilly *et al*., 2007; Jančová *et al*., 2010; Knights and Miners, 2012). The substrate is activated during the first step of amino acid conjugation and then combined with an amino acid during the second step to yield a conjugated product (Ioannides, 2002; Knights *et al*., 2007; Jančová *et al*., 2010).

2.4.1 Amino acid conjugation of benzoic acid with glycine: the first conjugation reaction demonstrated in humans

Day by day, benzoate, a preservative in food, is consumed in variable amounts by humans (Badenhorst *et al*., 2013). The conjugation of benzoic acid with glycine, yielding hippuric acid, is generally considered to be the first xenobiotic biotransformation reaction to be discovered (Table 2.2) (Figure 2.4) (Conti and Bickel, 1977; Ioannides, 2002; Liska *et al*., 2006; Beyoğlu *et al*., 2012).

Amino acid conjugation of exo- and endogenous carboxylic acids are based on the following two steps:

2.4.1.1 The first step: Activation

As indicated in Figure 2.4, the first step of amino acid conjugation consists of the initial activation of the carboxylic acid with adenosine triphosphate (ATP), generating an acyladenylate (AMP) and inorganic pyrophosphate (Knights *et al*., 2000; Ioannides, 2002; Knights *et al*., 2007; Knights and Miners, 2012). Followed by the reaction of the bound acyladenylate with CoA, catalysed by an ATP-dependent acid: CoA ligases. The medium-chain CoA synthetase (ACSM) (EC 6.2.1.2) has been identified in humans as HXM-A (xenobiotic/medium-chain fatty acid: CoA ligase) and is principally associated with the activation of benzoic acid. HXM-A is encoded by the *ACSM2B* gene (ENSG00000066813, acyl CoA synthetase medium-chain family member 2B), and leads to the formation of a “high energy” xenobiotic–CoA thioester intermediate. This will be discussed in more detail in Section 2.7. By the use of benzoic acid as substrate, ATP binds first to the enzyme (ACSM), indicated by the black circle followed by benzoic acid binding, pyrophosphate release, CoA binding, benzoyl CoA release, and AMP release (Knights *et al*., 2007).

In the event of the glycine conjugation of benzoic acid to yield hippuric acid, benzoic acid is in general considered to be the substrate for amino acid conjugation, but in actual fact,

benzoic acid is first a substrate for the conjugation with CoA followed by the acyl transfer of the xenobiotic CoA to the amino acid (Knights *et al*., 2007).

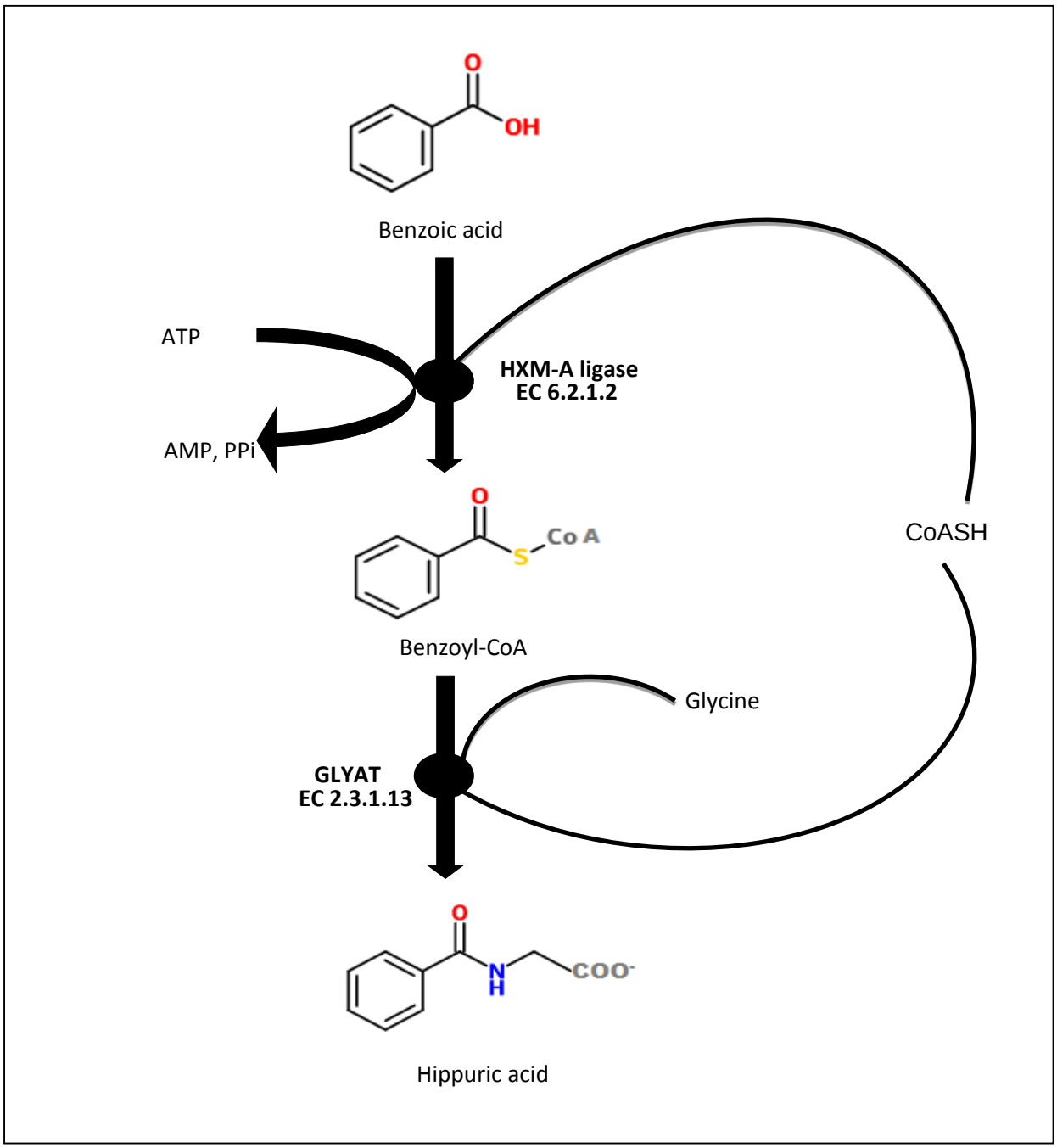


Figure 2.4: Description of glycine conjugation of benzoic acid leading to hippuric acid formation (Adapted from Badenhorst *et al*., 2013). Factors that may influence the overall rate of the glycine conjugation pathway include levels of ATP, CoASH, and glycine availability (The black circles indicate the ligase and GLYAT enzymes). Factors that may influence the overall functioning of HXM-A and GLYAT enzymes include expression or induction of the enzymes, substrate specificity, and genetic variation. AMP: Adenosine monophosphate; ATP: Adenosine triphosphate; CoASH:

Coenzyme A; GLYAT: Glycine N-acyltransferase; PPi: Pyrophosphate; EC: Enzyme commission number.

2.4.1.2 The second step

The next step in amino acid conjugation involves the linkage of the activated acyl group via an acyl CoA: amino acid N-acyltransferase (GLYAT) (EC 2.3.1.13) to the amino group of the acceptor amino acid (glycine, as indicated in Figure 2.4) (Knights *et al*., 2007; Knights and Miners, 2012). Because amino acid conjugation is a 2-step process it is important to bear in mind that a lack of amino acid conjugation in the second step, does not necessarily suggest a lack of xenobiotic CoA conjugation. Xenobiotic carboxylic acid is first a substrate for HXM-A (step 1), followed by the xenobiotic–CoA conjugate as a substrate for GLYAT (step 2) (Knights and Miners, 2012). Although several other acyl CoAs, such as salicylic-CoA, 4-aminobenzoyl-CoA, hexanoyl-CoA, and isovaleryl-CoA, can act as acyl donor substrates (with much less efficiently), benzoyl CoA is the preferred substrate for GLYAT (Nandi *et al*., 1979; Kolvraa and Gregersen, 1986; Badenhorst *et al*., 2012; Badenhorst *et al*., 2014).

Despite the fact that amino acid conjugation has been identified and studied since the 19th century, the exact mechanism of factors influencing the metabolism of xenobiotics is still in early stages of understanding (Knights *et al*., 2000). There are a few factors that may influence the overall rate of the glycine conjugation pathway including levels of glycine available in the body, ATP, CoASH, the amount of enzyme available to catalyse the reaction as well as substrate selectivity of ACSM and GLYAT, that can all impact the overall process and rate of amino acid conjugation (Figure 2.4; indicated in the boxes). This will be discussed further in Section 2.5.

Glycine conjugation, depending on the xenobiotic exposure, may occur in either the liver or the kidney contributing to either hepatic or renal acyl CoA formation (Badenhorst *et al*., 2013; Kasuya *et al*., 2000). During the first step of amino acid conjugation a xenobiotic acyl CoA is formed, but when this cannot be metabolised further during the amino acid conjugation process, it will accumulate, resulting in toxicity (Badenhorst *et al*., 2013). In the event of glycine conjugation of benzoic acid leading to hippuric acid formation, hippuric acid synthesis may be limited by depletion in CoA and glycine availability. On account of depletion in glycine levels, CoA will be trapped in the form of benzoyl CoA and a limit will be placed on the reaction rate, see Figure 2.4 (Badenhorst *et al*., 2013). GLYAT plays a major part in

restoring CoASH levels by conjugating xenobiotic acyl CoAs to glycine. These effects are caused primarily by accumulated CoA esters itself, as well as reduction of acetyl CoA and free CoA (CoASH) which will trigger a clinical disease, grouped together and referred to as CASTOR (Coenzyme A sequestration, toxicity or redistribution) disorders. These disorders mainly signify diseases caused by accumulation of acyl CoAs (Mitchell *et al*., 2008). In the conjugation of most xenobiotics, the XM-ligases (xenobiotic/medium-chain fatty acid:CoA) constitute the rate-determining step, as these ligases are dependent on the availability of substrate, CoA and ATP (Lohr *et al*., 1998; Vessey *et al*., 1999). The dual role of XM-ligases for fatty acid oxidation as well as xenobiotic activation is one of the reasons why mitochondrial accumulation of xenobiotic acyl CoA esters may interfere with beta-oxidation (β-oxidation) and disturb mitochondrial metabolism (Badenhorst *et al*., 2013). The metabolism of medium-chain fatty acids is poorly understood, but it is thought to play an important role in energy generation, given that the medium-chain fatty acids are probably generated from long- and very long-chain fatty acids by peroxisomal β-oxidation, and further degraded via mitochondrial β-oxidation after transportation into the mitochondrial matrix.

2.5 Variation in urinary hippuric acid excretion versus variation in the rate of glycine conjugation

It is important to understand the difference between variations in the amount of hippuric acid excreted as opposed to variations in the rate of glycine conjugation. An example of this can be seen from the work of Williams *et al*., (2010) on patients with Crohn's disease. A key characteristic of patients with Crohn's disease is a significant lower level of hippuric acid excreted in urine even though these individuals still show normal conversion of an oral dose of benzoic acid to hippuric acid. This is related to significant alterations in the gut microbiome, resulting in decreased fermentation of dietary phenols and lower production of phenylpropionate (Williams *et al*., 2010; Badenhorst *et al*., 2014).

Gut microbiota play a significant role in contributing to the excretion of a range of metabolites such as the transformation of dietary polyphenols into metabolically active antioxidative compounds, therefore these metabolites are often referred to as urinary mammalian-microbial cometabolites by co-existing with the human host (Heinken *et al.*, 2015). Substrates for glycine conjugation include the following; benzoic acid, salicylate, 4-hydroxybenzoate, 3-hydroxybenzoate, 4-aminobenzoate, 2-furoate, and microbial metabolites of polyphenols (Knights and Miners, 2012; Badenhorst *et al*., 2014). Polyphenol-

rich components are abundant in the human diet as vegetables, fruit, tea and coffee and the major families of phenolic compounds include flavan-3-ols, flavonols, flavanones, anthocyanins, and hydroxycinnamates (Manach *et al*., 2003; Tsao, 2010; Dueñas *et al*., 2015; Badenhorst *et al*., 2014). Polyphenolic molecules in the human diet have generated much attention due to their antioxidant potential linked to many health benefits (Tsao, 2010; Dueñas *et al*., 2015). Polyphenolic molecules are metabolised to simpler aromatic acids such as phenylpropionate, 3-hydroxyphenylpropionate, and 4-hydroxyphenylpropionate by gut microbiota (Manach *et al*., 2004; Tsao, 2010; Knights and Miners, 2012; Dueñas *et al*., 2015; Badenhorst *et al*., 2014). These aromatic acids are further metabolized to derivatives of benzoic acid as the microbial metabolites are absorbed and conjugated with glycine, glucuronic acid, or sulphate. Microbial degradation of dietary aromatic compounds in the intestine, such as polyphenols, purines, and aromatic organic acids and amino acids, produce simple carboxylic acids such as benzoyl CoA (Williams *et al*., 2010; Badenhorst *et al*., 2014).

It is well established that diet influences microbial fermentation and total bacteria in the intestine, so the amount of glycine conjugates excreted in the urine largely depends on the dietary intake of free aromatic acids and polyphenolic compounds, and the extent to which the polyphenols are fermented in the colon (Lees *et al*., 2013; Dueñas *et al*., 2015; Badenhorst *et al*., 2014). The polyphenol fermentation products produced in the colon depend on the type and amount of dietary content of the ingested polyphenols, its transit time through the digestive system, and individual colonic microbiota composition (Dueñas *et al*., 2015; Badenhorst *et al*., 2014). Although genetic and environmental factors are the main contributing factors of gut microbiota composition, interindividual variation in gut microbiota may, to some extent, indicate differences in dietary intake (Dueñas *et al*., 2015). Other factors that may contribute to the role of the gut microbiota in the metabolism of polyphenolic compounds and variation in the amount of hippuric acid excretion can be observed from studies conducted by Lees *et al* (2013). Results have shown that a reduction in urinary hippuric acid excretion and relevant metabolites such as phenylpropionate may be observed when orally-administered antibiotic-induced suppression of the gut microbiota takes place (Lees *et al*., 2013). Gut microbiota is central to the metabolism of polyphenolic molecules in the human diet and the consequent production of hippuric acid (Lees *et al*., 2013, Badenhorst *et al*., 2014).

The metabolism of benzoic acid is dose dependent since the glycine conjugation system is a saturable process (Andersen, 1989; Knights *et al*., 2007; Badenhorst *et al*., 2014). Suggesting that the amount of glycine conjugates excreted in the urine such as hippuric acid, largely depend on the dietary intake of free aromatic acids and polyphenolic

compounds. The major metabolite of benzoic acid is hippuric acid. However, benzoylcarnitine and benzoylglucuronide might be another important product of benzoic acid metabolism when large doses are administered, observed in a study conducted by Sakuma (1991). Sakuma *et al*., (1991) reported on patients suffering from carbamylphosphate synthetase deficiency type I on benzoic acid treatment. Hippuric acid excretion in urine is a possible cause of hypocarnitinaemia associated with long term benzoic acid treatment for hyperammonaemic patients (Sakuma, 1991; Badenhorst *et al*., 2014). This indicates that the formation of related metabolites depends on the dose of benzoic acid ingested as well as individual's glycine conjugation capacity, that may be influenced by several other factors such as gender, age, the species in question, as well as environmental factors such as diet and disease but this will not be discussed further in this study. Factors that influence the rate and capacity of the glycine conjugation pathway are discussed further in Section 2.6.

2.6 Glycine, CoA and ATP: The rate-limiting factors in glycine conjugation

Section 2.6 will focus on the factors influencing the rate and capacity of the glycine conjugation pathway.

2.6.1 Glycine and ATP availability

Glycine, CoA and ATP availability are rate-limiting factors for hippuric acid production after the administration of benzoic acid (Lees *et al*., 2013; Badenhorst *et al*., 2014). Hippuric acid synthesis in humans is a saturable process with large doses of benzoic acid, therefore benzoic acid clearance and with this the formation of hippuric acid will significantly increase with the co-administration of glycine (Knights and Miners, 2012; Lees *et al*., 2013; Badenhorst *et al*., 2014). Investigations by White (1941) support this theory by demonstrating that the administration of glycine supplementation normalized the growth inhibiting-effect of orally administered sodium benzoate in rats (White, 1941; Lees *et al*., 2013). Findings in this study concluded that the administration of sodium benzoate to a low-protein basal diet leads to the inhibition of growth in young white rats. The ability of various supplements to affect the growth-rate in rats was studied, with glycine supplements producing a prompt stimulation in the growth rate (White, 1941).

This theory was further supported by Beliveau and Brusilow (1987) with the co-administration of glycine with sodium benzoate resulting in a normalization of liver serine and glycine concentrations, and also reduced benzoyl CoA accumulation, compared to the control group (Beliveau and Brusilow, 1987). The study by Beliveau and Brusilow (1987) also examined factors that limit benzoic acid metabolism in growing rats, where the excretion of hippuric acid significantly increases with the supplementation of glycine in rats that are injected with growth-impairing doses of sodium benzoate. Growth-impaired rats tend to have decreased concentrations of glycine whereas liver benzoyl CoA concentration increases. Administration of glycine reverses these concentrations. Findings in these studies support the hypothesis that glycine availability for the reaction of benzoyl CoA transferases is a rate-limiting factor in hippuric acid synthesis (Beliveau and Brusilow, 1987). When glycine is not administered or supplies become extremely low, CoA is trapped as benzoyl CoA and free CoA is unavailable to be used for the first reaction in the glycine conjugation pathway, production of further benzoyl CoA is halted, therefore limiting the rate of hippuric acid production. Studies by Quick (1931) also supported this theory in man (Quick, 1931).

Additionally, the availability of ATP is also an important rate-limiting factor to bear in mind in the synthesis of benzoyl CoA as the conjugation of benzoic acid consumes two molecules of ATP. Therefore, the rate of the glycine conjugation pathway will be influenced by the ability to produce benzoyl CoA which is ATP dependent (Knights *et al*., 2000; Ioannides, 2002; Knights *et al*., 2007; Knights and Miners, 2012; Badenhorst *et al*., 2014).

2.6.2 Coenzyme A significance in metabolism

CoA is an extremely important molecule involved in hundreds of reactions and essential for several crucial mitochondrial reactions including pyruvate dehydrogenase, α-ketoglutarate dehydrogenase and fatty acid β-oxidation, and is a major regulator of energy metabolism (Zhang *et al*., 2007; Badenhorst *et al*., 2013). Since CoA is a thiol, reacting with carboxylic acids to form thioesters (thus functioning as an acyl group carrier), it is required for the metabolism of fatty acids, carbohydrates, amino acids and ketone bodies, as it assists in transferring fatty acids from the cytoplasm to mitochondria (Zhang *et al*., 2007; Badenhorst *et al*., 2013).

Acyl CoA esters are important intermediates in anabolic and catabolic reactions, so disturbances of CoA metabolism, changes in the relationships between CoASH and acyl CoAs, and deficiencies of enzymes acting on acyl CoAs, may result in many inborn errors of

metabolism and can have severe and far-reaching consequences for metabolism as a whole. Formation of a xenobiotic CoA within the mitochondrial matrix has the ability to disturb mitochondrial function as this enzyme is involved in both fatty acid and xenobiotic metabolism, for that reason, CoA metabolism needs to be tightly regulated (Knights, 2007; Zhang *et al*., 2007; Mitchell *et al*., 2008; Badenhorst *et al*., 2013).

2.6.3 Pathogenesis and toxicity of acyl CoA

Excluding the aspect of enzyme specificity, the only two remaining factors required for the effective functioning of the glycine conjugation pathway, are the CoA and glycine co-substrates. These two factors would enforce low capacity on the system if they are depleted or rates are low.

2.6.3.1 Depletion of CoASH

Disruptions of acyl CoA-related pathways originate ultimately from the accumulation of acyl CoA and/or from the associated decrease of acetyl CoA or CoASH (Mitchell *et al*., 2008). Availability and regulation of CoASH is essential for a number of important mitochondrial reactions. These include reactions such as fatty acid β-oxidation, pyruvate dehydrogenase and α-ketoglutarate dehydrogenase (Zhang *et al*., 2007; Mitchell *et al*., 2008; Knights and Miners, 2012; Badenhorst *et al*., 2013; Badenhorst *et al*., 2014). When CoASH availability becomes limited, energy metabolism will be impacted on numerous levels implicating important mitochondrial reactions such as the following:

1) The first step in the glycine conjugation pathway requires CoASH for the synthesis of benzoyl CoA, therefore the main factor leading to the accumulation of acyl CoA results from the depletion of CoASH, as indicated in Section 2.6.1.
2) Carnitine depletion is a consequence of acyl CoA accumulation acting negatively on energy metabolism. During the breakdown of lipids, carnitine functions as a transporter of fatty acids in the mitochondria from the intermembraneous space into the mitochondrial matrix so they can be broken down through β-oxidation to acetyl CoA required for the generation of metabolic energy via the citric acid cycle (TCA) (Zhang *et al*., 2007; Badenhorst *et al*., 2013). Carnitine binds acyl residues and assists in their elimination, reducing the amount of acyl residues conjugated with

CoA, furthermore increasing the ratio between acylated- and free-CoA levels (Foster, 2004). Depletion of CoASH leading to subsequent accumulation of acyl CoA esters in the mitochondria, consecutively affects the Krebs cycle, pyruvate oxidation, amino acid metabolism, mitochondrial and peroxisomal βoxidation pathways of intermediary metabolism that require CoA (Mitchell *et al*., 2008; Knights and Miners, 2012; Badenhorst *et al*., 2013; Badenhorst *et al*., 2014).

3) CoASH is required for the conversion of pyruvate to acetyl CoA by means of pyruvate dehydrogenase after the consumption of glucose as acetyl CoA acts as the entry-level substrate for the TCA cycle.
4) CoASH is also needed for the β-oxidation of fatty acids. Fatty acids are enzymatically broken down through β-oxidation to form acetyl CoA which is then further oxidized by the TCA and mitochondrial electron transport chain (ETC) releasing energy under normal circumstances. However, when CoASH is depleted, glucose and lipids cannot be effectively used for energy production (Eaton *et al*., 1996; Spiekerkoetter and Wood, 2010; Badenhorst *et al*., 2013*;* Fillmore *et al*., 2014).

2.6.3.2 CoASH restoration

Methods to restore CoASH sequestration mainly involve administration of L-carnitine and glycine agents, capable of accepting the transfer of acyl groups from acyl CoA, along with the restriction of dietary precursors (Michell *et al*., 2008). The primary mechanism in restoring depleted CoASH levels is glycine conjugation as well as conjugation to amino acids/carnitine. The hydrolysis of acyl CoAs by acyl CoA thioesterases to free fatty acid and CoASH regulates intracellular levels of CoASH, acyl CoAs and free fatty acids. Acyl CoA thioesterases regulate lipid metabolism through maintaining proper metabolic milieu by assuring appropriate levels of CoASH, acyl CoA, and free fatty acids (Hunt and Alexson, 2002; Badenhorst *et al*., 2013).

2.7 Formation of xenobiotic acyl CoAs

The formation of an acyl CoA thioester during the first step of amino acid conjugation is a familiar concept at this stage. It is well-known that several fatty acids and xenobiotic carboxylic acids that are conjugated to amino acids must first be activated to acyl CoAs by ATP-dependent acid: CoA ligases (Ioannides, 2002; Linder *et al*., 2006; Kochan *et al*., 2009;

Kasuya *et al*., 1996; Watkins *et al*., 2007). This family of enzymes are referred to as ligases based on their role of ligating a fatty acid to CoA and play a central role in the metabolism of fatty acids and xenobiotic carboxylic acids. Acyl CoA synthetases belong to a superfamily of adenylate-forming enzymes and exhibit selectivity for short-, medium-, long-, or very long-chain fatty acids (Kochan *et al*., 2009). Short-, medium-, long-, and very-long-chain fatty acids are activated to their corresponding acyl CoA thioesters by different families of acyl CoA synthetases (ACS), as seen during the first step of amino acid conjugation (Knights and Miners, 2012). Twenty-six distinct acyl CoA synthetase genes known in the human genome have been identified. These genes are classified into subfamilies of related enzymes based on fatty acid carbon chain length, indicated in Table 2.3.

Table 2.3: Subfamilies of acyl CoA synthetase genes

Ligase	Abbreviation	EC #	Common synonyms	Activation
Short-chain acyl CoA synthetase	ACSS	EC 6.2.1.1	Short-chain fatty acid: CoA ligase/ Acetate-CoA ligase/Acetyl-CoA synthetase	C2-C4 fatty acids, typically acetate, propionate, or butyrate
Medium-chain acyl CoA synthetase	ACSM	EC 6.2.1.2	Medium-chain fatty acid: CoA ligase/ Butyrate-CoA ligase/ Propionoyl-CoA synthetase	C4-C12 fatty acids
Long-chain acyl CoA synthetase	ACSL	EC 6.2.1.3	Long-chain fatty acid: CoA ligase/ Palmitoyl-CoA synthetase	C10-C22, typically palmitate and oleate
Very long-chain fatty acid	ACSVL	EC 6.2.1.3	Very Long-Chain Acyl CoA Synthetase -	Very long-chain fatty acids C22- and longer
Bubblegum acyl CoA synthetase	ACSBG	EC 6.2.1.3	Long-Chain-Fatty-Acid-CoA Ligase ACSBG1/ Long-Chain-Fatty-Acid-CoA Ligase ACSBG2/ ACSBG1/ ACSBG2	Long-and very long-chain fatty acid substrates

EC #: Enzyme Commission number; ACSS: Short-chain acyl CoA synthetase; ACSM: Medium-chain acyl CoA synthetase; ACSL: Long-chain acyl CoA synthetase; ACSVL: Very long-chain fatty acid; ACSBG1: Acyl CoA synthetase bubblegum family member; ACSBG2: Acyl CoA synthetase bubblegum family member 2.

Evident from above mentioned subfamilies of related enzymes of the acyl CoA synthetase genes are that these enzymes have overlapping fatty acid specificity (Knights, 2000). As a result, defining the substrate profile of these enzymes has therefore been challenging as overlapping substrate and fatty acid specificity, along with the use of crude tissue and subcellular preparations have made this a complicated task (Knights, 2000).

Table 2.4: Classification of the human acyl CoA synthetase family (Adapted from Watkins *et al*., 2007 and Watkins *et al*., 2012).

ACS Subfamily	Proposed Description	Approved Gene Name	Alternative Symbols
Short-chain	ACSS1	Acyl CoA synthetase short-chain family member 1	ACAS2L, AceCS2
	ACSS2	Acyl CoA synthetase short-chain family member 2	ACAS2, ACECS, ACS, ACSA
	ACSS3	Acyl CoA synthetase short-chain family member 3	N/A
Medium-chain	ACSM1	Acyl CoA synthetase medium-chain family member 1	BUCS1; MACS1; HXM-B
	ACSM2A	Acyl CoA synthetase medium-chain family member 2A	LOC123876; A-923A4.1
	ACSM2B	Acyl CoA synthetase medium-chain family member 2B	HXM-A; HYST1046
	ACSM3	Acyl CoA synthetase medium-chain family member 3	SAH; SA
	ACSM4	Acyl CoA synthetase medium-chain family member 4	LOC341392
	ACSM5	Acyl CoA synthetase medium-chain family member 5	FLJ20581
Long-chain	ACSL1	Acyl CoA synthetase long-chain family member 1	ACS1, FACL1, FACL2
	ACSL3	Acyl CoA synthetase long-chain family member 3	ACS3, FACL3
	ACSL4	Acyl CoA synthetase long-chain family member 4	ACS4, FACL4, LACS4
	ACSL5	Acyl CoA synthetase long-chain family member 5	ACS2, ACS5, FACL5
	ACSL6	Acyl CoA synthetase long-chain family member 6	ACS2, FACL6, KIAA0837, LACS 6, LACS2, LACS5
Very long-chain	SLC27A1	Solute carrier family 27 (fatty acid transporter), member 1	ACSVL5, FATP, FATP1
	SLC27A2	Solute carrier family 27 (fatty acid transporter), member 2	ACSVL1, FACVL1, FATP2, VLACS, VLCS, hFACVL1
	SLC27A3	Solute carrier family 27 (fatty acid transporter), member 3	ACSVL3, FATP3, VLCS-3
	SLC27A4	Solute carrier family 27 (fatty acid transporter), member 4	ACSVL4, FATP4, IPS
	SLC27A5	Solute carrier family 27 (fatty acid transporter), member 5	ACSB, ACSVL6, FACVL3, FATP5, VLACSR, VLCSH2
	SLC27A6	Solute carrier family 27 (fatty acid transporter), member 6	ACSVL2, FACVL2, FATP6, VLCS-H1
"Bubble-gum"	ACSBG1	Acyl CoA synthetase bubblegum family member 1	BG, BG1, BGM, GR-LACS, KIAA0631, LPD
	ACSGG2	Acyl CoA synthetase bubblegum family member 2	BGR, BRGL
Other	AACS	Acetoacetyl-CoA synthetase	ACSF1
	ACSF2	Acyl CoA synthetase family member 2	ACSF1
	ACSF3	Acyl CoA synthetase family member 3	N/A
	AASDH	Aminoadipate-semialdehyde dehydrogenase	ACSF4

N/A: Not available.

Table 2.4 describes the classification of the human acyl CoA synthetase family, six of these twenty six distinct human acyl CoA synthetases namely; ACSM1, ACSM2 (A/B), ACSM3, ACSM4 and ACSM5, form part of the medium-chain acyl CoA synthetases by catalysing the activation of medium-length fatty acids (C4-C12 fatty acids) as well as xenobiotic carboxylic acids such as benzoic acid (Badenhorst *et al*., 2013; Vessey *et al*.,1999; Vessey *et al*., 2003; Linder *et al*., 2006; Kochan *et al*., 2009; Watkins *et al*., 2007; Boomgaarden *et al*., 2009). The activation of xenobiotic and medium-chain fatty acids are facilitated by two XM-ligases namely; HXM-A and HXM-B, respectively. Therefore, these medium-chain acyl CoA synthetases are known as xenobiotic/medium-chain fatty acid: CoA ligases (Badenhorst *et al*., 2013; Vessey *et al*.,1999; Vessey *et al*., 2003; Linder *et al*., 2006; Kochan *et al*., 2009; Watkins *et al*., 2007; Boomgaarden *et al*., 2009). In 1999, HXM-A and HXM-B were characterized and isolated from human liver mitochondria, with activity towards a range of xenobiotics (Vessey *et al*., 1999; Iwai *et al*., 2003). In humans HXM-A, encoded by the *ACSM2B* gene has been identified as the active mitochondrial ATP dependent acid: CoA ligase, responsible for catalysing the initial reaction of amino acid conjugation (Vessey *et al*., 2003; Kasuya *et al*., 1996; Badenhorst *et al*., 2013). Multiplicity of the mitochondrial medium-chain xenobiotic CoA ligases and functions and characteristics of the HXM-A and HXM-B will be discussed in more detail in Section 2.8.

2.8 Medium-chain acyl CoA synthetases (HXM-A and HXM-B)

As discussed in Section 2.7, mitochondria from liver contain various amounts of diverse forms of carboxylic acid: CoA ligases. Xenobiotic medium-chain fatty acid: CoA ligases are of particular interest as these ligases have extremely broad substrate specificities. Not only do these ligases function as xenobiotic carboxylic acid: CoA ligases, but also medium-chain fatty acid: CoA ligases (Vessey and Kelley, 1997; Vessey and Kelley, 1998; Vessey *et al*., 1999).

Vessey and Hu (1995) stated that mitochondria from liver contain a number of different forms of carboxylic acid: CoA ligases, of which three distinct XM-ligases (XL-I, XL-II, XL-III) were separated and characterised from bovine liver mitochondria (Vessey and Hu, 1995; Vessey, 1997; Vessey and Kelley, 1997; Vessey and Kelley, 1998; Vessey *et al*., 1999; Knights and Drogemuller, 2000; Vessey *et al*., 2003). These ligases have overlapping substrate specificities and cation requirements, as well as slightly different molecular weights (Killenberg *et al*., 1971; Vessey and Hu, 1995; Vessey and Hu, 1996; Vessey, 1997; Vessey

and Kelley, 1997; Vessey and Kelley, 1998; Vessey *et al*., 1999; Knights and Drogemuller, 2000; Vessey *et al*., 2003).

The designation of these three distinctive ligases was based on their order of elution from a Diethylaminoethyl-cellulose chromatography column (DEAE-column) (Vessey *et al*., 1999). Additional to these studies, Kasuya *et al*., (1996) reported on the identification of an apparent fourth XM-ligase (XL-J), purified from bovine liver mitochondria (Kasuya *et al*., 1996; Knights and Drogemuller, 2000). Vessey and Kelley (1997) reported on the purification and partial sequencing of the XL-I form of XM-ligases (Vessey and Kelley, 1997; Vessey and Kelley, 1998). The study by Vessey and Kelley (1997) uncovered regions of high homology with the essential hypertension protein (SA protein) leading to the speculation of the role of endogenous carboxylic acids in the regulation of blood pressure (Vessey and Kelley, 1997). All forms of bovine XM-ligases (XL-I, XL-II, XL-III and XL-J) have been shown to activate C3-C10 fatty acids as well as a broad variety of aromatic carboxylic- and arylacetic acids, along with substrates common to all forms such as hexanoate, octanoate, propionate, benzoate, 4-chlorobenzoate, 4-aminobenzoate, 4-nitrobenzoate, 1-naphthylacetate and salicylate (Vessey and Hu, 1995; Kasuya *et al*., 1996; Vessey *et al*., 1998; Kasuya *et al*., 1998; Knights, 2000).

Bovine XM-ligases are very well characterised, but what is of more interest is that of human liver XM-ligases. Accordingly, Vessey *et al.,* (1999) had undertaken the isolation and characterization of these XM-ligases from human liver mitochondria as these enzymes have not been characterized (Vessey *et al*., 1999). The study by Vessey *et al*., (1999) reported that the matrix space of human liver mitochondria contain a particular short-chain fatty acid: CoA ligase (SC-ligase) as well as a number of ligases with broad specificities towards the activation of medium-chain fatty acids as well as a wide range of xenobiotic carboxylic acids (Badenhorst *et al*., 2013; Vessey *et al*.,1999; Iwai *et al*., 2003; Vessey *et al*., 2003; Linder *et al*., 2006; Kochan *et al*., 2009; Watkins *et al*., 2007; Boomgaarden *et al*., 2009). The activation of xenobiotic carboxylic acids and several medium-chain fatty acids to acyl CoAs in humans are facilitated by less well characterised ATP-dependant acid: CoA ligases of which two such HXM-ligases were isolated from liver mitochondria (Vessey *et al*., 1999; Vessey *et al*., 2003; Linder *et al*., 2006; Knights *et al*., 2007; Watkins *et al*., 2007; Boomgaarden *et al*., 2009; Kochan *et al*., 2009; Badenhorst *et al*., 2013). These two ligases consist of broad but differing substrate specificities and were initially termed the ‘Mahler form’ and the ‘Salicylate form’ (now referred to as HXM-A and HXM-B respectively), this will be discussed in more detail in Section 2.8.1 (Mahler, 1953; Killenberg *et al*., 1971; Groot, 1976; Vessey and Kelley, 1997).

2.8.1 Biochemical and enzymatic characteristics of HXM-A and HXM-B

Fatty acids (C4-C12), as well as xenobiotic and branched-chain fatty acids are activated by the Mahler form (HXM-A), while the Salicylate form (HXM-B) activates salicylate, hexanoate, p-aminosalicylate and o-methoxybenzoate (Killenberg *et al*., 1971; Groot, 1976; Vessey and Kelley, 1997). Vessey *et al.,* (1999) revealed through analysis of human liver mitochondrial lysates, the existence of these distinct forms of human liver XM-ligases, significantly different in substrate specificities from bovine XM-ligases (Vessey *et al*., 1999; Iwai *et al*., 2003). The designation of these ligases was also based on their order of elution from a DEAE-cellulose chromatography column. The first XM-ligase to elute from the DEAE-cellulose column was assigned as the human XM-ligase form A (a 48 kDa protein), and the second XM-ligase to elute from the DEAE-cellulose chromatography column was assigned as form B (a 49 kDa protein) (Table 2.5) (Vessey *et al*., 1999; Knights *et al*., 2007). Vessey *et al*., (1999) characterised by way of substrate specificity studies, the now well-known fact that these XM-ligases are capable of activating both xenobiotic and medium-chain fatty acids as proven by benzoate and hexanoate activation by both ligases (Vessey *et al*., 1999).

In a study conducted by Vessey *et al*., (1999), the HXM-A and HXM-B form of XM-ligases were characterized with regard to their substrate specificity, assayed with fifteen different carboxylic acids at a concentration of 100 µM in each case, as indicated by Table 2.5 (Benzoate; Propionate; Butyrate; Hexanoate; Octanoate; Decanoate; Laurate; p-Hydroxybenzoate; Phenylacetate; Naphthylacetate; Indoleacetate; Isobutyrate; Nicotinate; Salicylate; Valproate). Findings by Vessey *et al*., (1999) indicate that the XM-Ligases (A/B) differ significantly in their substrate specificities. HXM-A had the highest affinity for benzoate, and HXM-B had the highest affinity for hexanoate (Vessey *et al*., 1999). HXM-B appeared to be better at activating straight chain fatty acids (Butyrate; Hexanoate; Octanoate; Decanoate), and HXM-A better at activating xenobiotics. Kinetic analysis was conducted on each of the XM-ligases by Vessey *et al*., (1999), however it appeared that the two ligases were of different degrees of purity so values are only referred to as relatively apparent. Nonetheless, it is evident from the study by Vessey *et al*., (1999) that the best substrate for HXM-A is benzoate while benzoate values are significantly decreased in HXM-B compared to those for the medium-chain fatty acids. The best substrate for HXM-B appeared to be hexanoate (Vessey *et al*., 1999).

Table 2.5: A summary of the properties and substrate specificities of the HXM-ligases (Adapted from Vessey *et al*., 1999).

HXM-ligase		Human/xenobiotic medium-chain ligase form A	Human/xenobiotic medium-chain ligase form B
Abbreviation		HXM-A	HXM-B
Common Synonym		Mahler form	Salicylate form
Molecular weight		48 kDa	49 kDa
Activation		• Fatty acids (C4-C12) • Xenobiotics • Branched-chain fatty acids	• Straight chain fatty acids • Xenobiotics
Substrate specificity (%) (Vessey *et al*., 1999)	Benzoate	100	100
	Propionate	25±2	18±4
	Butyrate	18±2	33±3
	Hexanoate	78±4	407±20
	Octanoate	42±6	120±13
	Decanoate	20±4	38±5
	Laurate	1.0±0.3	1.3±0.4
	p-Hydroxybenzoate	15±3	8.9±0.6
	Phenylacetate	21±4	8.4±0.2
	Naphthylacetate	<1	2.3±0.2
	Indoleacetate	0	0
	Isobutyrate	6±3	0
	Nicotinate	0	<1
	Salicylate	1.5±0.5	<1
	Valproate	5±1	2.3±0.3

Initial substrate specificity for each of the substrates were normalised relative to the activity observed with benzoate.

The molecular mass of human XM-ligases, HXM-A and HXM-B, has been reported as 48 and 49 kDa respectively. Variations in the values for the molecular weight have been reported and may be partly explained by the different techniques used in the different studies (Vessey *et al*., 1999).

Fujino *et al*. (2001) characterised and reported on the cloning and expression of two medium-chain acyl CoA synthetases namely; ACSM1 and ACSM3. Findings suggested that

ACSM1 appeared to be similar to HXM-B, but this will not be discussed here as the focus will be on HXM-A (encoded by ACSM2B) (Kasuya *et al*., 1996; Vessey *et al*., 1999; Vessey *et al*., 2003; Watkins *et al*., 2007; Knights and Miners, 2012; Knights and Miners, 2012; Badenhorst *et al*., 2013). Subsequently, Vessey *et al*., (2003) reported on the isolation, sequencing and expression of a recombinant HXM-A form (in COS cells) (Vessey *et al*., 2003). Furthermore, it is revealed through enzyme activity assays with hexanoate, benzoate, and phenylacetate as carboxylic acid substrates, that the expressed enzyme had greater affinity for benzoate than phenylacetate, consistent with the known substrate specificity of HXM-A. HXM-A amino acid sequence comparison to the ACSM1 (similar to HXM-B) sequence reveals only a 56.2% amino acid homology, while ACSM1 appears to be highly homologous to the XL-III bovine form presumed to be the same as HXM-B (Vessey *et al*., 2000; Vessey *et al*., 2003).

2.9 The *ACSM2A* and *ACSM2B* genes and genetic variation

As discussed throughout Chapter 2, *ACSM2A* (also known as medium-chain acyl CoA synthetase 2A, butyryl-CoA synthetase 2A or butyrate-CoA ligase), and *ACSM2B* (also known as medium-chain acyl CoA synthetase 2B, butyryl-CoA synthetase 2B or butyrate-CoA ligase 2B), both encode for 577 amino acid proteins belonging to the ATP-dependent AMP-binding enzyme family (Vessey *et al*., 1999; Iwai *et al*., 2003; Vessey *et al*., 2003; Linder *et al*., 2006; Boomgaarden *et al*., 2009; Kochan *et al*., 2009).

Table 2.6: Properties of human acyl CoA synthetase medium-chain family members (Adapted form Watkins *et al*., 2007).

ACS	RefSeq Nucleotide	RefSeq Protein	Chromosomal Location	Strand	Exons
ACSM1	NM_052956	NP_443188	16p12.2	Reverse	13
ACSM2A	NM_001010845	NP_001010845	16p12.3	Forward	14
ACSM2B	NM_182617	NP_872423	16p12.3	Reverse	15
ACSM3	NM_005622	NP_005613	16p13.11	Forward	14
ACSM4	NM_001080454	NP_001073923	12p13.31	Forward	13
ACSM5	NM_017888	NP_060358	16p12.3	Forward	14

Medium-chain ACSs identified by Watkins *et al*., (2007) are listed along with the number of exons, the strand location and chromosomal location, the transcript variants for which Reference Sequences

(RefSeq) are found in National Center for Biotechnology Information (NCBI) databases, the number of amino acids was deduced from the NCBI data.

2.9.1 The *ACSM2A* gene

The *ACSM2A* gene (ENSG00000183747; www.ensembl.org) is located on chromosome 16 at position 16p12 on the forward strand, spans over 36 000 base pairs, and contains 14 exons, of which 13 are coding exons (Table 2.6). Within the *ACSM2A* gene, there are approximately 3270 known SNPs (www.ensembl.org, November, 2015). Of the 3270 known SNPs, 457 synonymous SNPs and 147 non-synonymous SNPs are identified within the coding region of the *ACSM2A* gene (Ensemble, v82, http://grch37. ensembl.org) (Cunningham *et al*., 2015). ACSM2A is one of 26 putative human acyl CoA synthetases. The reference transcript used in this study for the *ACSM2A* gene is ENST00000573854, as suggested by Watkins *et al*., (2007) (see Table 2.7).

2.9.2 The *ACSM2B* gene

The human *ACSM2B* gene (ENSG00000066813; www.ensembl.org) is located on chromosome 16 at position 16p12.3 on the reverse strand, spans over 40 000 base pairs, and contains 15 exons, of which 13 are coding exons (Table 2.6). Since the *ACSM2B* gene is located on the reverse strand, all variations and sequence data discussed during this study relating to *ACSM2B*, specifically in Chapter 4, reflect the negative ('−' or "reverse") strand. Two splice variants of human ACSM2B mRNA exist, coding for isoforms a (2935 residues) and b (1988 residues). The transcript for isoform b does not contain exon 15. Within the *ACSM2B* gene, there are approximately 3535 known SNPs (www.ensembl.org, November, 2015). Of the 3535 known SNPs, 432 synonymous SNPs and 164 non-synonymous SNPs are identified within the coding region of the *ACSM2B* gene (Ensemble, v82, http://grch37. ensembl.org) (Cunningham *et al*., 2015). ACSM2B is one of 26 putative human acyl CoA synthetases. The reference transcript used in this study for the *ACSM2B* gene is ENST00000567001, as suggested by Watkins *et al*., (2007) (see Table 2.6)

No information at this time is available suggesting that genetic variation in the acid: CoA ligase enzymes influence the metabolism of salicylate and also other benzoate derivatives.

2.9.3 Nomenclature of the *ACSM2* genes (*ACSM2A* and *ACSM2B*)

The well-described HXM-A form of XM-ligases was renamed by the Human Genome Organization (HUGO) first to ACSM2 and, more recently, to ACSM2B. ACSM2A was also recently renamed by HUGO as the new proposed designation of ACSM6, because of the high sequence similarity to ACSM2B. Due to chromosomal duplication, two different loci of the *ACSM2* gene exists, *ACSM2A* and *ACSM2B*, located on chromosome 16 at position 16p12.3 (Boomgaarden *et al*., 2009; Loftus *et al*., 1999; Watkins *et al*., 2007; Martin *et al*. 2004). In 2009, Boomgaarden *et al*. detected that *ACSM2B* is the most abundant transcript in the liver of the ACSM family. Due to the abundant transcript level of the *ACSM2B* gene compared to expression levels of the rest of the ACSM family, *ACSM2B* is recommended to be the more promising target for studies into the role of the ACSM family as disease risk genes. Numerous association studies described the ACSM family as candidate genes for development of hypertension, overweight, insulin resistance, as well as type 2 diabetes mellitus (Iwai *et al*., 2003; Lindner *et al*., 2006; Boomgaarden *et al*., 2009).

The *ACSM2A* and *ASCM2B* genes are almost identical and therefore difficult to distinguish, the coding sequences of these genes are 98.8% identical, and their amino acid sequences are 97.6% identical, so it would be likely to assume that experimentally determined differences were attributable to polymorphisms or sequencing errors (Watkins *et al*., 2007). Although the *ACSM2A* and *ASCM2B* genes are nearly identical, there is sufficient evidence supporting the existence of both genes. Both the *ACSM2A* gene and *ACSM2B* gene are located on chromosome 16p12.3, whereas *ACSM2A* is on the plus strand, *ACSM2B* is on the minus strand, indicated in Table 2.7.

The *ACSM2A* and *ACSM2B* genes have 21 nucleotide differences in the coding region. There are 17 non-synonymous substitutions and 4 synonymous substitutions. Only one of the amino acid changes resulting from a non-synonymous substitution lies within a conserved motif. Residue 463, found in Motif II, is Asn in the *ACSM2A* gene and Asp in the *ACSM2B* gene; this will be explained in full detail in Section 2.9.4 and Table 2.7.

2.9.4 Conserved sequence motifs used to identify candidate human ACSs

A previous study by Steinberg *et al*., (1999) described two highly conserved amino acid sequences for a group of 57 ACS sequences from a diverse group of organisms (including mammals, roundworms, fruit flies, bacteria, and yeast). These conserved Motifs were designated Motif I (20 amino acids) and Motif II (44–45 amino acids). 26 distinct acyl CoA synthetase genes in the human genome were identified in a study conducted by Watkins *et al*., (2007), as previously mentioned in Section 2.7 and Table 2.5.

2.9.4.1 Motif I

As pointed out in the previous section, work by Steinberg *et al*., (1999) and Watkins *et al*., (2007) revealed that at least two distinct, highly conserved amino acid sequence motifs are present in proteins with ACS activity. Motif I is highly conserved from bacteria to humans ACSs and consist of a 10 amino acid sequence, starting at about residue 260 (Table 2.8) (Steinberg *et al*., 1999; Watkins *et al*., 2007). Motif I is found within the 11 amino acid putative AMP binding domain signature in the PROSITE database (PDOC00427, PS00455). The PROSITE database describes protein domains and families, functional sites, as well as related patterns and profiles to identify these domains (http://expasy.org/prosite).

2.9.4.2 Motif II

Previous studies conducted by Steinberg *et al*., (1999) and Watkins *et al*., (2007) on the amino acid sequences of 57 ACSs also identified a second region of residues, highly conserved. Motif II consists of 44–45 residues partially overlapping with the signature motif identified by Black *et al*., (1997) (Table 2.7). A 25 amino acid consensus sequence; DGWLHTGDIGXWXPXGXLKIIDRKK, was identified by Black *et al*., (1997) who examined a group of ACSLs from bacteria, yeast, and mammals. This sequence appeared to be common to all fatty acyl CoA synthetases. Black *et al*., (1997) proposed that this consensus sequence signifies the fatty acyl CoA synthetase signature motif (FACS signature motif). The signature motif can be used for two main objectives: i) for the identification of candidate

ACSs and, ii) for the organisation of the ACSs into subfamilies as described in Table 2.4. An arginine (Arg) residue was found within Motif II as well as the signature motif in studies conducted by Black *et al*., (1997), Steinberg *et al*., (1999), and Watkins *et al.,* (2007). This Arg residue was found to be nearly invariant in ACS sequences (Table 2.7).

Table 2.7: Conserved amino acid sequence motif I and II identified in human ACSs (Adapted from Watkins *et al*., 2007).

	ACS	MOTIF I	MOTIF II
Short-chain ACS	**ACSS1**	293-YTSGSTGMPK-302	531-TGDGAYRTEGGYYQITG**R**MDDVINISGHRLGTAEIE-566
	ACSS2	315-YTSGSTGKPK-324	550-TGDGCQRDQDGYYWITG**R**IDDMLNVSGHLLSTAEVE-585
	ACSS3	298-YTSGTTGLPK-307	537-TMDAGYMDEEGYLYVMS**R**VDDVINVAGHRISAGAIE-572
Medium-chain ACS family	**ACSM1**	225-FTSGTTGFPK-234	450-TGDRGKMDEEGYICFLG**R**SDDIINASGYRIGPAEVE-485
	ACSM2A	220-FTSGTSGLPK-229	444-LGDRGIKDEDGYFQFMG**R**ANDIINSSGYRIGPSEVE-479
	ACSM2B	220-FTSGTSGLPK-229	444-LGDRGIKDEDGYFQFMG**R**ADDIINSSGYRIGPSEVE-479
	ACSM3	234-FTSGTSGYPK-243	459-TGDRGYMDKDGYFWFVA**R**ADDVILSSGYRIGPFEVE-494
	ACSM4	228-FTSGTTGFPK-237	453-TGDRGVMDSDGYFWFVG**R**ADDVIISSGYRIGPFEVE-488
	ACSM5	229-FTSGTTGAPK-238	453-TGDRARMDKDGYFWFMG**R**NDDVINSSSYRIGPVEVE-488
Long-chain ACS family	**ACSL1**	275-FTSGTTGNPK-284	538-TGDIGKWLPNGTLKIID**R**KKHIFKLAQGEYIAPEKIE-574
	ACSL3	286-YTSGSTGLPK-295	562-TGDIGEFEPDGCLKIID**R**KKDLVKLQAGEYVSLGKVE-598
	ACSL4	236-YTSGSTGRPK-245	512-TGDIGEFHPDGCLQIID**R**KKDLVKLQAGEYVSLGKVE-548
	ACSL5	316-FTSGTTGDPK-325	579-TGDIGRWLPNGTLKIID**R**KKNIFKLAQGEYIAPEKIE-615
	ACSL6	300-FTSGTTGNPK-309	563-TGDIGKWLPAGTLKIID**R**KKHIFKLAQGEYVAPEKIE-599
Very long-chain ACS family	**ACSVL1**	224-YTSGTTGLPK-233	463-SGDLLMVDHENFIYFHD**R**VGDTFRWKGENVATTEVA-498
	ACSVL2	223-FTSGTTGLPK-232	462-TGDLIVQDQDNFLYFWD**R**TGDTFRWKGENVATTEVA-497
	ACSVL3	334-FTSGTTGLPK-343	573-TGDLLVCDDQGFLRFHD**R**TGDTFRWKGENVATTEVA-608
	ACSVL4	248-YTSGTTGLPK-257	490-SGDVLVMDELGYMYFRD**R**SGDTFRWRGENVSTTEVE-525
	ACSVL5	245-YTSGTTGLPK-254	487-TGDVLVMDELGYLYFRD**R**TGDTFRWKGENVSTTEVE-522
	ACSVL6	294-YTSGTTGLPK-303	533-TGDVLAMDREGFLYFRD**R**LGDTFRWKGENVSTHEVE-568
	Consensus	[Y,F]TSG[T,S]TGXPK	TGDX(7)GX(3)hX(2)**R**X(4)hX(3,4)GX(2)hX(4)hE

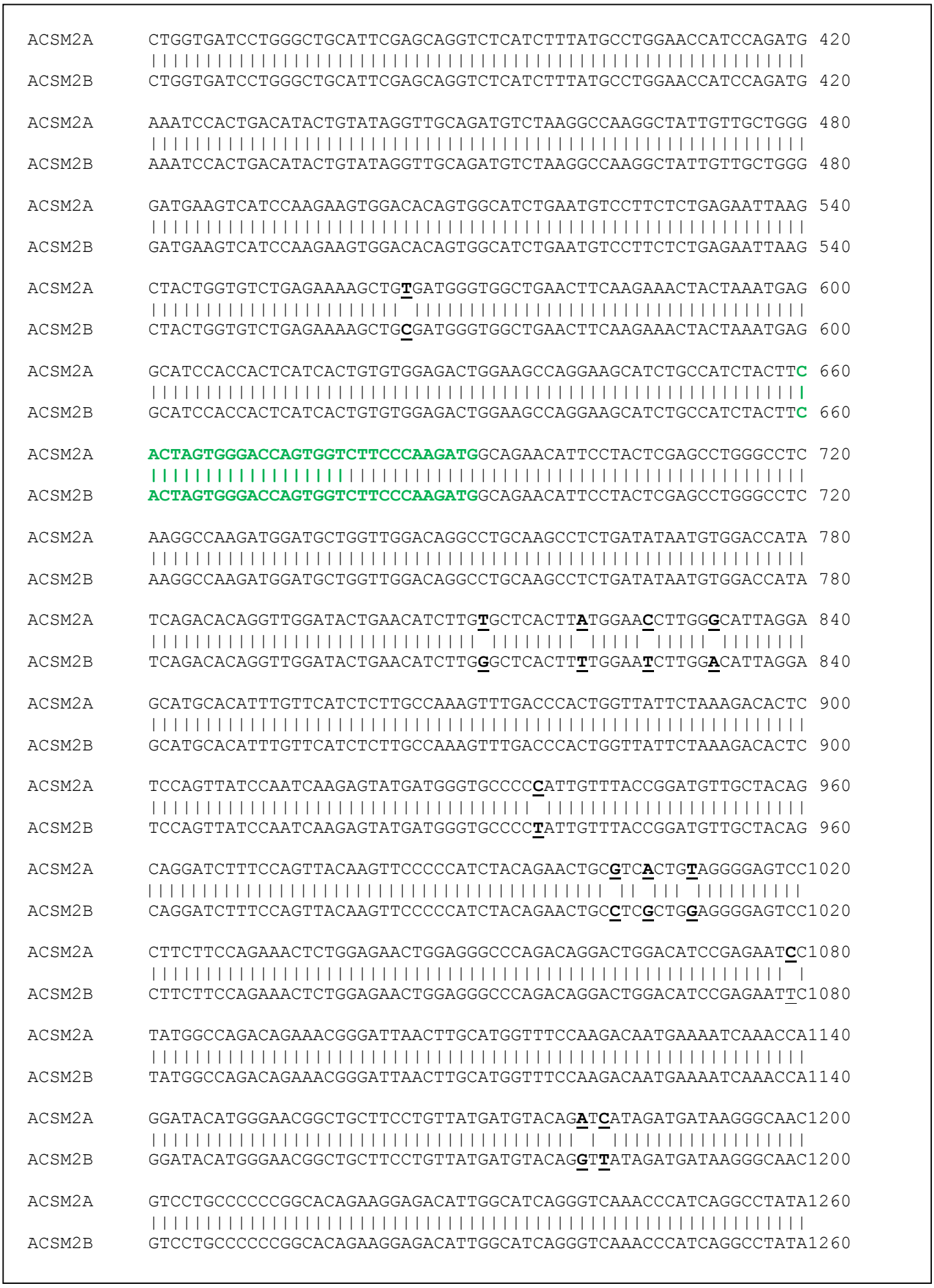
ACS: acyl coenzyme A synthetase; ACSL: long-chain ACS; ACSM; medium-chain ACS; ACSS; short-chain ACS; ACSVL: very long-chain ACS. Conserved domains were identified as described in Section 2.8.4. The position of each motif within the open reading frame is indicated. The presence of Motif II in human ACSs is indicated by a highly conserved arginine residue **(R)** (shown in red boldface). The symbol *X* in the consensus sequence indicates any amino acid, while the symbol *h* indicates any of the hydrophobic amino acids (I; L; V; M; F).

Before, fatty acid activation activity was only characterized biochemically by their chain length preference. As mentioned in Section 2.7, these enzymes have overlapping fatty acid specificity (Knights, 2000). As a result, defining the substrate profile of these enzymes has therefore been challenging. The discovery of these conserved motifs is therefore extremely beneficial in not only the classification of these ACSs, but also the discovery of new candidate human ACS genes and proteins.

2.9.5 Non-synonymous substitution essential for the distinction between the ACSM2A and *ACSM2B* genes

As discussed in Section 2.9.3, The *ACSM2A* and *ASCM2B* genes are extremely difficult to distinguish as the coding sequences of these genes are 98.8% identical, and their amino acid sequences are 97.6% identical. The *ACSM2A* and *ACSM2B* gene have 21 nucleotide differences in the coding region. There are 17 non-synonymous substitutions and 4 synonymous substitutions (Table 2.8) (Figure 2.5). Only one of the amino acid changes resulting from a non-synonymous substitution lies within a conserved motif. Residue 463, found in Motif II, is Asn in the *ACSM2A* gene and Asp in the *ACSM2B* gene (Table 2.9)

```
ACSM2A     ATGCATTGGCTGCGAAAAGTTCAGGGACTTTGCACCCTGTGGGGTACTCAGATGTCCAGC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ACSM2B     ATGCATTGGCTGCGAAAAGTTCAGGGACTTTGCACCCTGTGGGGTACTCAGATGTCCAGC 60

ACSM2A     CGCACTCTCTACATTAATAGTAGGCAACTGGTGTCCCTGCAGTGGGGCCACCAGGAAGTG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ACSM2B     CGCACTCTCTACATTAATAGTAGGCAACTGGTGTCCCTGCAGTGGGGCCACCAGGAAGTG 120

ACSM2A     CCGGCCAAGTTTAACTTTGCTAGTGATGTGTTGGATCACTGGGCTGACATGGAGAAGGCT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ACSM2B     CCGGCCAAGTTTAACTTTGCTAGTGATGTGTTGGATCACTGGGCTGACATGGAGAAGGCT 180

ACSM2A     GGCAAGCGACTCCCAAGCCCAGCCCTGTGGTGGGTGAATGGGAAGGGGAAGGAATTAATG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ACSM2B     GGCAAGCGACTCCCAAGCCCAGCCCTGTGGTGGGTGAATGGGAAGGGGAAGGAATTAATG 240

ACSM2A     TGGAATTTCAGAGAACTGAGTGAAAACAGCCAGCAGGCAGCCAACGTCCTCTCGGGAGCC 300
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
ACSM2B     TGGAATTTCAGAGAACTGAGTGAAAACAGCCAGCAGGCAGCCAACATCCTCTCGGGAGCC 300

ACSM2A     TGTGGCCTGCAGCGTGGGGATCGTGTGGCAGTGGTGCTGCCCCGAGTGCCTGAGTGGTGG 360
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
ACSM2B     TGTGGCCTGCAGCGTGGGGATCGTGTGGCAGTGATGCTGCCCCGAGTGCCTGAGTGGTGG 360
```


ACSM2A CTGGTGATCCTGGGCTGCATTCGAGCAGGTCTCATCTTTATGCCTGGAACCATCCAGATG 420
ACSM2B CTGGTGATCCTGGGCTGCATTCGAGCAGGTCTCATCTTTATGCCTGGAACCATCCAGATG 420
ACSM2A AAATCCACTGACATACTGTATAGGTTGCAGATGTCTAAGGCCAAGGCTATTGTTGCTGGG 480
ACSM2B AAATCCACTGACATACTGTATAGGTTGCAGATGTCTAAGGCCAAGGCTATTGTTGCTGGG 480
ACSM2A GATGAAGTCATCCAAGAAGTGGACACAGTGGCATCTGAATGTCCTTCTCTGAGAATTAAG 540
ACSM2B GATGAAGTCATCCAAGAAGTGGACACAGTGGCATCTGAATGTCCTTCTCTGAGAATTAAG 540
ACSM2A CTACTGGTGTCTGAGAAAAGCTGTGATGGGTGGCTGAACTTCAAGAAACTACTAAATGAG 600
ACSM2B CTACTGGTGTCTGAGAAAAGCTGCGATGGGTGGCTGAACTTCAAGAAACTACTAAATGAG 600
ACSM2A GCATCCACCACTCATCACTGTGTGGAGACTGGAAGCCAGGAAGCATCTGCCATCTACTTC 660
ACSM2B GCATCCACCACTCATCACTGTGTGGAGACTGGAAGCCAGGAAGCATCTGCCATCTACTTC 660
ACSM2A ACTAGTGGGACCAGTGGTCTTCCCAAGATGGCAGAACATTCCTACTCGAGCCTGGGCCTC 720
ACSM2B ACTAGTGGGACCAGTGGTCTTCCCAAGATGGCAGAACATTCCTACTCGAGCCTGGGCCTC 720
ACSM2A AAGGCCAAGATGGATGCTGGTTGGACAGGCCTGCAAGCCTCTGATATAATGTGGACCATA 780
ACSM2B AAGGCCAAGATGGATGCTGGTTGGACAGGCCTGCAAGCCTCTGATATAATGTGGACCATA 780
ACSM2A TCAGACACAGGTTGGATACTGAACATCTTGTGCTCACTTATGGAACCTTGGGCATTAGGA 840
ACSM2B TCAGACACAGGTTGGATACTGAACATCTTGGGCTCACTTTTGGAATCTTGGACATTAGGA 840
ACSM2A GCATGCACATTTGTTCATCTCTTGCCAAAGTTTGACCCACTGGTTATTCTAAAGACACTC 900
ACSM2B GCATGCACATTTGTTCATCTCTTGCCAAAGTTTGACCCACTGGTTATTCTAAAGACACTC 900
ACSM2A TCCAGTTATCCAATCAAGAGTATGATGGGTGCCCCCATTGTTTACCGGATGTTGCTACAG 960
ACSM2B TCCAGTTATCCAATCAAGAGTATGATGGGTGCCCCTATTGTTTACCGGATGTTGCTACAG 960
ACSM2A CAGGATCTTTCCAGTTACAAGTTCCCCCATCTACAGAACTGCGTCACTGTAGGGGAGTCC1020
ACSM2B CAGGATCTTTCCAGTTACAAGTTCCCCCATCTACAGAACTGCCTCGCTGGAGGGGAGTCC1020
ACSM2A CTTCTTCCAGAAACTCTGGAGAACTGGAGGGCCCAGACAGGACTGGACATCCGAGAATCC1080
ACSM2B CTTCTTCCAGAAACTCTGGAGAACTGGAGGGCCCAGACAGGACTGGACATCCGAGAATTC1080
ACSM2A TATGGCCAGACAGAAACGGGATTAACTTGCATGGTTTCCAAGACAATGAAAATCAAACCA1140
ACSM2B TATGGCCAGACAGAAACGGGATTAACTTGCATGGTTTCCAAGACAATGAAAATCAAACCA1140
ACSM2A GGATACATGGGAACGGCTGCTTCCTGTTATGATGTACAGATCATAGATGATAAGGGCAAC1200
ACSM2B GGATACATGGGAACGGCTGCTTCCTGTTATGATGTACAGGTTATAGATGATAAGGGCAAC1200
ACSM2A GTCCTGCCCCCCGGCACAGAAGGAGACATTGGCATCAGGGTCAAACCCATCAGGCCTATA1260
ACSM2B GTCCTGCCCCCCGGCACAGAAGGAGACATTGGCATCAGGGTCAAACCCATCAGGCCTATA1260

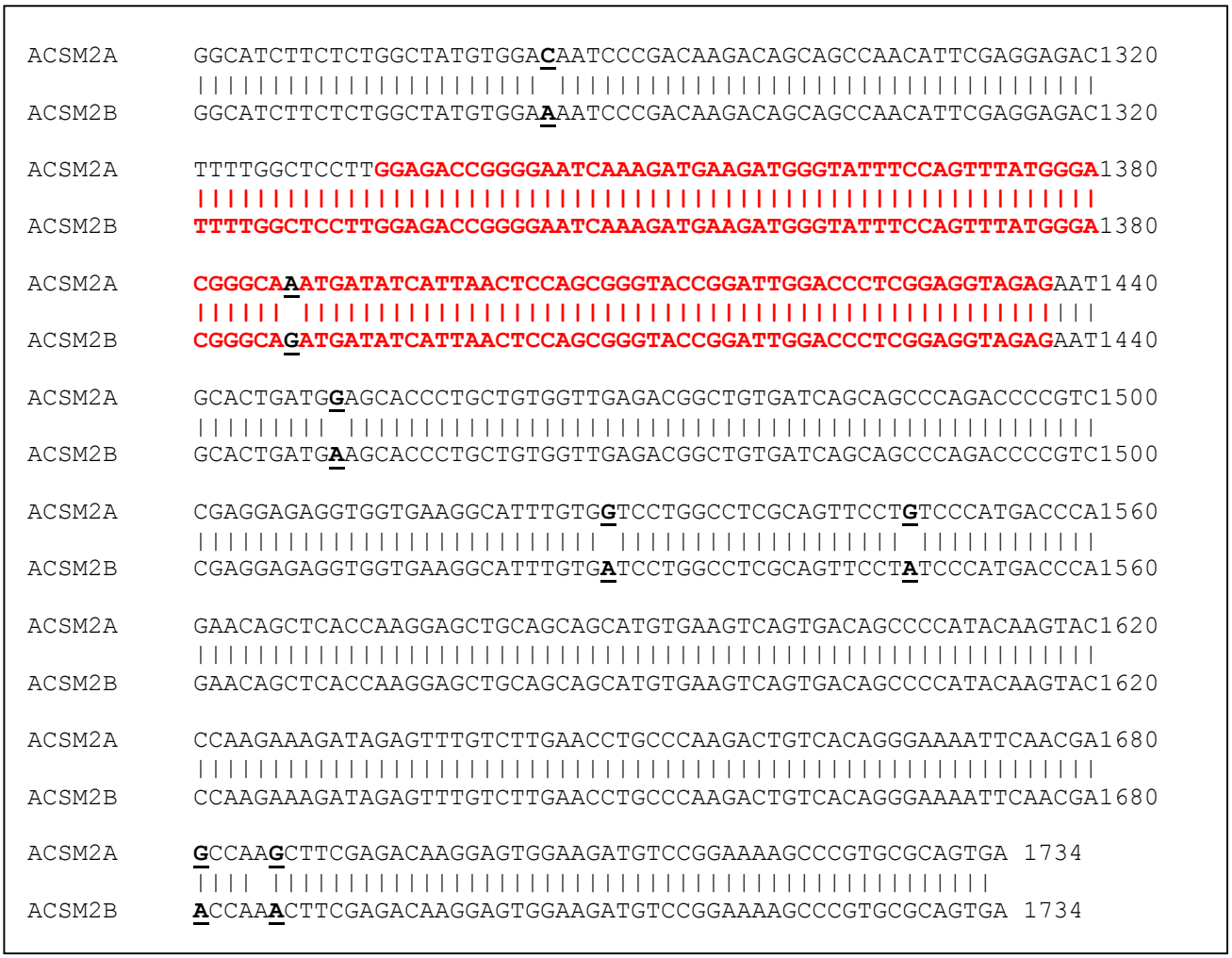


Figure 2.5: Alignment of the human *ACSM2A* gene (ENST00000066813) and the *ACSM2B* gene (ENST00000567001) amino acid coding sequences, indicating the 21 nucleotide differences in the coding region. Shown in black boldface is the 21 nucleotide differences in the coding region (performed by use of CLUSTAL multiple sequence alignment) (Li *et al*., 2015, https://www.ebi.ac.uk/Tools/msa/). Motif I is tremendously well conserved in bacteria to human ACSs and consists of a 10 amino acid sequence, starting at about residue 260 (Table 2.7), indicated in green boldface) (Figure 2.5). Motif II consists of 44–45 residues (Table 2.7) indicated in red boldface.

Table 2.8: Nucleotide differences in the coding region of *ACSM2A* and *ACSM2B* indicating the 17 non-synonymous substitutions and 4 synonymous substitutions.

Position	Alleles	Amino acids	Codons	Consequence
285	T/C	I/V	Atc/Gtc	Non-synonymous
334	G/A	V/M	Gtg/Atg	Non-synonymous
811	T/G	C/G	Tgc/Ggc	Non-synonymous
820	A/T	M/L	Atg/Ttg	Non-synonymous
826	C/T	P/S	Cct/Tct	Non-synonymous
832	G/A	A/T	Gca/Aca	Non-synonymous
1003	G/A/C	V/I/L	Gtc/Atc/Ctc	Non-synonymous
1006	A/G	T/A	Act/Gct	Non-synonymous
1010	T/G	V/G	gTa/gGa	Non-synonymous
1079	A/G	F/S	tTc/tCc	Non-synonymous
1180	A/G	I/V	Atc/Gtt	Non-synonymous
1182	C/T	I/V	atC/gtT	Non-synonymous
1284	C/A	D/E	gaC/gaA	Non-synonymous
1387	A/G	N/D	Aat/Gat	Non-synonymous
1450	G/A	E/K	Gag/Aag	Non-synonymous
1528	G/A	V/I	Gtc/Atc	Non-synonymous
1681	G/A	A/T	Gcc/Acc	Non-synonymous
564	T/C	C/C	tgT/TgC	Synonymous
936	C/T	P/P	ccC/ccT	Synonymous
1548	G/A	L/L	ctG/ctA	Synonymous
1586	G/A	K/K	aaG/aaA	Synonymous

Table 2.8 represents the nucleotide differences in the coding region of *ACSM2A* and *ACSM2B* indicating the 17 non-synonymous substitutions and 4 synonymous substitutions as well as the position, alleles, amino acid changes, codons and the corresponding consequence.

Table 2.9: Non-synonymous substitutions identified within the conserved amino acid sequence motif II identified in human ACSs, essential for the distinction between the *ACSM2A* and *ACSM2B* genes (Adapted from Watkins *et al*., 2007).

ACS	MOTIF I	MOTIF II
Medium-chain ACS family		
ACSM1	225-FTSGTTGFPK-234	450-TGDRGKMDEEGYICFLGRSDDIINASGYRIGPAEVE-485
ACSM2A	220-FTSGTSGLPK-229	444-LGDRGIKDEDGYFQFMGRANDIINSSGYRIGPSEVE-479
ACSM2B	220-FTSGTSGLPK-229	444-LGDRGIKDEDGYFQFMGRADDIINSSGYRIGPSEVE-479
ACSM3	234-FTSGTSGYPK-243	459-TGDRGYMDKDGYFWFVARADDVILSSGYRIGPFEVE-494
ACSM4	228-FTSGTTGFPK-237	453-TGDRGVMDSDGYFWFVGRADDVIISSGYRIGPFEVE-488
ACSM5	229-FTSGTTGAPK-238	453-TGDRARMDKDGYFWFMGRNDDVINSSSYRIGPVEVE-488
Consensus	[Y,F]TSG[T,S]TGXPK	TGDX(7)GX(3)hX(2)RX(4)hX(3,4)GX(2)hX(4)hE

ACS: acyl coenzyme A synthetase; ACSM: medium-chain ACS; Conserved domains were identified as described in Section 2.7.4. The position of motif I and motif II within the open reading frame of each medium-chain ACS family member is indicated in the table. The presence of Motif II is indicated by a highly conserved arginine residue **(R)** shown in boldface. The non-synonymous substitution identified within Motif II is indicated in Residue 463 as Asn **(N)** in the *ACSM2A* gene and Asp **(D)** in the *ACSM2B* gene (indicated in yellow boldface).

2.10 Human genetic variation studies

The human genome consists of 3.2 billion base pairs. From the 23 chromosomes, only 1.5% of the three billion base pairs codes for proteins (Lander *et al.,* 2001). DNA is only as powerful as the proteins it produces and nucleotide variations between individuals are expected every 300 base pairs (Mancinelli *et al.,* 2000; Feuk *et al.,* 2006). Hence, between the genomes of two individuals there would exist approximately 10 million variations (Mancinelli *et al.,* 2000; Feuk *et al.,* 2006). A single nucleotide polymorphism (SNP) is non-randomly distributed and defined as a single nucleotide alteration within a DNA sequence that occurs in more than 1% of the population, while a mutation is defined as damage or alterations within a DNA sequence that occurs in less than 1% of the population. Variations are not only due to single nucleotide changes, but can also be the result of insertions, deletions, nucleotide repeats, inversions, and base pair changes within a DNA sequence. During the replication of DNA, mutations or SNPs can be introduced which are permanent and can be associated with disease (Mancinelli *et al.,* 2000; Chorley *et al.* 2008). It is not

only SNPs and mutations that can be associated with disease, but copy number variations (CNV) as well. A CNV is a segment of DNA which is either deleted or duplicated and can lead to copy number variations in a genetic sequence.

Each gene consists of introns and exons. The exons are spliced when DNA is transcribed to mRNA which can be translated into amino acids to form a protein. Incorrectly spliced nucleotides can change the amino acid sequence and thus the possibility exists that the activity of the enzyme can also change. If the splice site or branch point was changed by either a SNP or a mutation, a sequence can be incorrectly spliced. When a variation occurs in the exons, branch point or splice site, enzyme activity could be altered or the enzyme function could be terminated (McCarroll and Altshuler, 2007). When a variation causes a key amino acid to change in terms of folding within the affected protein, it can alter the function of the protein and increase susceptibility to disease (Pritchard, 2001; Chorley *et al*., 2008). Variants that occur in protein coding regions are further classified into synonymous and nonsynonymous variants. Synonymous SNPs are defined as DNA changes with no related change in the amino acid sequence, and nonsynonymous SNPs are defined as DNA changes that result in an amino acid change or insertion of a stop codon (Jaffe *et al*., 2010). The study of genetic diversity classifies an individual or population compared to other individuals or populations, the identification and characterization of specific genetic variations among individuals within a population can be detected through various available methods. Genetic variation can also be identified by examining variation at the level of enzymes using the process of protein electrophoresis. The traditional approach to molecular analysis of defects is based on identifiable clinical symptoms followed by screening for mutations, sequencing of specific known candidate genes one at a time or, more recently, in combination using next-generation sequencing (NGS) technology, for focused and genome-wide discovery (Mardis, 2008). This study will focus largely on the identification of genetic variation within the coding region of a gene. A more detailed overview of the sequencing approaches will be given in Section 2.11 to support the experimental strategy that was used in this study.

2.10.1 Next-generation Sequencing

The ability of next-generation high-throughput DNA sequencing (NGS) techniques to process millions of sequence reads in parallel has transformed genomics research in recent years by enabling vast data generation of large amounts of interpretable data to ultimately

discover genetic variants that causes rare complex diseases (Mardis, 2008; Meldrum *et al*., 2011; Pabinger *et al*., 2014). A new era of high-throughput genomic analysis (now referred to as next-generation sequencing) was introduced in 2005 with the commercial launch of the first massively parallel pyrosequencing platform, a technique based on sequencing by synthesis (Voelkerding *et al*., 2009). NGS has essentially transformed genomic research in this rather short time frame since 2005, allowing researchers to perform experiments not technically practical or affordable in the past (Mardis, 2008; Voelkerding *et al*., 2009).

Instead of sequencing a single DNA fragment, NGS extends this process across millions of fragments in a massively parallel fashion. Even though several second-generation sequencing platforms are available today that differ in their engineering configurations and sequencing chemistries, they share a common technological feature by making use of high throughput parallel sequencing through target gene enrichment, library preparation, and sequencing, generating hundreds of megabases to gigabases of nucleotide sequence output in a single instrument run (Mardis, 2008; Voelkerding *et al*., 2009; Wong, 2013). To perform massive parallel sequencing, second-generation sequencing platforms is defined by the methods used to achieve the following two principles:

1. The amplification of clonally amplified DNA templates or single DNA molecules usually through a polymerase-based clonal replication process on a solid support matrix.
2. Followed by cyclic sequencing chemistries (Margulies *et al*., 2005; Shendure *et al*., 2005; Voelkerding *et al*., 2009; Meldrum *et al*., 2011; Rothberg *et al*., 2011; Wong, 2013; Zhang *et al*., 2014).

Front-end library preparation methods across all second-generation sequencing platforms are similar, involving the addition of universal adapter sequences to the terminal ends of the DNA fragment (Meldrum *et al*., 2011). Target gene enrichment selectively enrich the coding regions of the gene of interest utilising a method of preference depending on the number and size of the genes or the total size of the targeted genes in megabases (Wong, 2013). This enriched target genes are then ultimately sequenced via repeated sequencing chemistries (Voelkerding *et al*., 2009; Meldrum *et al*., 2011). Each of the platforms available today makes use of various sequencing chemistries and methods for signal detection (Meldrum *et al*., 2011). All aspects and performance metrics, for instance cost per base, coverage fold, simplicity of the data analysis, speed and accuracy should be taken into consideration when selecting the most suitable platform for a sequencing project.

The use of exome sequencing, by only sequencing the ~1% of the genome that codes for specific protein sequences, has also become commonly used in major diagnostic centres by reducing high cost and excessive data produced from whole-genome sequencing. The amplification of a specific target region from the complete genome can easily be accomplished, which contributes greatly to clinical application in understanding human health (Meldrum *et al*., 2011; Pabinger *et al*., 2014) but this will be discussed further in Section 2.10.1.1.

2.10.1.1 Exome Sequencing

Recent technological advances have allowed researchers to study individual genomes at a base-pair resolution and have established that the protein-coding genes of the genome, known as the exomes harbours more than 15 000 genetic variants (Jaffe *et al*., 2010). Mutations in the protein-coding regions in human genes are much more likely to have severe consequences as these sequences harbour 85% of the mutations associated with disease-related traits (Liu *et al*., 2012). Compared to whole-genome sequencing of complex samples, exome sequencing can sequence samples to much higher depth than could be achieved with whole genome sequencing. By only sequencing the 1% of the genome that codes for specific protein sequences, reliable variant calls can be achieved at a significantly reduced cost (Liu *et al*., 2012; Lupski *et al*., 2013). Exome sequencing is one of the most cost-efficient sequencing approaches and perhaps as the most widely used targeted sequencing approach enabling researches to identify common single nucleotide variants (SNVs), copy number variations (CNVs), and small insertions or deletions (indels), responsible for Mendelian diseases, as well as rare de novo mutations that explain the heritability of complex diseases (Samuels *et al*., 2013). Target-enrichment methods allow one to selectively capture genomic regions of interest from a DNA sample prior to sequencing (Samuels *et al*., 2013). Capturing techniques enables the enrichment of specific sequences and can enrich just the exome, exome plus 3’ and 5’ untranslated regions (UTRs) and other functionally annotated regions of interest such as miRNA genes and various noncoding RNAs, which can significantly expand the amount of captured material (Lupski *et al*., 2013; Samuels *et al*., 2013). Exome sequencing approaches can efficiently identify variants across a wide range of applications, including population genetics, genetic disease and cancer studies.

2.11 Ion Proton Next-generation Sequencer

The Ion Proton System for Next-generation Sequencing, developed by Life Technologies, was available, suitable and thus used during this study. The Ion Proton system (referred to as the “next-generation” of semiconductor sequencing instrumentation) is an integrated, easy-to-implement workflow solution for high-throughput sequencing of exomes, transcriptomes and genomes (Boland *et al*., 2013). The Ion Proton System has been designed with affordability in mind, the system combines semiconductor sequencing technology with biochemistry to directly translate chemical information into digital data, therefore eliminating the need for expensive optics and complex sequencing chemistries. The Ion Proton NGS system combines simple bioinformatics through point-and-click run setup and data analysis in Torrent Suite Software v5.0 combined with exome specific workflows in Ion Reporter Software. High-throughput chips enable the Ion Proton Sequencer to sequence a human exome with similar run times and single-day workflow as the Ion Torrent Personal Genome Machine (PGM), a benchtop sequencing system that is still the fastest and most affordable solution for the sequencing of sets of genes, small genomes and small RNA (Boland *et al*., 2013). Users of the Ion Proton System can sequence exomes and whole transcriptomes on the Ion Proton I Chip.

2.11.1 Technology of the Ion Proton NGS System

The Ion Proton System is a complementary platform to the Ion PGM System. Granting that the PGM System produces megabases (Mbs) of DNA sequence within hours, its net output is insufficient to generate whole exome sequencing (WES) data with a sufficient high coverage per base (Boland *et al*., 2013). With chip densities up to 1,000-fold greater than the Ion PGM Sequencer, the Ion Proton System is designed to generate gigabases (Gbs) of data of high quality sequencing results from DNA library to variants in a single day to identify, prioritize, and report on the most biologically interesting variants in real-time (Boland *et al*., 2013). Combined with the enormous amount of computing power in the Ion Proton Sequencer as well as the Ion Proton Torrent Server, the Ion Proton System relies on the basic principle of “sequencing-by-synthesis”, through detecting chemical changes within a reaction on a semiconductor chip. The incorporation of a nucleotide into a strand of DNA by a polymerase, releases a hydrogen ion (H^+) from the elongating 3’OH end of the complementary DNA strand as a by-product. This hydrogen ion carries a charge that the

Proton System's ion sensor, acting essentially as n solid-state pH meter, can detect. As the sequencer floods the chip (Ion PI Chip with 165 million wells) with one nucleotide after another, any nucleotide added to a DNA template will be detected as a recordable voltage change, and the Ion Proton System will call the base, if a nucleotide is not a match for a particular template, no voltage change will be detected and no base will be called for that template (Niedringhaus *et al.,* 2011, Merriman *et al.,* 2012; Boland *et al*., 2013).

2.12 Classification of variation

This section gives a brief description of the nomenclature used in this dissertation. The consistent use of uniform nomenclature as described by the Human Genome Variation Society (HGVS) (HGVS, http://www.hgvs.org/mutnomen) in the management of changes (mutations and polymorphisms) in DNA and protein sequences is essentially important for efficient and accurate reporting (den Dunnen and Antonarakis, 2000). As the terminology to describe genetic variants can be confusing, the Human Genome Organization (HUGO) Mutation Database Initiative (MDI) initiated an ad-hoc committee to construct rules for the description of sequence variants as the nomenclature used by different laboratories can vary (den Dunnen and Paalman, 2003). A number of discussion papers describing the rules for the description of sequence changes have been published by the nomenclature committee that are widely accepted at this time (Antonarakis *et al*., 1998; den Dunnen and Antonarakis, 2000; den Dunnen and Paalman, 2003).

The frequently used term 'variant' in this dissertation indicates a small genetic variation identified in a genetic sequence, either with a benign or deleterious genetic effect. The following standards used for terminology of the variants identified are briefly described:

- Nucleotide substitutions are identified as e.g. c.26G>T, where the prefix 'c' signifies the sequencing type, coding DNA reference sequence (cDNA), or the standard reference referring to genomic (gDNA) as used for sequences in this study. The symbol greater than (>) indicates the base substitution of the 26nd nucleotide G to T.
- Protein substitutions are expressed as e.g. p.Ser43His, where the prefix 'p' indicates a protein change, followed by the codon number of the serine to histidine amino acid change caused by the variant.

- More than one variant on an allele in the same gene are designated as compound heterozygous variants, expressed as e.g. c.[26G>T] + [648G>T] and p.[Ser43His] + [Met52Ala].
- The reference SNP (RefSNP) refers to the database of SNPs that have previously been uniquely identified and reported and have an applied reference number (rs). The rs number found in the NCBI dbSNP (database for SNP variants) (http://www.ncbi.nml.nih.gov/snp/) specify the frequency of the SNP identified in specific populations, providing additional information if available.

2.13 Problem statement

Recent studies conducted by Badenhorst *et al*., (2013), Badenhorst *et al*., (2014) and Van der Sluis *et al*., (2015) emphasised the essential role of the *GLYAT* gene and also called attention to the glycine conjugation pathway as an essential detoxification pathway in humans. A study by Badenhorst *et al*., (2014) on the importance of glycine conjugation in the metabolism of aromatic acids, presented a schematic representation of the glycine conjugation pathway (Figure 2.6). Badenhorst *et al*., (2014) suggested that the glycine conjugation pathway should rather be viewed as a part of the phenylpropionate catabolism pathway. The part of the pathway on the grey background (Figure 2.6) represents the natural glycine conjugation pathway, as suggested by Badenhorst *el al*., (2014):

1) Dietary polyphenols are converted to simple aromatic acids such as phenylpropionate, cinnamate, and benzoate by the gut microbiota.
2) After absorption and transport to the liver, aromatic acids are activated to acyl CoA thioesters by the action of ATP dependent acid: CoA ligases.
3) Phenylpropionyl-CoA is converted to cinnamoyl-CoA by medium-chain acyl CoA dehydrogenase (the first step of the β-oxidation cycle).
4) Cinnamoyl-CoA is further oxidized to benzoyl CoA and acetyl CoA.
5) Benzoyl-CoA is then converted to hippuric acid by glycine N-acyltransferase.
6) The β-oxidation of phenylpropionate generates $FADH_2$, NADH, and acetyl CoA, which are used by the TCA-cycle and the ETC to produce ATP (As discussed in Section 2.6)
7) When the capacity of the glycine conjugation pathway is exceeded, benzoyl CoA will accumulate, resulting in CoASH sequestration (As discussed in Section 2.6.3).

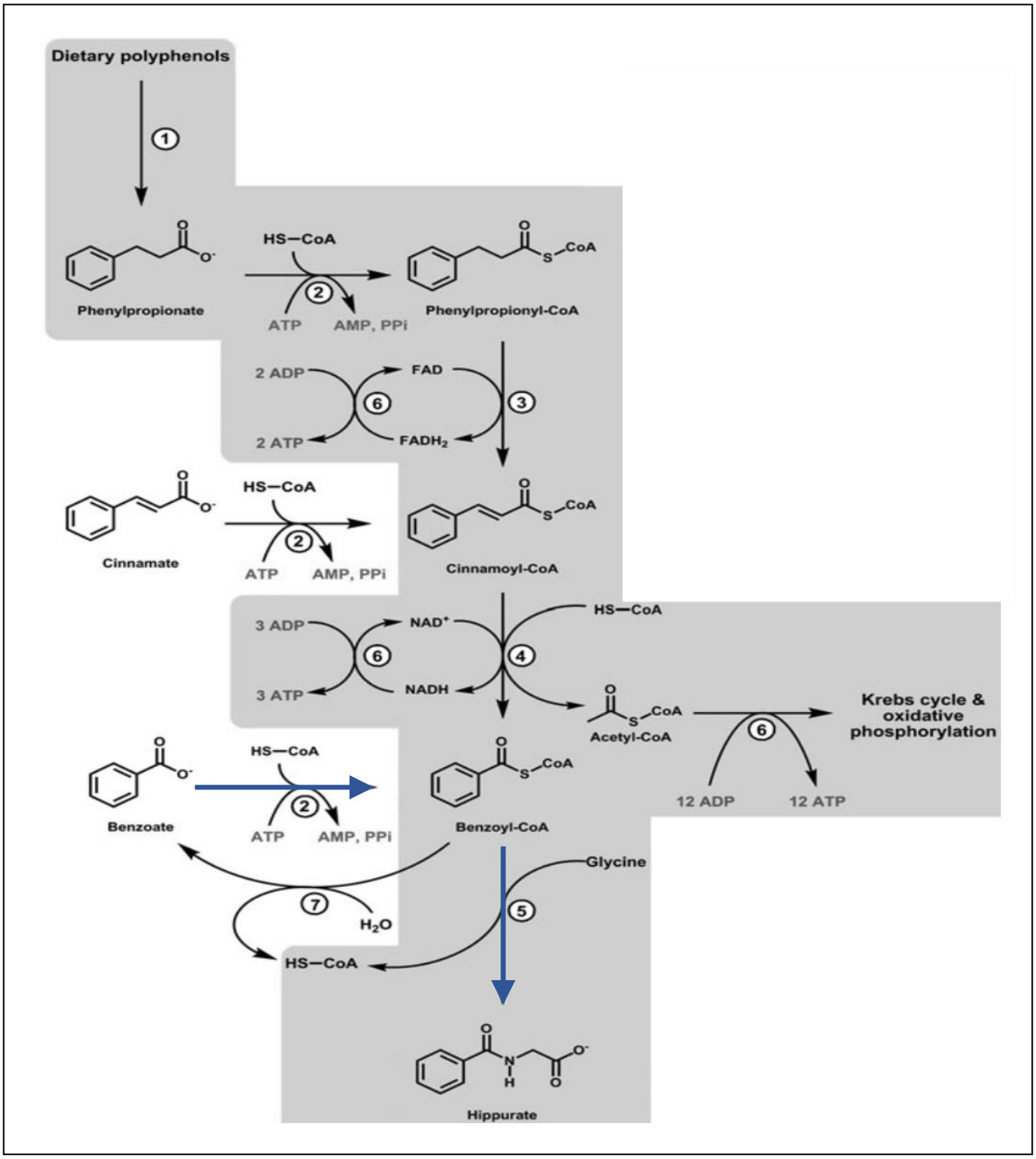


Figure 2.6: Schematic representation of the evolutionary role of GLYAT within the glycine conjugation pathway (Adapted from Badenhorst *et al*., 2014). The part of the pathway on the grey background represents the natural glycine conjugation pathway, as suggested by Badenhorst *el al*., (2014). The part of the pathway indicated by the blue arrows represents the glycine conjugation of benzoic acid leading to hippuric acid formation as presented in the present study. The first step consists of the initial activation of the carboxylic acid with ATP, generating AMP and inorganic pyrophosphate. This is followed by the reaction of the bound acyladenylate with CoASH catalysed by an ATP-dependent acid: CoA ligases. The medium-chain CoA synthetase (ACSM) (EC

6.2.1.2) has been identified in humans as HXM-A. Benzoyl-CoA is then converted to hippuric acid by glycine N-acyltransferase (GLYAT) (EC 2.3.1.13).

No defects for the glycine conjugation pathway have been reported, and studies by Badenhorst *et al*., (2013), Badenhorst *et al*., (2014) and Van der Sluis *et al*., (2015), indicated that given that the *GLYAT* gene is essential for survival, the *GLYAT* ORF should be highly conserved (Badenhorst *et al*., 2013, Badenhorst *et al*., 2014, Van der Sluis *et al*., 2015).

As mentioned in Section 2.4.1 and evident in Figure 2.6 (Pathway indicated by the blue arrows), hippuric acid is formed from benzoic acid in two steps, i) Benzoic acid is initially activated to benzoyl CoA by the mitochondrial ATP dependent acid: CoA ligases, and has been identified in humans as HXM-A (EC 6.2.1.2), encoded by the *ACSM2B* gene, ii) Next, glycine N-acyltransferase (GLYAT, EC 2.3.1.13) binds benzoyl CoA and catalyses the acylation of glycine to form hippuric acid and CoASH.

In humans, HXM-A, encoded by the *ACSM2B* gene has been identified as the active mitochondrial ATP dependent acid: CoA ligase, responsible for catalysing the initial reaction of amino acid conjugation. It has become significantly clear that little to no research have been conducted on the Acyl CoA synthetase genes (*ACSM2A* and *ACSM2B*) and their importance in the glycine conjugation pathway. Supporting this statement is the fact that current literature regarding the acyl CoA synthetase genes does not assign the correct reference transcript of the *ACSM2A* and *ACSM2B* genes to identified variants, by only identifying the gene as *ACSM2*. Before this study no information existed on the identification of genetic variation in the acyl CoA synthetase genes (*ACSM2A* and *ACSM2B*) specifically related to the South African Afrikaner Caucasian population (Kruse *et al*., 2009; Xu *et al*., 2009; Xu *et al*., 2011; Xu *et al*., 2012; Heathfield *et al*., 2013; Rodriguez-Murillo *et al*., 2014; Van der Sluis *et al*., 2015). This is also the first worldwide population analysis study of the non-synonymous SNPs within the *ACSM2A* and *ACSM2B* genes.

As no variations have been reported for the *ACSM2B* gene, and with this the increased levels of the exposure of humans to benzoic acid in recent times, the question exists on whether the *ACSM2B* ORF is also as highly conserved as *GLYAT* (Badenhorst *et al*., 2014).

2.14 Aims, objectives and experimental strategy

In an effort to further investigate the importance of the glycine conjugation pathway in humans, the focal point of this study centred on the acyl CoA synthetase genes. As recent studies by Badenhorst *et al*., (2013), Badenhorst *et al*., (2014) and Van der Sluis *et al*., (2015) particularly focused on the important role of GLYAT in the glycine conjugation pathway, this study subsequently focused specifically on the *ACSM2B* gene encoding HXM-A, responsible for the first step in the glycine conjugation pathway of benzoic acid. However, the *ACSM2A* gene was also investigated, as there is some confusion in current literature regarding the distinction of these two highly similar genes.

In an effort to investigate if the *ACSM2B* ORF is as highly conserved as *GLYAT* the present study was formulated. Figure 2.7 illustrates the schematic strategy devised to address the aim of this study which was to identify genetic variation in the acyl CoA synthetase genes (*ACSM2A* and *ACMS2B*).

The following objectives were formulated to achieve this aim:

1) Analysis of the genetic variation data of the human acyl CoA synthetase medium chain family member 2A (*ACSM2A*) and the acyl CoA synthetase medium-chain family member 2B (*ACSM2B*) ORF available on public databases.
2) Generation of DNA sequence data of the coding region of a small cohort of South African Afrikaner Caucasian individuals using the Ion Torrent Proton platform for NGS.
3) Identification of previously reported and potential novel SNPs by using a bioinformatics workflow.
4) Comparison of all population data available in this study against the South African Afrikaner Caucasian population.

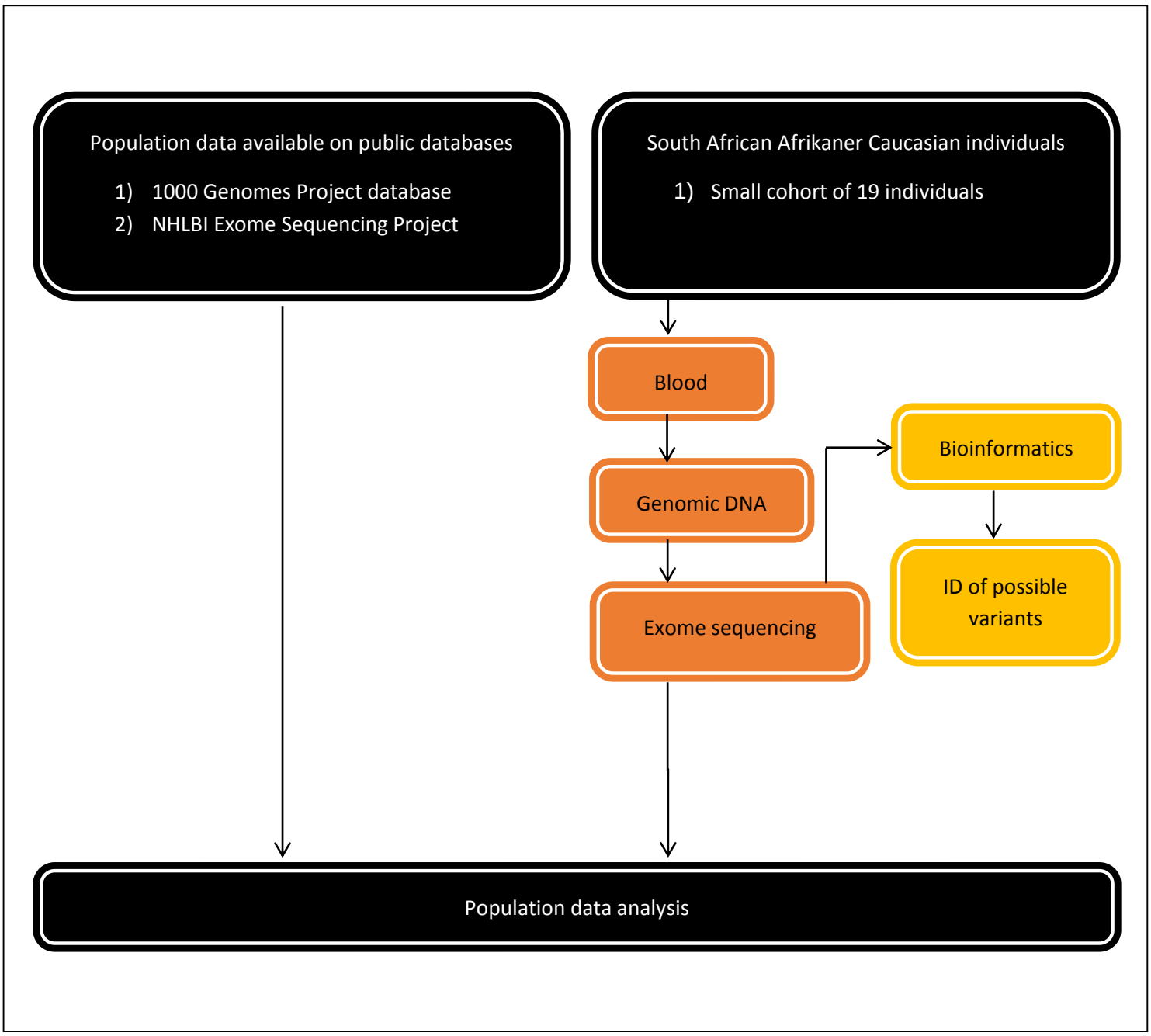


Figure 2.7: Schematic representation of the experimental strategy used to achieve the aim and objective. This study consisted of four main objectives. Starting with the analysis of the genetic variation data available on public databases; generation of DNA sequence data using Ion Proton NGS for exome sequencing and the Ion Torrent Proton platform for NGS; bioinformatics workflow and lastly, ultimate interpretation of the results. NHLBI: National Heart, Lung and Blood Institute.

CHAPTER 3: MATERIALS AND METHODS

3.1 Introduction

Chapter 3 presents further detail of the study design and materials, along with the methods and technologies that were used to conduct this study. As discussed in Chapter 2 the identification of genetic variation in the acyl CoA synthetase genes (*ACSM2A* and *ACSM2B*) requires a variety of investigations. In this study, data available in public databases was analysed to identify genetic variation in the *ACSM2A* and *ACSM2B* ORF available in public databases. Data from the 1000 Genomes Project and the NHLBI Exome Sequencing Project was used. A small cohort of healthy South African Afrikaner Caucasian (SA) individuals was also included in this study, by sequencing the coding regions of their *ACSM2A* and *ACSM2B* genes. This is an important addition to this study as there is no data available on genetic variation of the *ACSM2A* and *ACSM2B* genes in the SA population, as well as the fact that only a small number of studies have concentrated on the analysis of genetic variation within the South African Afrikaner Caucasian population (Kruse *et al*., 2009; Xu *et al*., 2009; Xu *et al*., 2011; Xu *et al*., 2012; Heathfield *et al*., 2013; Rodriguez-Murillo *et al*., 2014; Van der Sluis *et al*., 2015).

Fresh blood samples were obtained from 19 healthy SA individuals and exome sequencing was performed on the Ion Proton NGS platform, followed by a bioinformatics workflow for gene variant identification. The methods used to produce the key information of this study are described in full detail in Chapter 3. The results and discussion is presented in Chapter 4.

3.2 Variation data obtained from public databases

Variation data of the *ACSM2A* and *ACSM2B* genes were obtained from the following public databases: 1) the 1000 Genomes Project (Abecasis *et al*., 2012), and 2) the NHLBI Exome Sequencing Project (Fu *et al*., 2013)

3.2.1 The 1000 Genomes Project

The 1000 genomes project (http://www.1000genomes.org) launched in January 2008, is an international research attempt to establish the most detailed catalogue of human genetic variation designed to assist in the understanding of the genetic contribution to disease. The genomes of 2015 (1000 Genomes Project Phase 3 data) individuals from 14 populations (obtained from Europe, East Asia, sub-Saharan Africa and America) (Table 3.1) were included in the project, structured using a low-coverage whole-genome and exome sequencing combination, in order to provide perspective on genetic variations in diverse populations (Abecasis *et al*., 2012).

3.2.2 The NHLBI Exome Sequencing Project

The exomes of 6,503 individuals including 2,203 European-Americans (EAs) and 4,300 African-Americans (AAs) (Table 3.1) were sequenced as part of the NHLBI Exome Sequencing Project (http://evs.gs.washington.edu/EVS/). The overall aim of the Exome Sequencing Project (ESP) is to communicate to the scientific community the discovery of novel genes and mechanisms contributing to heart, lung and blood disorders by the application of NGS of the protein-coding regions of the genome (Tennessen *et al*., 2012; Fu *et al*., 2013).

Table 3.1: Summary of the population data used in the present study obtained from public databases (The 1000 genomes project and the NHLBI ESP).

Public database	Population	Sub-Population	
1000 Genomes Project	AFR (African)	ACB	African Caribbean in Barbados
		ASW	African Ancestry in Southwest US
		ESN	Esan in Nigeria
		LWK	Luhya in Webuye, Kenya
		MAG	Mandinka in The Gambia
		MSL	Mende in Sierra Leone
		YRI	Yoruba in Ibadan, Nigeria
	AMR (American)	CLM	Colombian in Medellin, Colombia
		MXL	Mexican Ancestry in Los Angeles, California
		PEL	Peruvian in Lima, Peru
		PUR	Puerto Rican in Puerto Rico
	EAS (East Asian)	CDX	Chinese Dai in Xishuangbanna, China
		CHB	Han Chinese in Beijing, China
		CHS	Southern Han Chinese, China
		JPT	Japanese in Tokyo, Japan
		KHV	Kinh in Ho Chi Minh City, Vietnam
	EUR (European)	CEU	Utah residents with Northern and Western European ancestry
		FIN	Finnish in Finland
		GBR	British in England and Scotland
		IBS	Iberian population in Spain
		TSI	Toscani in Italy
NHLBI ESP	AA (African American)	N/A	N/A
	EU (European American)	N/A	N/A

N/A: Not available

3.3 Ethics and participants (South African Afrikaner Caucasian population samples)

Ethical approval was acquired by Dr. R Van der Sluis, Division for Biochemistry, School of Physical and Chemical Sciences, North-West University, Potchefstroom Campus, Potchefstroom for the collection of blood from all individuals used in this study. Ethical approval under project number NWU-0096-08-A1, was obtained from the Ethics Committee of the NWU for the work described here which was performed at the NWU.

For this study, exome sequencing data for a total of nineteen healthy individuals from the Biotransformation and Oxidative Stress (BOSS) Laboratory at the Centre for Human

Metabolomics, Biochemistry Division, North-West University, Potchefstroom, South Africa, was included. Participant age and gender were not included as contributing factors to variation in the glycine conjugation pathway and will not be discussed during this study. All participants gave written informed consent.

3.4 DNA isolation and Quantification

The reagents, components, and instruments used during DNA isolation and quantification were provided in the FlexiGene DNA kit (Qiagen, Cat. No. 51204) and NanoDrop 1000 Spectrophotometer system respectively. The use of these are discussed below.

3.4.1 DNA isolation

Genomic DNA was isolated from human whole blood using the FlexiGene DNA kit (Qiagen, Cat. No. 51204) according to the protocols provided. 250 µl of Buffer FG1 were added to 100 µl of the whole blood samples into 1.5 ml centrifuge tubes and mixed by inverting each tube 5 times. The samples were centrifuged for 20 s at 10,000 × *g* in a fixed-angle rotor. The supernatant was discarded and the tubes were left inverted on a clean sheet of absorbent paper for 2 min, taking care that the pellet remained in the tube. 50 µl of Buffer FG2/QIAGEN Protease were added to each tube and vortexed until the pellet was completely homogenized. Next the tubes were briefly centrifuged for a few seconds and placed in a heating bock or water bath to incubate for 5 min at 65˚C. 50 µl isopropanol (100%) were added to each sample and mixed thoroughly by inversion until the DNA precipitate became visible as threads or a clump. Next, the samples were centrifuged for 3 min at 10,000 × *g* and the supernatant discarded, taking care again that the pellet remained in the tube. 50 µl of 70% ethanol were added and vortexed for 5 s. Next, the samples were centrifuged for 3 min at 10,000 × g, the supernatant was discarded and the tubes were left inverted for 5 min on a clean piece of absorbent paper, taking care that the pellet remained in the tube. The DNA pellet was then left to air-dry for at least 5 min until all of the liquid had evaporated. Finally, 100 µl of water was added and vortexed at a low speed for 5 s and incubated for 1 hour ar 65˚C in a heating block or water bath to dissolve the DNA.

3.4.2 DNA quantification

A final volume of 100 µl was obtained after DNA isolation and the total DNA was quantified before sequencing on the Ion Proton NGS sequencer. The DNA concentration was determined using a NanoDrop 1000 Spectrophotometer system by measuring 1 µl samples with high accuracy and reproducibility. This is essential for the determination of the concentration and purity of nucleic acid. This method relies on the principle of the Beer Lambert Law. The ratio of sample absorbance at 260 and 280 nm is used to assess the purity of DNA and RNA (A ratio of ~1.8 is overall accepted as “pure”) the ng/uL sample concentration is based on the absorbance at 260 nm and the selected analysis constant. Samples were sent to the Central Analytical Facility (CAF) of the University of Stellenbosch for exome sequencing.

3.5 Next-Generation Sequencing

This study made use of the Ion Proton System (Thermo Fisher Scientific, 4476610) for NGS as previously mentioned in Section 2.11. This instrument and associated chemistry was chosen as it was available at the Stellenbosch University CAF unit and provides a benchtop sequencing system for accurate, affordable high-throughput sequencing of exomes from DNA-to-variants called in a single day.

3.6 Data analysis and bioinformatics workflow

The data analysis and bioinformatics workflow used in this study is represented in Figure 3.1.

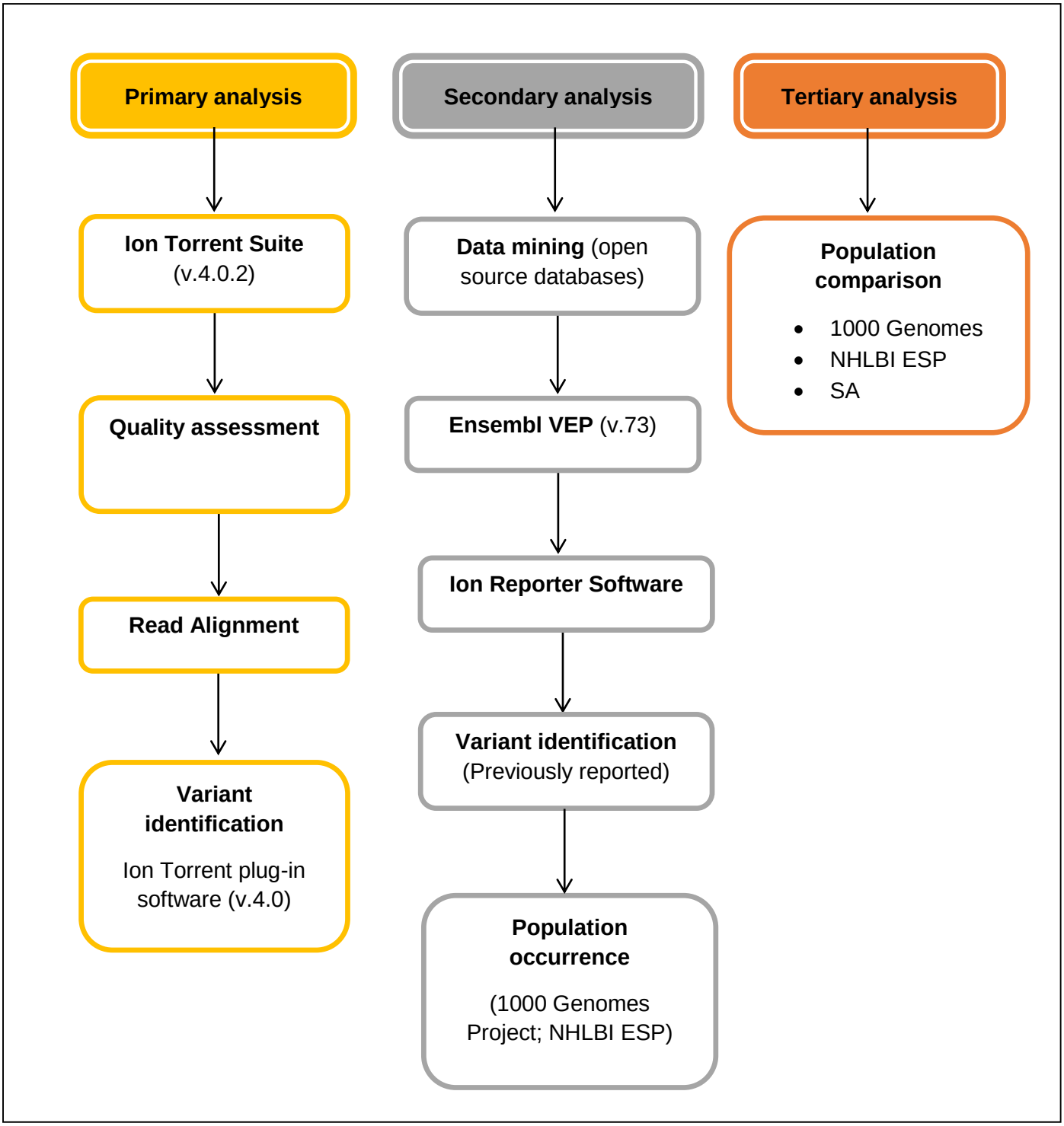


Figure 3.1: Schematic presentation of the workflow followed to perform data analysis. The workflow consists of primary analysis performed using the Ion Torrent suite, followed by data mining as secondary analysis by making use of various available open source databases, leading to tertiary data analysis for the identification of conserved alleles across populations.

Data generated by the Ion Proton NGS was automatically transferred to the standard Ion Proton System Torrent Suite Software (v4.0.2). The DNA sequences for primary data analyses consisting of basic interpretations of raw sequence data obtained, was produced by Torrent Suite software. This included the total number of amplicon -reads, -lengths, -depths, coverage obtained and quality scores of the data acquired. Low quality sequences were eliminated by quality assessment automatically trimming and filtering the raw data. The high quality reads for each of the nineteen individuals were aligned against the reference human genome (Homo sapiens, GrCh37). This allowed information to be generated on the variants identified in each individual as well as an estimation of allele frequencies (%), and information regarding all detected variants identified within the nineteen individuals by use of Torrent Variant Caller plug-in software (v4.0).

The secondary part of this workflow consisted of an in-house developed pipeline for data mining and the identification of possible novel and/or previously reported variants. Available open source databases were used during this next part of the workflow to run VCF data files obtained from the Torrent Suite server (v4.0.2) through the Ensemble online VEP runner (variant effect predictor, v73, http://www.ensembl.org) (McLaren *et al*., 2010). Ion Reporter software was also used to access the genotype information of each variant identified in the 19 individuals (https://ionreporter.thermofisher.com/). The output files contained the variants of significance in a variant caller file (VCF). Each previously identified variant is present in the NCBI dbSNP (http://www.ncbi.nml.nih.gov/snp/) and thus has a unique SNP rs number (Sherry *et al*., 2001) as mentioned in Section 2.11.

Further analysis of the data of the *ACSM2A* and *ACSM2B* ORF analysed the variants occurring in specific populations. This data was retrieved from the 1000 Genomes Project (1000 Genomes Project Consortium, 2012) and, the NHLBI Exome sequencing project (Tennessen *et al*., 2012; Fu *et al*., 2013).

Tertiary analysis and ultimate interpretation of the results were done by making use of population comparison studies of identified variants to ultimately prove the hypothesis that genetic variation in the open reading frame (ORF) of the acyl CoA synthetase genes (*ACSM2A* and *ACSM2B*) should be low.

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Introduction

Nucleotide variation of the *ACSM2A* and *ACSM2B* genes needs to be characterized to identify whether the ORF is as conserved as previously shown for the ORF of *GLYAT* (ENST00000344743). The aim of this study was to identify non-synonymous SNPs in the *ACSM2A* and *ACSM2B* genes using publicly available datasets, including a new cohort of South African Afrikaner Caucasian individuals. Genetic variation was characterised within populations as presented in Table 4.1 (1000 Genomes Project, NHLBI ESP, and SA). A detailed explanation of the experimental design, methods, and data analysis workflow used to generate the results obtained during this study, was presented in Chapter 3.

4.2 Inter-population data analyses

Table 4.1: Population data used for the present study

Source	Population	Description	n[1]
Present study	SA	South African Caucasian Afrikaner population	19
1000 Genomes Project	AFR	African	661
	EUR	European	503
	EAS	East Asian	504
	AMR	Admixed American	347
NHLBI Exome Sequencing Project	AA	African American	2203
	EA	European American	4300

[1]Number of individuals in the test sample; SA: South African Caucasian Afrikaner population; AFR: African; EUR: European; EAS: East Asian; AMR: Admixed American; NHLBI: National Heart, Lung and Blood Institute; AA: African American; EA: European American.

4.2.1 The 1000 Genomes Project data

The data generated by the 1000 Genomes Project included the first number of population groups (AFR; EUR; EAS; AMR) available in this study. The data received from the 1000 Genome database was used to analyse the genotype frequencies of the 47 (*ACSM2A* gene) and 43 (*ACSM2B* gene) non-synonymous SNPs identified by the 1000 Genomes Project (Table 4.2; Table 4.3).

Table 4.2: Non-synonymous SNPs and genotype frequencies identified in the 1000 Genomes sequencing data of 2015 individuals within the coding region of *ACSM2A*.

SNP	Genotype Frequency (%) of all 2015 individuals			Genotype frequency (%) of the different population groups											
				African (AFR)			American (AMR)			East Asian (EAS)			European (EUR)		
	hmz (wt)[1]	htz[1]	hmz (snp)[1]	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)
R5Q rs59292608	70.8 (G/G)	24.4 (G/A)	4.8 (AA)	97.6	2.4	0	68.7	12.4	0.9	34.5	50.6	14.9	68.0	29.6	2.4
G9R rs565557207	99.9 (G/G)	0.1 (G/A)	0 (A/A)	99.7	0.3	0	100	0	0	100	0	0	100	0	0
L10R rs192229205	97.2 (T/T)	2.8 (T/G)	0 (G/G)	88.6	11.0	0.4	100	0	0	100	0	0	99.7	0.3	0
Q17H rs137947890	96.7 (G/G)	3.2 (G/C)	0.1 (C/C)	87.6	12.1	0.3	99.7	0.3	0	100	0	0	100	0	0
Y24H rs140129240	99.8 (T/T)	0.2 (T/C)	0 (C/C)	100	0	0	100	0	0	100	0	0	98.8	1.2	0
R28K rs548664955	99.95 (G/G)	0.05 (G/A)	0 (A/A)	99.8	0.2	0	100	0	0	100	0	0	100	0	0
D52E rs552124276	99.95 (T/T)	0.05 (T/G)	0 (G/G)	99.8	0.2	0	100	0	0	100	0	0	100	0	0
K59Q rs534659594	99.95 (A/A)	0.05 (A/C)	0 (C/C)	100	0	0	99.8	0.2	0	100	0	0	100	0	0
R63Q rs535873217	99.95 (G/G)	0.05 (G/A)	0 (A/A)	100	0	0	100	0	0	100	0	0	99.8	0.2	0
L64P rs7187246	26.9 (T/T)	45.1 (T/C)	28.0 (C/C)	10.4	50.2	39.4	54.5	37.5	8.1	4.8	39.9	55.4	52.1	42.7	5.2
R105C rs13332099	98.8 (C/C)	1.2 (C/T)	0 (T/T)	95.8	4.2	0	99.4	0.6	0	100	0	0	100	0	0
R108H rs550591285	99.8 (G/G)	0.1 (G/A)	0.1 (A/A)	99.7	0.3	0	100	0	0	100	0	0	100	0	0
R108P rs550591285	99.9 (G/G)	0.1 (G/A)	0 (A/A)	99.7	0.3	0	100	0	0	100	0	0	100	0	0
R115Q rs142505782	99.95 (G/G)	0.05 (G/A)	0 (A/A)	99.8	0.2	0	100	0	0	100	0	0	100	0	0
L131V rs181910435	99.9 (C/C)	0.1 (C/G)	0 (G/G)	100	0	0	100	0	0	99.6	0.4	0	100	0	0
M151K rs536440408	99.9 (T/T)	0.1 (T/A)	0 (A/A)	100	0	0	100	0	0	99.8	0.2	0	100	0	0

[1]hmz (wt): Homozygote (wild-type); htz: Heterozygote; hmz (SNP): Homozygote (SNP); Rs: Variant identification.

SNP	Genotype Frequency (%) of all 2015 individuals			Genotype frequency (%) of the different population groups											
				African (AFR)			American (AMR)			East Asian (EAS)			European (EUR)		
	hmz (wt)[1]	htz[1]	hmz (snp)[1]	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)
G160V rs548812070	99.92 (G/G)	0.075 (G/A)	0 (A/A)	100	0	0	99.7	0.3	0	100	0	0	100	0	0
R178I rs185396407	99.92 (G/G)	0.075 (G/T)	0 (T/T)	100	0	0	99.7	0.3	0	100	0	0	100	0	0
E185K rs559312334	99.92 (G/G)	0.075 (G/A)	0 (A/A)	100	0	0	99.7	0.3	0	100	0	0	100	0	0
S202P rs200518351	99.9 (T/T)	0.1 (T/C)	0 (C/C)	99.8	0.2	0	100	0	0	99.8	0.2	0	100	0	0
A242T rs200202580	99.95 (G/G)	0.05 (G/A)	0 (A/A)	100	0	0	100	0	0	100	0	0	99.8	0.2	0
L270F rs147894332	99.8 (G/G)	0.2 (G/C)	0 (C/C)	100	0	0	100	0	0	99.8	0.2	0	99.8	0.2	0
L271G rs111621042	99.7 (T/T)	0.3 (T/G)	0 (G/G)	99.4	0.6	0	99.7	0.3	0	99.8	0.2	0	100	0	0
R316W rs367739019	99.95 (C/C)	0.05 (C/T)	0 (T/T)	100	0	0	100	0	0	100	0	0	99.8	0.2	0
V335I rs4643305	99.95 (G/G)	0.05 (G/A)	0 (A/A)	100	0	0	100	0	0	99.8	0.2	0	100	0	0
V335L rs4643305	99.95 (G/G)	0.05 (G/A)	0 (A/A)	100	0	0	100	0	0	99.8	0.2	0	100	0	0
R358P rs188648395	99.92 (G/G)	0.075 (G/C)	0 (C/C)	100	0	0	99.7	0.3	0	100	0	0	100	0	0
T366K rs150888398	99.95 (C/C)	0.05 (C/A)	0 (A/A)	100	0	0	100	0	0	100	0	0	99.8	0.2	0
T366M rs150888398	99.95 (C/C)	0.05 (C/A)	0 (A/A)	100	0	0	100	0	0	100	0	0	99.8	0.2	0
I394L rs200201528	99.95 (A/A)	0.05 (A/C)	0 (C/C)	100	0	0	100	0	0	99.8	0.2	0	100	0	0
V401I rs147314845	99.9 (G/G)	0.1 (A/G)	0 (A/A)	100	0	0	99.7	0.3	0	100	0	0	99.8	0.2	0
G405S rs570810199	99.95 (G/G)	0.05 (G/A)	0 (A/A)	100	0	0	100	0	0	99.8	0.2	0	100	0	0
G411R rs534821524	99.95 (G/G)	0.05 (G/C)	0 (C/C)	99.8	0.2	0	100	0	0	100	0	0	100	0	0
T433I rs556550170	99.95 (C/C)	0.05 (C/G)	0 (G/G)	100	0	0	100	0	0	99.8	0.2	0	100	0	0
A435P rs577931949	99.9 (G/G)	0.1 (G/C)	0 (C/C)	100	0	0	99.4	0.6	0	100	0	0	100	0	0
R438Q rs138191656	95.4 (G/G)	4.5 (G/A)	0.1 (A/A)	83.8	15.9	0.3	98.3	1.7	0	99.6	0.4	0	100	0	0
R447Q rs145697504	99.95 (G/G)	0.05 (G/A)	0 (A/A)	99.8	0.2	0	100	0	0	100	0	0	100	0	0
D451H rs572272646	99.2 (G/G)	0.075 (G/C)	0 (C/C)	100	0	0	99.7	0.3	0	100	0	0	100	0	0
A462S rs141811117	99.95 (G/G)	0.05 (G/T)	0 (T/T)	100	0	0	100	0	0	99.8	0.2	0	100	0	0
N463D rs146233716	77.8 (A/A)	9.7 (A/G)	12.5 (G/G)	92.7	5.9	1.4	82.4	9.5	8.1	75.6	6.9	17.5	71.6	7.8	20.7
I466V rs200695626	99.9 (A/A)	0.1 (A/G)	0 (G/G)	100	0	0	100	0	0	100	0	0	99.6	0.4	0
G470R rs531975924	99.9 (G/G)	0.1 (G/A)	0 (A/A)	100	0	0	100	0	0	99.6	0.4	0	100	0	0

[1]hmz (wt): Homozygote (wild-type); htz: Heterozygote; hmz (SNP): Homozygote (SNP); Rs: Variant identification.

SNP	Genotype Frequency (%) of all 2015 individuals			Genotype frequency (%) of the different population groups											
				African (AFR)			American (AMR)			East Asian (EAS)			European (EUR)		
	hmz (wt)[1]	htz[1]	hmz (snp)[1]	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)
T491M rs369633543	99.9 (C/C)	0.1 (C/T)	0 (T/T)	100	0	0	100	0	0	99.6	0.4	0	100	0	0
P499A rs543127018	99.95 (C/C)	0.05 (C/G)	0 (G/G)	99.8	0.2	0	100	0	0	100	0	0	100	0	0
V505L rs184553350	99.95 (G/G)	0.05 (G/C)	0 (C/C)	100	0	0	100	0	0	99.8	0.2	0	100	0	0
S513L rs1133607	74.6 (C/C)	22.7 (C/T)	2.7 (T/T)	75.5	23	1.5	75.8	21	3.2	65.5	29.4	5.2	85.3	14.5	0.2
K552R rs181555433	99.2 (A/A)	0.075 (A/G)	0 (G/G)	100	0	0	99.7	0.3	0	100	0	0	100	0	0

[1]hmz (wt): Homozygote (wild-type); htz: Heterozygote; hmz (SNP): Homozygote (SNP); Rs: Reference SNP cluster identification number. The 1000 Genomes Project Phase 3 data sequenced the complete genome of 2015 individuals of African, Admixed American, East Asian and European descent (Table 4.1). In this study, the data generated by the 1000 Genomes Project was used to analyse the genotype frequencies of the 47 non-synonymous SNPs identified in the human *ACSM2A* gene (Table 4.2).

From the 47 non-synonymous SNPs identified, the L64P polymorphism had the highest homozygous genotype frequency of 28.0%, followed by N463D (12.5%), R5Q (4.8%), S513L (2.7%), Q17H (0.1%), R108H (0.1%), R108P (0.1%), and, R438Q (0.1%). The remaining 39 non-synonymous SNPs were only detected as heterozygotes. Genotype frequency assessment within the different population groups (African, Admixed American, East Asian and European) in Table 4.2 revealed that most of the variants (G9R, Y24H, R28K, D52E, K59Q, R63Q, R108H, R108P, R115Q, L131V, M151K, G160V, R178I, E185K, A242T, R316W, V335I, V335L, R358P, T366K, T366M, I394L, G405S, G411R, T433I, A435P, R447Q, D451H, A462S, I466V, G470R, T491M, P499A, V505L, K552R) were only found in one of the four population groups. L10R (AFR; EUR), Q17H (AFR; AMR), R105C (AFR; AMR), S202P (AFR; EAS), L270F (EAS; EUR), and V401I (AMR; EUR) polymorphisms were found in two of the population groups. The L271G (AFR; AMR; EAS) and R438Q (AFR; AMR; EAS) polymorphisms were identified in three population groups, while the R5Q, L64P, and S513L polymorphism were found in all four population groups.

No information at this time is available suggesting that genetic variation in the acid: CoA ligase enzymes influences the metabolism of salicylate and also other benzoate derivatives.

Table 4.3: Non-synonymous SNPs and genotype frequencies identified in the 1000 Genomes sequencing data of 2015 individuals within the coding region of *ACSM2B*.

SNP	Genotype Frequency (%) of all 2015 individuals			Genotype frequency (%) of the different population groups											
				African (AFR)			American (AMR)			East Asian (EAS)			European (EUR)		
	hmz (wt)[1]	htz[1]	hmz (snp)[1]	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)
R5Q rs564472730	99.9 (G/G)	0.1 (G/A)	0 (AA)	100	0	0	99.7	0.3	0	100	0	0	99.8	0.2	0
R21H rs141648973	97.9 (G/G)	2.1 (G/A)	0 (A/A)	97.7	2.3	0	99.7	0.3	0	99.2	0.8	0	100	0	0
K77R rs563151418	99.95 (A/A)	0.05 (A/G)	0 (G/G)	99.8	0.2	0	100	0	0	100	0	0	100	0	0
S98A rs553128923	99.8 (T/T)	0.2 (T/G)	0 (G/G)	100	0	0	100	0	0	99.2	0.8	0	100	0	0
R105H rs577119435	99.95 (G/G)	0.05 (G/A)	0 (A/A)	9.8	0.2	0	100	0	0	100	0	0	100	0	0
D107G rs559044619	99.95 (A/A)	0.05 (A/G)	0 (G/G)	100	0	0	100	0	0	99.8	0.2	0	100	0	0
M112V rs374648082	99.95 (A/A)	0.05 (A/G)	0 (G/G)	99.8	0.2	0	100	0	0	100	0	0	100	0	0
P114L rs554694059	99.95 (C/C)	0.05 (C/T)	0 (T/T)	100	0	0	100	0	0	99.8	0.2	0	100	0	0
W120R rs566287967	99.95 (T/T)	0.05 (T/A)	0 (A/A)	100	0	0	100	0	0	99.8	0.2	0	100	0	0
R128Q rs538802177	99.95 (G/G)	0.05 (G/A)	0 (A/A)	100	0	0	100	0	0	99.8	0.2	0	100	0	0
Q252K rs540152177	99.8 (C/C)	0.1 (C/A)	0 (A/A)	100	0	0	99.7	0.3	0	99.8	0.2	0	100	0	0
N268T rs373399717	99.95 (A/A)	0.05 (A/C)	0 (C/C)	99.8	0.2	0	100	0	0	100	0	0	100	0	0
I269V rs142151089	99.95 (A/A)	0.05 (A/G)	0 (G/G)	100	0	0	100	0	0	100	0	0	99.8	0.2	0
T278A rs77863699	73.6 (A/A)	22.6 (A/G)	3.8 (G/G)	45.2	43.6	11.2	85.0	14.7	0.3	71	25.8	3.2	95.4	4.4	0.2
K298R rs557429764	99.95 (A/A)	0.05 (A/G)	0 (G/G)	100	0	0	100	0	0	100	0	0	99.8	0.2	0
I305V rs80135299	93.3 (A/A)	6.1 (A/G)	0.6 (G/G)	74.6	21.3	2.3	97.1	2.9	0	100	0	0	100	0	0
D322N rs138313532	99.4 (G/G)	0.6 (G/A)	0 (A/A)	100	0	0	98.6	1.4	0	100	0	0	98.2	1.8	0
P329S rs183078161	99.9 (C/C)	0.1 (C/T)	0 (T/T)	100	0	0	100	0	0	100	0	0	99.4	0.6	0
G337V rs539831165	99.95 (G/G)	0.05 (G/T)	0 (T/T)	100	0	0	100	0	0	99.8	0.2	0	100	0	0
E339K rs58301506	99.4 (G/G)	0.5 (G/T) 0.1 (G/A)	0 (A/A)	98.0	2.0 (G/T)	0	99.4	0.6 (G/A)	0	100	0	0	100	0	0
E339A rs142106780	99.9 (G/G)	0.1 (G/A)	0 (A/A)	100	0	0	99.7	0.3	0	100	0	0	99.8	0.2	0
E347Q rs562296082	99.95 (G/A)	0.05 (G/C)	0 (C/C)	100	0	0	100	0	0	100	0	0	99.8	0.2	0

[1]hmz (wt): Homozygote (wild-type); htz: Heterozygote; hmz (SNP): Homozygote (SNP); Rs: Variant identification.

SNP	Genotype Frequency (%) of all 2015 individuals			Genotype frequency (%) of the different population groups											
				African (AFR)			American (AMR)			East Asian (EAS)			European (EUR)		
	hmz (wt)[1]	htz[1]	hmz (snp)[1]	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)
L355P rs200390240	99.9 (T/T)	0.1 (T/C)	0 (C/C)	100	0	0	100	0	0	100	0	0	99.8	0.2	0
R358Q rs374731787	99.9 (G/G)	0.1 (G/A)	0 (A/A)	99.1	0.6	0	100	0	0	100	0	0	100	0	0
T385M rs146865904	99.9 (C/C)	0.1 (C/T)	0 (T/T)	99.6	0.4	0	100	0	0	100	0	0	100	0	0
F360S rs202090220	99.9 (T/T)	0.1 (T/C)	0 (C/C)	100	0	0	100	0	0	99.6	0.4	0	99.8	0.2	0
Q393E rs185232542	99.95 (C/C)	0.05 (C/G)	0 (G/C)	99.8	0.2	0	100	0	0	100	0	0	100	0	0
P404S rs139892532	99.8 (C/C)	0.2 (C/T)	0 (T/T)	99.2	0.8	0	100	0	0	100	0	0	100	0	0
G405S rs146499503	99.95 (G/G)	0.05 (G/A)	0 (A/A)	99.8	0.2	0	100	0	0	100	0	0	100	0	0
G411D rs142553905	99.8 (G/G)	0.2 (G/A)	0 (A/A)	100	0	0	99.7	0.3	0	100	0	0	99.2	0.8	0
G421D rs138313327	99.8 (G/G)	0.2 (G/A)	0 (A/A)	99.4	0.6	0	100	0	0	99.8	0.2	0	100	0	0
G439R rs143364212	99.9 (G/G)	0.1 (G/A)	0 (A/A)	100	0	0	99.7	0.3	0	100	0	0	99.6	0.4	0
R447W rs373308064	99.9 (C/C)	0.1 (C/T)	0 (T/T)	99.8	0.2	0	100	0	0	100	0	0	100	0	0
R472W rs140049721	99.9 (C/C)	0.1 (C/T)	0 (T/T)	99.7	0.3	0	100	0	0	100	0	0	99.8	0.2	0
T491M rs370065320	99.9 (C/C)	0.1 (C/G)	0 (C/C)	99.8	0.2	0	100	0	0	99.8	0.2	0	100	0	0
V504A rs200676297	99.9 (T/T)	0.1 (T/C)	0 (C/C)	100	0	0	100	0	0	99.6	0.4	0	100	0	0
K506R rs199930809	99.9 (A/A)	0.1 (A/G)	0 (G/G)	100	0	0	100	0	0	99.8	0.2	0	100	0	0
V531L rs201732013	99.95 (G/G)	0.05 (G/T)	0 (T/T)	100	0	0	100	0	0	99.8	0.2	0	100	0	0
P537L rs150010732	99.2 (C/C)	0.8 (C/T)	0 (T/T)	100	0	0	98.6	1.4	0	97.8	2.2	0	100	0	0
Y540N rs139253186	99.8 (T/T)	0.2 (T/A)	0 (A/A)	100	0	0	100	0	0	100	0	0	98.8	1.2	0
K543N rs200472857	99.8 (G/G)	0.2 (G/C)	0 (C/C)	100	0	0	100	0	0	99.4	0.6	0	100	0	0
P551R rs201364830	99.9 (C/C)	0.1 (C/G)	0 (G/G)	100	0	0	100	0	0	99.4	0.6	0	100	0	0
G572R rs562863484	99.9 (G/G)	0.1 (G/A)	0 (A/A)	100	0	0	100	0	0	99.6	0.4	0	100	0	0

[1]hmz (wt): Homozygote (wild-type); htz: Heterozygote; hmz (SNP): Homozygote (SNP). The 1000 Genomes Project Phase 3 data sequenced the complete genome of 2015 individuals of African, Admixed American, East Asian and European descent (Table 4.1). In this study, the data generated by the 1000 Genomes Project was used to analyse the genotype frequencies of the 43 non-synonymous SNPs identified in the human *ACSM2B* gene (Table 4.3).

The non-synonymous SNPs and genotype frequencies identified in the *ACSM2B* gene by the 1000 Genomes data are presented in Table 4.3. From the 43 non-synonymous SNPs identified, the T278A polymorphism had the highest homozygous genotype frequency of 3.8%, followed by I305V (0.6%). The remaining 41 non-synonymous SNPs were only detected as heterozygotes. Genotype frequency assessment within the different population groups (African, Admixed American, East Asian and European) in Table 4.3 revealed that most of the variants (K77R, S98A, R105H, D107G, M112V, P114L; W120R, R128Q, N268T, I269V, K298R, P329S, G337V, E347Q, L355P, R358Q, T385M, Q393E, P404S, G405S, R447W, V504A, K506R, V531L, Y540N, K543N, P551R, G572R) were only found in one of the four population groups. R5Q (AMR; EUR), Q252K (AMR; EAS), I305V (AFR; AMR), D322N (AMR; EUR), E339K (AFR; AMR), E339K (AFR; AMR), E339A (AFR; AMR), F360S (EAS; EUR), G411D (AMR, EUR), G421D (AFR; EAS), G439R (AMR; EUR), R472W (AMR; EUR), T491M (AMR; EAS), and P537L (AMR; EAS) polymorphisms were found in two of the population groups. The R21H (AFR; AMR; EAS) polymorphism was identified in three population groups, while the T278A polymorphism was found in all four population groups.

No previous studies have linked any association between non-synonymous SNPs identified in the *ACSM2B* gene and the metabolism of salicylate and also other benzoate derivatives,

4.2.2 The NHLBI Exome Sequencing Project data

The data generated by the NHLBI ESP provided the next two population groups available in this study. The data received from the NHLBI ESP was used to analyse the allele frequencies of the 15 (*ACSM2A*), and 15 (*ACSM2B*) non-synonymous SNPs identified in the AA and EA individuals (Table 4.4; Table 4.5).

Table 4.4: Non-synonymous SNPs and allele frequencies identified within the coding region of *ACSM2A*, genotyped by the NHLBI ESP in 6,503 individuals.

Non-synonymous SNPs		Population allele frequencies (%)	
		AA	EA
R5Q rs59292608	WT (G)	96.9	85.2
	SNP (A)	3.1	14.8
V7I rs139231029	WT (G)	100	99.9
	SNP (A)	0	0.1
Q17H rs137947890	WT (G)	95.0	99.9
	SNP (C)	5.0	0.1
Y24H rs140129240	WT (T)	100	99.9
	SNP (C)	0	0.1
A60V rs142474503	WT (C)	100	99.0
	SNP (T)	0	0.1
L64P rs7187246	WT (T)	40.6	71.9
	SNP (C)	59.4	28.1
E78K rs34655000	WT (G)	99.5	97.5
	SNP (A)	0.5	2.5
R115Q rs142505782	WT (G)	99.9	100
	SNP (A)	0.1	0
S222T rs142969595	WT (G)	100	99.9
	SNP (C)	0	0.1
C271G rs111621042	WT (T)	99.7	100
	SNP (G)	0.3	0
I305V rs144701679	WT (A)	94.0	100
	SNP (G)	6.0	0
T366K rs150888398	WT (C)	100	99.8
	SNP (A)	0	0.2
T366M rs150888398	WT (C)	100	99.8
	SNP (A)	0	0.2
R438Q rs138191656	WT (G)	94.9	100
	SNP (A)	5.1	0
S513L rs1133607	WT (C)	87.5	89.9
	SNP (T)	12.5	10.1

WT: Wild-type; SNP: Single nucleotide polymorphism; AA: African American; EA: European American.

The allele frequencies of the 15 non-synonymous SNPs identified in the coding region of the *ACSM2A* gene in 6,503 individuals are summarised in Table 4.4. Allele frequency investigations within the AA and EA population groups of Table 4.4 revealed that in all population groups (AA and EA), the L64P variant had the highest allele frequency of 59.4% in AA and 28.1% in EA, the R5Q polymorphism had the second highest allele frequency of 14.8% in EA and 3.1% in AA, followed by S513L (12.5% in AA and 10.1% in EA), I305V (6.0% in AA), R438Q (5.1% in AA), Q17H (5.0% in AA and 0.1% in EA), E78K (2.5% in EA and 0.5% in AA), C271G (0.3% in AA),T366K (0.2% in EA), T366M (0.2% in EA), V7I (0.1% in EA), Y24H (0.1% in EA), A60V (0.1% in EA), R115Q (0.1% in AA), S222T (0.1% in EA).

Table 4.5: Non-synonymous SNPs and allele frequencies identified within the coding region of *ACSM2B*, genotyped by the NHLBI ESP in 6,503 individuals.

Non-synonymous SNPs		Population allele frequencies (%)	
		AA	EA
R21H rs141648973	WT (C)	98.9	100
	SNP (T)	1.1	0
R108P rs143779927	WT (G)	99.9	100
	SNP (A)	0.1	0
T278A rs77863699	WT (T)	71.8	97.6
	SNP (C)	28.2	2.4
I305V rs80135299	WT (T)	90.3	99.9
	SNP (C)	9.7	0.1
D322N rs138313532	WT (C)	99.8	98.9
	SNP (T)	0.2	1,1
E339K rs58301506	WT (G)	99.5	100
	SNP (A/T)	0.5	0
G362R rs143459487	WT (C)	100	99.8
	SNP (G)	0	0.2
T385M rs146865904	WT (G)	99.9	99.9
	SNP (A)	0.1	0.1
P404S rs139892532	WT (G)	99.6	100
	SNP (A)	0.4	0
G411D rs142553905	WT (C)	100	99.8
	SNP (T)	0	0.2
G421D rs138313327	WT (C)	99.9	100
	SNP (T)	0.1	0
N436Y rs148478247	WT (T)	99.9	100
	SNP (A)	0.1	0
W442R rs139233971	WT (A)	100	99.8
	SNP (G)	0	0.2
R472W rs140049721	WT (G)	99.8	100
	SNP (A)	0.2	0
Y540N rs139253186	WT (A)	99.9	99.4
	SNP (T)	0.1	0.6

WT: Wild-type; SNP: Single nucleotide polymorphism; AA: African American; EA: European American.

The allele frequencies of the 15 non-synonymous SNPs identified in the coding region of the *ACSM2B* gene in 6,503 individuals are summarised in Table 4.5. Allele frequency investigations within the AA and EA population groups of Table 4.5 revealed that in all

population groups (AA and EA), the T278A polymorphism had the highest allele frequency of 28.2% (AA) and 2.4% (EA). The I305V variant had the second highest allele frequency of 9.7% (AA) and 0.1% (EA), followed by D322N (1.1% in EA and 0.2% in AA), R21H of 1.1% (AA), Y540N (0.6% in EA and 0.1% in AA), E339K (0.5% in AA), P404S (0.4% in AA), G362R, (0.2% in EA), W442R (0.2% in AA), R472W (0.2% in AA). G411D (0.2% in EA), G421D (0.1% in AA), N436Y (0.1% in AA) R108P (0.1% in AA), T385M (0.1% in AA and EA).

4.2.3 Exome sequencing data of the SA population

The exomes of 19 South African Afrikaner Caucasian individuals for the present study was sequenced. This was done using an Ion Proton NGS approach at the CAF unit in Stellenbosch, as described in Section 2.10. From the sequencing run, and after comparison against reference sequences, a total of 25 variants were identified within the coding region of *ACSM2A* and *ACSM2B* in 19 individuals.

Section 3.6 gives a detailed description of the bioinformatics workflow used in the present study to analyse the substantial amount of data acquired from sequencing runs. By means of primary and secondary data mining processes, single nucleotide variants were filtered from the data to ultimately present non-synonymous variants identified within the coding region of the *ACSM2A* and *ACSM2B* genes, important to this study (Table 4.6, 4.7)

4.2.3.1 South African Afrikaner Caucasian population

The secondary part of the bioinformatics workflow (Figure 3.1) consisted of an in-house developed pipeline for data mining and the identification of possible novel and/or previously reported variants by means of available open source databases.

To investigate whether genetic variation in the ORF of the acyl CoA synthetase genes of the South African Afrikaner Caucasian population are low, a small cohort of 19 individuals were chosen and the exomes were sequenced. These results are indicated in Table 4.6 and 4.7.

Table 4.6: Summary of the polymorphisms identified in the nucleotide sequence of the *ACSM2A* gene in 19 South African Afrikaner Caucasian individuals.

Polymorphism (rs)	Variation	Location (gene)	Exon/Intron Location	Effect (protein)	Allele frequency		Genotype frequency		
					Allele	%	Genotype	n[1]	%
rs59292608	Missense	g.20471450 G>A	Exon 2	R5Q p.Arg5Gln	G A	84.2 15.8	GG GA AA	13 6 0	68.4 31.6 0
rs7187246	Missense	g.20476852 T>C	Exon 3	L64P p.Leu64Pro	T C	71.1 28.9	TT TC CC	9 9 1	47.4 47.4 5.3
rs34655000	Missense	g.20476893 G>A	Exon 3	E78K p.Glu78Lys	G A	97.4 2.6	GG GA AA	18 1 0	94.7 5.3 0
rs1133607	Missense	g.20494408 C>T	Exon 13	S513L p.Ser513Leu	C T	89.5 10.5	CC CT TT	15 4 0	78.9 21.1 0
rs1700805	Synonymous	g.20488696 A>G	Exon 9	No effect	A G	76.3 23.7	AA AG GG	11 7 1	57.9 36.8 5.3
rs1634306	Intron variant	g.20488803 T>A	Intron 9	No effect	T A	89.5 10.5	TT TA AA	17 0 2	89.5 0 10.5
rs4783532	Intron variant	g.20497775 G>C	Intron 13	No effect	G C	86.8 13.2	GG GC CC	15 3 1	78.9 15.8 5.3
rs28615302	Intron variant	g.20480818 T>C	Intron 3	No effect	T C	94.7 5.3	TT TC CC	17 2 0	89.5 10.5 0
rs1827711	Intron variant	g.20482368 A>G	Intron 4	No effect	A G	94.7 5.3	AA AG GG	17 2 0	89.5 10.5 0
rs1629314	Intron variant	g.20487125 T>G	Intron 8	No effect	T G	94.7 5.3	TT TG GG	17 2 0	89.5 10.5 0
rs1634305	Intron variant	g.20486913 G>T	Intron 6	No effect	G T	97.4 2.6	GG GT TT	18 1 0	94.7 5.3 0
rs2072080	3'UTR	g.20498079 T>G	Exon 14	No effect	T G	81.6 18.4	TT TG GG	15 1 3	78.9 5.3 15.8
N/A	Novel variation	g.20494512 A>G	Intron 13	No effect	A G	97.4 2.6	AA AG GG	18 1 0	94.7 5.3 0
N/A	Novel variation	g.20489868 A>C	Intron 9	No effect	A C	97.4 2.6	AA AC CC	18 1 0	94.7 5.3 0

[1]Number of individuals; N/A: Not available. A summary of the polymorphisms identified in the nucleotide sequence of the *ACSM2A* gene in 19 South African Afrikaner Caucasian individuals are indicated in Table 4.6.

Two novel and twelve known polymorphisms of the *ACSM2A* gene were identified. Of the twelve known polymorphisms, four (R5Q, L64P, E78K, and S513L) were non-synonymous. Only one (rs1700805) was synonymous, and six (rs1634306, rs4783532, rs28615302, rs1827711, rs1634305, rs1629314) were located in the intron regions (Table 4.6). Of the non-synonymous SNPs, the L64P variant was homozygous in one individual and heterozygous in 9 individuals resulting in the non-synonymous SNP with the highest allele frequency (28.9%). This is followed by the R5Q variant, heterozygous in 6 individuals (15.8%), the S513L variant, heterozygous in 4 individuals (10.5%), and the E78K variant, heterozygous in one individual (2.6%) (Table 4.6). The two novel variants identified in the 19 South African Afrikaner Caucasian individuals were both located in the intron regions and not of any importance to the present study.

Table 4.7: Summary of the polymorphisms identified in the nucleotide sequence of the *ACSM2B* gene in 19 South African Afrikaner Caucasian individuals.

Polymorphism (rs)	Variation	Location (gene)	Exon/Intron Location	Effect (protein)	Allele frequency		Genotype frequency		
					Allele	%	Genotype	n[1]	%
rs141648973	Missense	g.20576106 G>A	Exon 3	R21H p.Arg21His	G A	97.4 2.6	GG GA AA	18 1 0	94.7 5.3 0
rs16970280	Synonymous	g.20552075 C>A	Exon 14	No effect	C A	13.2 86.8	CC CA AA	1 3 15	5.3 15.8 78.9
rs112352583	Synonymous	g.20576048 G>T	Exon 3	No effect	G T	76.3 23.7	GG GT TT	11 7 1	75.9 36.8 5.3
rs148194016	Synonymous	g.20554317 G>A	Exon 13	No effect	G A	97.4 2.6	GG GA AA	18 1 0	94.7 5.3 0
rs62035054	Intron variant	g.20554613 C>T	Intron 11	No effect	C T	94.7 5.3	CC CT TT	18 0 1	94.7 0 5.3
rs9929350	Intron variant	g.20557533 A>G	Intron 10	No effect	A G	97.4 2.6	AA AG GG	18 1 0	94.7 5.3 0
rs12927522	Intron variant	g.20557688 A>G	Intron 10	No effect	A G	89.5 10.5	AA AG GG	16 2 1	84.2 10.5 5.3
rs7198108	Intron variant	g.20557935 G>A	Intron 9	No effect	G A	97.4 2.6	GG GA AA	18 1 0	94.7 5.3 0
N/A	Novel variant	g.20552166 T>G	Intron 13	No effect	T G	97.4 2.6	TT TG GG	18 1 0	94.7 5.3 0
N/A	Novel variant	g.20554204 A>C	Intron 11	No effect	A C	97.4 2.6	AA AC CC	18 1 0	94.7 5.3 0
N/A	Novel variant	g.20556642 T>C	Intron 10	No effect	T C	97.4 2.6	TT TC CC	18 1 0	94.7 5.3 0

[1]Number of individuals. A summary of the polymorphisms identified in the nucleotide sequence of the *ACSM2B* gene in 19 South African Afrikaner Caucasian individuals are indicated in Table 4.7.

Three novel and eight known polymorphisms of the *ACSM2B* gene were identified. Of the eight known polymorphisms, only one (R21H) was non-synonymous, three (rs16970280, rs112352583, rs148194016) were synonymous and four (rs62035054, rs9929350, rs12927522, rs7198108) were located in the intron regions (Table 4.7). Only one non-synonymous SNP was identified within the coding region of the *ACSM2B* gene of the 19 South African Afrikaner Caucasian individuals. The R21H variant was heterozygous in only one individual with an allele frequency of 2.6% (Table 4.7). The three novel variations identified in the *ACSM2B* gene are all located in the intron regions and are therefore not relevant to the present study.

Table 4.8 represents a summary of the non-synonymous SNPs and homozygous SNP genotype frequencies (%) identified within the coding region of the *ACSM2A* gene in the 1000 Genomes Project, NHLBI ESP and the Present study. If the specific non-synonymous SNP was not identified within a population group, it is indicated by a grey background in Table 4.8.

Nine non-synonymous SNPs (R5Q, Q17H, L64P, E78K, R108H, R108P, N463D, R438Q, and S513L) represent the variants with the highest homozygous SNP frequency identified in the AFR, AMR EAS, EUR, AA, EA, and SA population groups.

The L64P variant had the highest homozygous SNP genotype frequency of 29.0%, identified in all population groups (AFR, AMR EAS, EUR, AA, EA, SA), followed by the N463D variant of 12.5% identified in the AFR, AMR, EAS, and EUR population. The R5Q variant had the third highest genotype frequency of 5.5% (AFR, AMR EAS, EUR, AA, EA, SA), followed by the S513L variant with a genotype frequency of 1.6% (AFR, AMR EAS, EUR, AA, EA, SA), the E78K variant of 0.5% (AA, EA, SA), Q17H of 0.1% (AFR, AMR EAS, EUR, AA, EA), followed by R438Q, R108H and R108P with a frequency of 0.1%, identified in the AFR, AMR EAS, and EUR population.

Table 4.9 represents a summary of the non-synonymous SNPs and highest identified homozygous SNP genotype frequencies (%) identified within the coding region of the *ACSM2B* gene in the 1000 Genomes Project, NHLBI ESP and the Present study. If the specific non-synonymous SNP was not identified within a population group, it is indicated by a grey background in Table 4.9. Three non-synonymous SNPs (T278A, I305V, D322N) represent the variants with the highest homozygous SNP frequency identified in the AFR, AMR EAS, EUR, AA, EA, and SA population groups. The T278A variant had the highest homozygous SNP genotype frequency of 4.0%, followed by the I305V variant of 0.7%, and the D322N variant of 0.1%, identified in the AFR, AMR EAS, EUR, AA, and EA population.

Table 4.8: Summary of the non-synonymous SNPs and genotype frequencies (%) identified within the coding region of the *ACSM2A* gene in the 1000 Genomes Project, NHLBI ESP and the present study.

SNP	All			Genotype frequency (%) of the different population groups																				
				1000 Genomes Data												NHLBI Exome Sequencing Project						Present Study		
				African (AFR)			American (AMR)			East Asian (EAS)			European (EUR)			African American (AA)			European American (EA)			South African Caucasian Afrikaner		
	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)
L64P rs7187246	34.0 T/T	37.0 T/C	29.0 C/C	10.4	50.2	39.4	54.5	37.5	8.1	4.8	39.9	55.4	52.1	42.7	5.2	16.4	48.4	35.1	51.3	41.2	7.5	47.4	47.4	5.3
N463D rs146233716	77.8 A/A	9.7 A/G	12.5 G/G	92.7	5.9	1.4	82.4	9.5	8.1	75.6	6.9	17.5	71.6	7.8	20.7									
R5Q rs59292608	72.0 G/G	22.5 G/A	5.5 A/A	97.6	2.4	0	68.7	12.4	0.9	34.5	50.6	14.9	68.0	29.6	2.4	93.7	6.3	0	72.8	24.9	2.3	68.4	31.6	0
S513L rs1133607	77.0 C/C	21.4 C/T	1.6 T/T	75.5	23.0	1.5	75.8	21.0	3.2	65.5	29.4	5.2	85.3	14.5	0.2	76.5	22.0	1.5	80.4	18.8	0.7	78.9	21.1	0
E78K rs34655000	95.0 G/G	4.5 G/A	0.5 A/A													94	5.0	1	96.3	3.4	0.3	94.7	5.3	0
Q17H rs137947890	96.3 G/G	3.6 G/C	0.1 C/C	87.6	12.1	0.3	99.7	0.3	0	100	0	0	100	0	0	90.4	9.4	0.2	99.9	0.1	0			
R438Q rs138191656	95.4 G/G	4.5 G/A	0.1 A/A	83.8	15.9	0.3	98.3	1.7	0	99.6	0.4	0	100	0	0									
R108H rs550591285	99.8 G/G	0.1 G/A	0.1 A/A	99.7	0.3	0	100	0	0	100	0	0	100	0	0									
R108P rs550591285	99.8 G/G	0.1 G/A	0.1 A/A	99.7	0.3	0	100	0	0	100	0	0	100	0	0									

[1]hmz (wt): Homozygote (wild-type); htz: Heterozygote; hmz (SNP): Homozygote (SNP). If the specific non-synonymous SNP was not identified within a population group, it is indicated by a grey background.

Table 4.9: Summary of the non-synonymous SNPs and genotype frequencies (%) identified within the coding region of the *ACSM2B* gene in the 1000 Genomes Project, NHLBI ESP and the Present study.

SNP	All			Genotype frequency (%) of the different population groups																				
				1000 Genomes Data												NHLBI Exome Sequencing Project						Present Study		
				African (AFR)			American (AMR)			East Asian (EAS)			European (EUR)			African American (AA)			European American (EA)			South African Caucasian Afrikaner		
	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)	hmz (wt)	htz	hmz (snp)
T278A rs77863699	74 A/A	22 A/G	4.0 G/G	45.2	43.6	11.2	85.0	14.7	0.3	71	25.8	3.2	95.4	4.4	0.2	52.0	40.0	8.0	95.4	4.4	0.2			
I305V rs80135299	92.3 A/A	7.0 A/G	0.7 G/G	74.6	21.3	2.3	97.1	2.9	0	100	0	0	100	0	0	82.2	15.8	2.0	99.9	0.1	0			
D322N rs138313532	99.1 G/G	0.8 G/A	0.1 A/A	100	0	0	98.6	1.4	0	100	0	0	98.2	1.8	0	99.5	0.4	0.1	97.8	2.1	0.1			

[1]hmz (wt): Homozygote (wild-type); htz: Heterozygote; hmz (SNP): Homozygote (SNP). If the specific non-synonymous SNP was not identified within a population group, it is indicated by a grey background.

Noticeable from Table 4.8 as well as Table 4.9 is that the genotype frequencies for the non-synonymous SNPs identified within the coding region of the *ACSM2A* and *ACSM2B* genes of the AFR, AMR EAS, EUR, AA, EA and SA population vary a great deal between the different population groups. However, it is evident from the genotype frequencies that the AFR and AA population has similar genotype frequencies, along with the AMR and EA population. The EUR population has similar genotype frequencies to those identified for the SA population, as can be expected. It is clear that the AFR and SA population vary significantly in genotype frequency and non-synonymous SNPs identified within the specific populations. This may be due to the fact that the SA population used in the present study is not African in recent descent, but rather European. It is for this reason that the AFR and SA populations vary in such a way.

From Table 4.9 it is evident that very low genotype frequencies exist for the SNPs identified within the coding region of the *ACSM2B* gene (T278A: 4.0%, I305V: 0.7%, D322N: 0.1%) compared to genotype frequencies identified for *GLYAT* (N156S: 90%, S17T: 4.6%; R131H: 0.1%); from a study conducted by Van der Sluis *et al*., (2015).

Opposed to the three non-synonymous SNP identified with the highest homozygous SNP genotype frequency for the *ACSM2B* gene (Table 4.9), nine non-synonymous SNP were identified for the *ACSM2A* gene consisting of the highest homozygous SNP genotype frequency. This might indicate that the *ACSM2B* gene is more conserved than the *ACSM2A* gene.

From the allele frequencies analysed in this study it is clear that both the ACSM2A and *ACSM2B* genes are highly conserved with SNPs occurring at low frequencies and mostly as heterozygotes. This study supports the reference sequences chosen by Watkins *et al*., (2007) as no allele were found at high enough frequencies to indicate that it might be the wild-type allele.

CHAPTER 5: CONCLUSIONS AND FUTURE PROSPECTS

5.1 Introduction

Over the last few years, research on amino acid conjugation and specifically the glycine conjugation pathway has seriously been neglected. As the development of man-made chemicals and xenobiotics increases, research on the detoxification of these compounds needs to continue. The consumption of benzoic acid in humans has increased considerably over the last few years, lack of mechanisms capable of biotransformation and/or metabolism of toxic compounds like benzoic acid can lead to the alteration of a variety of physiological processes and serious adverse health consequences (Piper, 1999; Nair, 2001; Tfouni and Toledo, 2002; Lees *et al*., 2013; Van der Sluis *et al*., 2015). Very little information is available on the ligase enzyme responsible for the CoA activation step in the glycine conjugation pathway. Even though the *ACSM2B* gene is proposed to be involved in the metabolism of aspirin, no association studies have been done investigating the effect of non-synonymous SNPs on aspirin metabolism (Agundez *et al*., 2009).

Chapter 1 of this study provided a basic introduction to biotransformation while Chapter 2 presented a literature background concerning aspects of acyl CoA metabolism, toxicity and its central role in the glycine conjugation pathway, together with a thorough investigation on the acyl CoA synthetase family and nomenclature. A focus on the acyl CoA synthetase medium chain family members, *ACSM2A* and *ACSM2B*, forming the basis and motivation behind the study were also included in Chapter 2. In addition, the problem statement, aims and objectives were also presented in Chapter 2, while the materials and methods used to perform this study were provided in Chapter 3. The final results obtained were given and discussed in Chapter 4.

In Chapter 5, the approach, methodology used, along with the results, will be summarised to come to substantiated conclusions, outcomes and suggestions for future studies.

5.2 Summary

5.2.1 Problem statement, aim and objectives

In an effort to further investigate the importance of the glycine conjugation pathway in humans, the focal point of this study centred on the acyl CoA synthetase genes. As recent studies by Badenhorst *et al*., (2013), Badenhorst *et al*., (2014) and Van der Sluis *et al*., (2015) particularly focused on the important role of GLYAT in the glycine conjugation pathway, this study subsequently focused specifically on the *ACSM2B* gene encoding HXM-A, responsible for the activation step in the glycine conjugation pathway of benzoic acid. The *ACSM2A* gene was also investigated because very little information is available on the *ACSM2A* gene which results in considerable confusion in current literature regarding the distinction of these two highly similar genes. Studies by Van der Sluis *et al*., (2015) stated that the ORF of the *GLYAT* gene is highly conserved among humans, in an effort to identify if the *ACSM2B* ORF is also as highly conserved as *GLYAT*, the present study was formulated. Figure 2.7 illustrated the schematic strategy devised to address the aim of this study which was to identify genetic variation in the acyl CoA synthetase genes (*ACSM2A* and *ACMS2B*) in not only the present study (South African Afrikaner Caucasian Population), but also the 1000 Genomes Project (AFR, AMR EAS, EUR), and the NHLBI ESP (AA, EA).

The following objectives were performed to achieve the aim of this study:

1) Analysis of the genetic variation data of the human acyl CoA synthetase medium-chain family member 2A (*ACSM2A*) and the acyl CoA synthetase medium-chain family member 2B (*ACSM2B*) ORF available on public databases.
2) Generation of DNA sequence data of the coding region of a small cohort of South African Afrikaner Caucasian individuals using the Ion Proton NGS system for exome sequencing.
3) Identification of previously reported and potential novel SNPs by using a bioinformatics workflow.
4) Comparison of all population data available in this study against the South African Afrikaner Caucasian population.

5.2.2 Lack in basic understanding of the importance of the glycine conjugation pathway

A general lack in basic understanding exists regarding the significance of the glycine conjugation pathway in metabolism. The functional importance of the glycine conjugation pathway has particularly been underestimated by researchers in the past, and only recently has interest picked up on this subject (Badenhorst *et al*., 2014; Van der Sluis *et al*., 2015).

The importance of the glycine conjugation pathway was thoroughly discussed throughout this study, especially in Section 2.6 regarding all factors acquired for the efficient functioning of the glycine conjugation system. Studies by Badenhorst *et al*., (2013; 2014) and Van der Sluis *et al*., (2015) contributed to this topic, but the focus was mainly on the GLYAT enzyme.

As discussed in Section 2.7, several fatty acids and xenobiotic carboxylic acids that are conjugated to amino acids must first be activated to acyl CoAs by ATP-dependent acid: CoA ligases. The classification of these ligases are based on their role of ligating a fatty acid to CoA and exhibit selectivity for short-, medium-, long-, or very long-chain fatty acids. Fatty acid activation activity was only characterized biochemically in recent years by their chain length preference, so the discovery of the conserved motifs I and II, discussed in Section 2.9.4 is therefore extremely beneficial in not only the classification of these ACSs, but also the discovery of new candidate human ACS genes and proteins (Watkins *et al*., 2007).

Current literature regarding the HXM-A ligase substrate specificities and genetic variations in the *ACSM2B* gene is limited. One of the main objectives of this study was to inform on the aspect of the *ACSM2A* and *ACSM2B* gene nomenclature and highlight the basic differences between these genes to clear up the confusion in the literature.

The discovery of the conserved motif I and II and specifically the non-synonymous substitution identified within the conserved amino acid sequence motif II identified in human ACSs, play an essential part in the distinction between the *ACSM2A* and *ACSM2B* genes.

The following Section 5.2.2.1 will point out discrepancies in the current literature regarding the *ACSM2A* and *ACSM2B* genes.

5.2.2.1 Discrepancies in the current literature regarding the *ACSM2A* and *ACSM2B* genes

The following literature is discussed that currently adds to the confusion surrounding the *ACSM2A* and *ACSM2B* genes:

i. Iwai *et al*., (2003); Linder *et al*., (2006)

Past literature does not differentiate clearly between the two different forms of the *ACSM2* gene, namely *ACSM2A* and *ACSM2B*, by only referring to these genes as *ACSM2*.

Iwai *et al*., (2003) reported on a study conducted on an acyl CoA synthetase gene family in chromosome 16p12 contributing to multiple risk factors in a Japanese Suita cohort. The L513S polymorphism in the *MACS2* (Synonym: *ACSM2*) gene was found to contribute to multiple risk factors of the metabolic syndrome. Linder *et al*., (2006) also reported on the association between the L513S polymorphism in MACS2 and different risk factors in the Caucasian Metabolic Intervention Cohort Kiel.

Linder *et al*., (2006), as well as Iwai *et al*., (2003), fail to distinguish between the *ACSM2A* and *ACSM2B* genes, and also report on the incorrect SNP. The S513L polymorphism is a non-synonymous SNP identified in the *ACSM2A* gene, and not the L513S polymorphism as stated by Linder *et al*., (2006). Linder *et al*., (2006) and Iwai *et al*., (2003) also neglect to indicate the corresponding variant identification number for the S513L polymorphism as well as the reference sequence.

ii. Boomgaarden *et al*., (2009)

Boomgaarden *et al*., (2009) conducted a study on the comparative analyses of disease risk genes belonging to the ACSM family in human liver and cell lines, where they have shown that the *ACSM2B* gene is the predominant transcript in human liver. Therefore, suggesting that genetic variations of the *ACSM2B* gene play an important role in disease susceptibility.

The transcript levels of *ACSM2A* and *ACSM2B* in human liver samples were quantified using a TaqMan qPCR assay. Assays were designed to detect *ACSM2A/B* Total and *ACSM2B* (probe used: ACATCCTGAGAGGACTTG). It was found that *ACSM2A* is one of the least expressed transcripts with 0.24% of *ACSM2* total expression levels. They concluded that the results therefore suggest that *ACSM2B* is the most abundant transcript. When the probe

designed by Boomgaarden *et al*., (2009) is used as input sequence on NCBI nucleotide BLAST, this probe aligned perfectly with both the *ACSM2A* (ENST00000573854) and *ACSM2B* (ENST000000567001) reference sequences, as proposed by Watkins *et al*., (2007).

The question arises on whether the *ACSM2B* gene is, as Boomgaarden *et al*., (2009) suggests, in fact the predominant transcript in human liver. This study needs to be repeated to verify the results.

Twenty-one nucleotide differences (17 non-synonymous substitutions and 4 synonymous substitutions) exist in the coding region for the *ACSM2A* and *ACSM2B* genes, as indicated in Figure 2.5. Only one non-synonymous substitution lies within a conserved motif. Residue 463, found in Motif II, is Asn in the *ACSM2A* gene and Asp in the *ACSM2B* gene (Table 2.9) (Watkins *et al*., 2007).

For future studies, a fluorescent probe for SNP detection needs to be designed for *ACSM2B*, distinctive of that for *ACSM2A* in order to distinguish between *ACSM2A* and *ACSM2B*.

iii. Agundez *et al*., (2009)

A study by Agundez *et al* (2009) on pharmacogenomics in Aspirin intolerance reported on the importance of genetic variations in enzymes involved in the metabolism of aspirin as well as common polymorphisms related in aspirin biodisposition. UGT1A6, ACSM2B, and CYP2C9, are the three major enzymes involved in aspirin metabolism.

Inconsistencies related to the abstract and the conclusion of the study conducted by Agundez *et al.,* (2009) exists. As stated in the abstract by Agundez *et al.,* (2009), a common polymorphism (rs1133607, S513L) for the *ACSM2A* gene was identified as a major polymorphism related to aspirin biodisposition. Suggesting that genetic variation in the acid: CoA ligase enzymes can influence the metabolism of salicylate and other benzoate derivatives. On the other hand, the conclusion made by Agundez *et al.,* (2009), states that no common and functionally significant non-synonymous SNP with regard to aspirin intolerance has, as of yet, been described for the *ACSM2B* gene.

The glycine conjugation of salicylate is an important pathway in aspirin metabolism, so future studies investigating the potential role of polymorphisms in the *ACSM2B* gene may contribute to the understanding of the conversion of salicylic acid to salicyluric acid.

iv. Badenhorst *et al*., (2013, 2014)

Review articles by Badenhorst *et al*., (2014) on the importance of glycine conjugation in the metabolism of aromatic acids, and Badenhorst *et al*., (2013) on the role of glycine N-acyltransferase, and factors that influence interindividual variation, refers to HXM-A (EC 6.2.1.2) as encoded by the *ACSM2A* gene. However, it is still not clear which protein is encoded by the *ACSM2A* gene or what the physiological role is.

Badenhorst *et al*., (2014) also report on an SNP (rs1133607), associated with aspirin intolerance, in the *ACSM2A* gene, which encodes HXM-A. Although the rs1133607 SNP does occur in the *ACSM2A* gene, HXM-A (EC 6.2.1.2) is encoded by the *ACSM2B* gene, as stated in the current study.

5.2.3 Investigation of genetic variation on available public databases and the SA population of the present study

The aim of this study was to identify genetic variation in the acyl CoA synthetase genes (*ACSM2A* and *ACMS2B*) in not only the present study (19 South African Afrikaner Caucasian individuals), but also the 1000 Genomes Project (2015 individuals of the AFR, AMR EAS, and EUR population), and the NHLBI ESP (6503 individuals of the AA, and EA population) in order to identify if the ORF is conserved. These publicly available databases were used to investigate the non-synonymous SNPs identified in the *ACSM2A* and *ACSM2B* genes.

The second objective of the present study was to generate DNA sequence data of the coding region of 19 South African Afrikaner Caucasian individuals for the present study. This was done using an Ion Proton NGS approach at the CAF unit in Stellenbosch, to identify known and/or possible novel variants of the *ACSM2A* and *ACSM2B* genes. From the sequencing run, and after comparison against reference sequences, a total of 25 variants were identified.

The bioinformatics pipeline identified two novel and twelve known polymorphisms of the *ACSM2A* gene, and three novel and eight known polymorphisms of the *ACSM2B* gene in the 19 South African Afrikaner Caucasian individuals. All non-synonymous variants identified in

both genes were found at low frequencies or as heterozygotes. Novel variants identified were located in the intron regions.

It is essential to understand that not one NGS workflow is without error and that various databases should be used to compare results. To ensure the most reliable results the in-house developed bioinformatics pipeline used in the present study consisted of available open source databases to run VCF data files obtained from the Torrent Suite server (v4.0.2) through the Ensemble online VEP runner (variant effect predictor, v73, http://www.ensembl.org) (McLaren *et al*., 2010), and genotype detail obtained from Ion Reporter software (lifetechnologies.com/ionreporter), after which results were compared to data files obtained from the Ion Reporter software to ensure consistency in reported variants.

Of the 47 (1000 Genomes Project, Table 4.2), 15 (National Heart, Lung and Blood Institute Exome Sequencing Project, NHLBI ESP, Table 4.4), and 4 (South African Afrikaner Caucasian Population, SA, Table 4.6) non-synonymous SNPs identified within the coding region of the *ACSM2A* gene, the L64P variant had the highest homozygous SNP genotype frequency (29.0%), followed by the N463D (12.5%), and the R5Q variant (5.5%) (Table 4.8). All other variants were found at frequencies <5%.

Of the 43 (1000 Genomes Project, Table 4.3), 15 (National Heart, Lung and Blood Institute Exome Sequencing Project, NHLBI ESP, Table 4.5), and 1 (South African Afrikaner Caucasian Population, SA, Table 4.7) non-synonymous SNPs identified within the coding region of the *ACSM2B* gene, the T278A variant had the highest homozygous SNP genotype frequency (4.0%), followed by the I305V variant (0.7%), and the D322N variant (0.1%) (Table 4.9).

The results of this study indicated that variation within the *ACSM2A* and *ACSM2B* ORF is very low. It is evident from Table 4.9 that very low genotype frequencies exist for the SNPs identified within the coding region of the *ACSM2B* gene (T278A: 4.0%, I305V: 0.7%, D322N: 0.1%) compared to genotype frequencies identified for *GLYAT* (N156S: 90%, S17T: 4.6%; R131H: 0.1%); from a study conducted by Van der Sluis *et al*., (2015). With increased levels of benzoic acid exposure in humans, the HXM-A protein might not be able to effectively detoxify such large amounts. Thus, findings underline the importance of future investigations into the *ACSM2A* and *ACSM2B* genes, and their proteins to better understand the effect of SNPs on protein function.

5.3 Concluding remarks

For years, amino acid conjugation was regarded as relatively unimportant in humans and was only studied as a single-step process with emphasis on the final step (GLYAT). The glycine conjugation pathway was significantly overshadowed by research on the P450 system, responsible for Phase I of the biotransformation process.

However, a lack of amino acid conjugation contributes significantly to the formation of xenobiotic CoA within the mitochondrial matrix as discussed throughout Chapter 2 of the present study. The continued studying and examination into the role of the HXM-ligases will contribute significantly to the characterization of these ligases, their substrate/inhibitor profiles and the role of these enzymes in metabolism.

Badenhorst *et al*., (2014) and Van der Sluis *et al*., (2015) concluded that seeing as the *GLYAT* gene is part of the original phenylpropionate catabolism pathway, and the fact that no defects for the glycine conjugation pathway have been reported, the *GLYAT* gene is essential for survival and the ORF is highly conserved. The present study was formulated to characterise the genetic variation in the ORF of the *ACSM2A* and *ACSM2B* genes in order to determine if the *ACSM2B* gene is also as highly conserved as GLYAT. Results obtained in the present study indicated that non-synonymous SNPs identified in the acyl CoA synthetase genes (*ACSM2A* and *ACSM2B*) in the 1000 Genomes data, the NHLBI ESP and the present study, were detected at low frequencies or as heterozygotes, and are generally only present in a single population group.

This further supports the hypothesis that the glycine conjugation pathway is therefore an essential detoxification mechanism. From the allele frequencies identified in this study, it can be proposed that the current reference sequence used in the present study for the *ACSM2A* and *ACSM2B* genes should probably be considered as the wild-type as no clear allele of high frequency was identified in all populations as for *GLYAT* (S156 haplotype) (Van der Sluis *et al*., 2015).

The present study contributed significantly to better understanding the nomenclature regarding the acyl CoA synthetases, especially confusion regarding the *ACSM2A* and *ACSM2B* genes. The likelihood of incorrectly reported SNPs in previous studies may be due to an incorrect reference transcript used. In order to avoid any future confusion regarding the *ACSM2A* and *ACSM2B* genes, it is suggested that nomenclature and transcripts proposed in the study conducted by Watkins *et al*., (2007) should be used in any future research.

These findings underline the importance of future investigations into these genes.

5.4 Future prospects

The following suggestions are made to contribute to future research on the *ACSM2A* and *ACSM2B* genes:

1. Future investigations may include the expression of the recombinant *ACSM2A* and *ACSM2B* genes and investigations of substrate specificity.
2. Possible future investigations adapted from a previous study on the N-acyltransferase gene conducted by Van der Sluis *et al*., (2015) may include the following:
   I. The inclusion of *ACSM2A* and *ACSM2B* sequences obtained from the Khoi-San individuals and one Bantu individual from southern Africa (Delaneau *et al*., 2012) available on the Sequencing read Archive (SRA) database (www.ncbi.nlm.nih.gov/sra). As mentioned in Section 4.2.3.1, it is clear that the AFR and SA population vary significantly in genotype frequency and non-synonymous SNPs identified within the specific populations. This may be due to the fact that the SA population used in the present study is not African in recent descent, but rather European. For this reason, the inclusion of sequences obtained from the Khoi-San individuals and one Bantu individual from Southern Africa (Delaneau *et al*., 2012), is necessary. On the other hand, the inclusion of a broader range of South Africans to the study is also necessary, and not just the inclusion of individuals from European descent.
   II. Identification and phasing of haplotypes of the 1000 Genomes data, the Present study and the Bantu and Khoisan data, may be another aspect included in future studies as haplotype distributions among population groups present insights in the diversity, history as well as migration patterns of human populations (Delaneao *et al*., 2012, Delaneau *et al*., 2013, Van der Sluis *et al*., 2015). Phased haplotypes of 1000 Genomes data are available on the USCS Genome Browser (genome.ucsc.edu/).
   III. A last prospect for future studies is the inclusion of phylogenetic analyses contributing to the evolutionary conservation of the coding region of the *ACSM2A* and *ACSM2B* genes. Phylogenetic analysis may assist in providing insight to which haplotype represents the ancestral allele (Tamura *et al*., 2013; Edgar, 2004; Van der Sluis *et al*., 2015).
3. Future investigation of the present study includes validation of sequencing results (SA population) by Sanger sequencing.
4. Additional research on the *ACSM2A* gene is needed as very little is presented in current literature about the gene function and no information exist on the specific protein coding for *ACSM2A* as well as substrate specificities.

Bibliography

1000 GENOMES PROJECT CONSORTIUM. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature, 491(7422):56-65.

ABECASIS, G.R., *et al*., 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature,* 491 (7422), 56–65.

AGUNDEZ, JOSE, A.G., *et al*., 2009. Pharmacogenomics in aspirin intolerance. *Current drug metabolism,* 10 (9): 998-1008.

AKIRA, K., NEGISHI, E., IMACHI, M. & HASHIMOTO, T. 2001. Direct nuclear magnetic resonance spectroscopic analysis of 13C-labeled antipyrine metabolites in human urine. *Drug metabolism and disposition*, 29, 903-907.

ANDERSEN, M. E. & MCKENNA, M. J. 1981. Saturable metabolism and its relationship to toxicity. *CRC Critical Reviews in Toxicology*, 9, 105-150.

ANTONARAKIS, S. E. 1998. Recommendations for a nomenclature system for human gene mutations. *Human mutation*, 11, 1-3.

BADENHORST, C. P. S., ERASMUS, E., VAN DER SLUIS, R., NORTJE, C. & VAN DIJK, A. A. 2014. A new perspective on the importance of glycine conjugation in the metabolism of aromatic acids. *Drug metabolism reviews*, 46, 343-361.

BADENHORST, C. P. S., JOOSTE, M. & VAN DIJK, A. A. 2012. Enzymatic characterization and elucidation of the catalytic mechanism of a recombinant bovine glycine N-acyltransferase. *Drug Metab Dispos,* 40**,** 346-52.

BADENHORST, C. P. S., VAN DER SLUIS, R., ERASMUS, E. & VAN DIJK, A. A. 2013. Glycine conjugation: importance in metabolism, the role of glycine N-acyltransferase, and factors that influence interindividual variation. *Expert opinion on drug metabolism & toxicology*, 9, 1139-1153.

BALES, J. R., HIGHAM, D. P., HOWE, I., NICHOLSON, J. K. & SADLER, P. J. 1984. Use of high-resolution proton nuclear magnetic resonance spectroscopy for rapid multi-component analysis of urine. *Clinical chemistry*, 30, 426-432.

BANDMANN, O., VAUGHAN, J., HOLMANS, P., MARSDEN, C.D., WOOD, N.W. 1997. Association of slow acetylator genotype for N-acyltransferase 2 with familial Parkinson's disease. *Lancet,* 350, 1136-1139.

BELIVEAU, G. P. & BRUSILOW, S. W. 1987. Glycine availability limits maximum hippurate synthesis in growing rats. *The Journal of nutrition*, 117, 36.

BENET, L.Z. 27th Gordon research conference on drug metabolism. July 6-13, 1997, *personal communication.*

BERNAUDIN, J. F. *et al*. 2009. ABC Proteins and Oncology: Expression, Detection and Implication of ABC Proteins in Solid Tumors. in ABC Transporters and Multidrug Resistance

(eds Boumendjel, A., Boutonnat, J. & Robert, J.), *Wiley, & Sons Inc*., Hoboken, New Jersey, USA, 143–176.

BEYOGLU, D., SMITH, R. L. & IDLE, J. R. 2012. Dog bites man or man bites dog? The enigma of the amino acid conjugations. *Biochem Pharmacol*, 83, 1331-9.

BLACK, P. N., Q. ZHANG, J. D. WEIMAR, and C. C. DIRUSSO. 1997. Mutational analysis of a fatty acyl-acyl coenzyme A synthetase signature motif identifies seven amino acid residues that modulate fatty acid substrate specificity. *J. Biol. Chem*, 272: 4896–4903.

BOLAND, J. F., CHUNG, C. C., ROBERSON, D., MITCHELL, J., ZHANG, X., IM, K. M., HE, J., CHANOCK, S. J., YEAGER, M. & DEAN, M. 2013. The new sequencer on the block: comparison of Life Technology's Proton sequencer to an Illumina HiSeq for whole-exome sequencing. *Human genetics*, 132, 1153-1163.

BOOMGAARDEN, I., VOCK, C., KLAPPER, M. & DORING, F. 2009. Comparative analyses of disease risk genes belonging to the acyl-acyl CoA synthetase medium-chain (ACSM) family in human liver and cell lines. *Biochem Genet,* 47**,** 739-48.

BRINKMANN, U., ROOTS, I. & EICHELBAUM, M. 2001. Pharmacogenetics of the human drug-transporter gene MDR1: impact of polymorphisms on pharmacotherapy. *Drug discovery today*, 6, 835-839.

BUCHET, J. & LAUWERYS, R. 1973. Measurement of urinary hippuric and m-methylhippuric acids by gas chromatography. *British journal of industrial medicine*, 30, 125-128.

CHIN, K.V., PASTAN, I., GOTTESMAN, M.M. 1993. Function and regulation of the multidrug resistance gene. *Adv Cancer Res*, 60:157-180.

CHORLEY, B. N., X. WANG, M. R. CAMPBELL, G. S. PITTMAN, M. A. NOUREDDINE *et al*., 2008. Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies. *Mutat Res,* 659: 147-157.

CONTI, A. & BICKEL, M. H. 1977. History of Drug Metabolism: Discoveries of the Major Pathways in the 19th Century. Drug Metabolism Reviews, 6**,** 1-50. DELANEAU, O., MARCHINI, J. & ZAGURY, J. F. 2012. A linear complexity phasing method for thousands of genomes. *Nat Methods*, 9, 179-81.

CUNNINGHAM, F., AMODE, M. R., BARRELL, D., BEAL, K., BILLIS, K., BRENT, S., CARVALHO-SILVA, D., CLAPHAM, P., COATES, G. & FITZGERALD, S. 2015. Ensemble 2015. *Nucleic acids research*, 43, D662-D669.

DELANEAU, O., ZAGURY, J. F. & MARCHINI, J. 2013. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods*, 10, 5-6.

DEN DUNNEN, J.T. & ANTONARAKIS, S.E. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: A discussion. *Human mutation*, 15(1):7-12.

DEN DUNNEN1N, J. T. & PAALMAN, M. H. 2003. Standardizing mutation nomenclature: why bother? *Changes*, 100, 123-126.

DUEÑAS, M., MUÑOZ-GONZÁLEZ, I., CUEVA, C., JIMÉNEZ-GIRÓN, A., SÁNCHEZ-PATÁN, F., SANTOS-BUELGA, C., MORENO-ARRIBAS, M. & BARTOLOMÉ, B. 2015. A survey of modulation of gut microbiota by dietary polyphenols. *BioMed research international, 2015.*

EATON, S., BARTLETT, K., POURFARZAM, M. 1996. Mammalian mitochondrial b-oxidation. *Biochem J*; 320:345-57.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res,* 32**,** 1792-7.

EPIS, S., PORRETTA, D., MASTRANTONIO, V., COMANDATORE, F., SASSERA, D., ROSSI, P., CAFARCHIA, C., OTRANTO, D., FAVIA, G. & GENCHI, C. 2014. ABC transporters are involved in defense against permethrin insecticide in the malaria vector Anopheles stephensi. *Parasites & vectors*, 7, 349.

EPIS, S., PORRETTA, D., MASTRANTONIO, V., URBANELLI, S., SASSERA, D., DE MARCO, L., MEREGHETTI, V., MONTAGNA, M., RICCI, I. & FAVIA, G. 2014. Temporal dynamics of the ABC transporter response to insecticide treatment: insights from the malaria vector Anopheles stephensi. *Scientific reports*, 4.

FEUK, L., A. R. CARSON and S. W. SCHERER, 2006. Structural variation in the human genome. Nat *Rev Genet* 7: 85-97.

FILLMORE, N., MORI, J. & LOPASCHUK, G. 2014. Mitochondrial fatty acid oxidation alterations in heart failure, ischaemic heart disease and diabetic cardiomyopathy. *British journal of pharmacology*, 171, 2080-2090.

FOSTER, D. W. 2004. The role of the carnitine system in human metabolism. *Annals of the New York Academy of Sciences,* 1033, 1-16.

FU, W., O'CONNOR, T.D., JUN, G., KANG, H.M., ABECASIS, G., LEAL, S.M., GABRIEL, S., ALTSHULER, D., SHENDURE, J., NICKERSON, D.A. *et al.,* .2013. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature*, 493, 216–220. 21. UniProt Consortium. (2013) Update on activities at the Universal.

FUJINO, T., TAKEI, Y. A., SONE, H., IOKA, R. X., KAMATAKI, A., MAGOORI, K., TAKAHASHI, S., SAKAI, J. & YAMAMOTO, T. T. 2001. Molecular identification and characterization of two medium-chain acyl -CoA synthetases, MACS1 and the Sa gene product. *J Biol Chem*, 276, 35961-6.

FURA, A., SHU, Y.-Z., ZHU, M., HANSON, R. L., ROONGTA, V. & HUMPHREYS, W. G. 2004. Discovering drugs through biological transformation: role of pharmacologically active metabolites in drug discovery. *Journal of medicinal chemistry*, 47, 4339-4351.

GIBSON, G. & SKETT, P. 2001. Assay for glutathione-S-transferase activity. Introduction to drug metabolism. Cheltenham, UK: *Nelson Thornes*, 47-49.

GONZALEZ, F. J. & TUKEY, R. H. 2006. Drug metabolism. Goodman & Gilman's The Pharmacological Basis of Therapeutics. 11th ed. New York, NY: *McGraw-Hill*, 71-91.

GREGERSEN, N., KOLVRAA, S. & MORTENSEN, P. B. 1986. Acyl-Acyl CoA: glycine N-acyltransferase: in vitro studies on the glycine conjugation of straight- and branched-chained acyl-acyl CoA esters in human liver. *Biochem.Med.Metab Biol*., 35, 210-218.

GRONAU, S., KOENIG-GREGER, D., JERG, M. & RIECHELMANN, H. 2003. Gene polymorphisms in detoxification enzymes as susceptibility factor for head and neck cancer? *Otolaryngology--Head and Neck Surgery*, 128, 674-680.

GROOT, P. H. 1976. Acyl-Acyl CoA synthetases in guinea-pig liver mitochondria. Purification and characterization of a distinct propionyl-CoA synthetase. *Biochim Biophys Acta*, 441, 260-7.

HEATHFIELD, L., LACERDA, M., NOSSEK, C., ROBERTS, L. & RAMESAR, R. S. 2013. Stargardt disease: towards developing a model to predict phenotype. *Eur J Hum Genet*, 21, 1173-6.

HUNT, M. C. & ALEXSON, S. E. 2008. Novel functions of acyl -CoA thioesterases and acyltransferases as auxiliary enzymes in peroxisomal lipid metabolism. *Progress in lipid research*, 47, 405-421.

IOANNIDES, C. 2002. Enzyme Systems that Metabolise Drugs and Other Xenobiotics, *Wiley,* New York, 1–578.

IOANNIDES, C., BARNETT, C., IRIZAR, A. & FLATT, P. 1996. Expression of cytochrome P450 proteins in disease. Cytochromes P450: *Metabolic and Toxicological Aspects*, 301-327.

IWAI, N., MANNAMI, T., TOMOIKE, H., ONO, K. & IWANAGA, Y. 2003. An acyl -CoA synthetase gene family in chromosome 16p12 may contribute to multiple risk factors. *Hypertension*, 41, 1041-1046.

J. A., MARTINEZ, C., PEREZ-SALA, D., CARBALLO, M., TORRES, M. J. & GARCIA-MARTIN, E. 2009. Pharmacogenomics in aspirin intolerance. *Curr Drug Metab*, 10, 998-1008.

JAFFE, A., WOJCIK, G., CHU, A., GOLOZAR, A., MAROO, A., DUGGAL, P. & KLEIN, A. P. Identification of functional genetic variation in exome sequence analysis. BMC proceedings, 2011. *BioMed Central Ltd*, 5(9).

JANČOVÁ, P., ANZENBACHER, P. & ANZENBACHEROVA, E. 2010. Phase II drug metabolizing enzymes. *Biomed Pap Med*, Fac Univ Palacky Olomouc Czech Repub, 154, 103-16.

JANČOVÁ, P., ŠILLER, M. Phase II Drug Metabolism, in: J. Paxton (Ed.), *Topics on DrugMetabolism, InTech*., 35–60.

JULIANO, R. L. & LING, V. 1976. A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 455, 152-162.

KASUYA, F., IGARASHI, K., FUKUI, M., NOKIHARA, K. Purification and characterization of a medium chain acyl-acyl coenzyme A synthetase. *Drug Metab Dispos*, 24(8):879-83.

KASUYA, F., YAMAOKA, Y., OSAWA, E., *et al*. 2000. Difference of the liver and kidney in glycine conjugation of ortho-substituted benzoic acids. *Chem Biol Interact*; 125:39-50.

KAWAI, T., UKAI, H., INOUE, O., MAEJIMA, Y., FUKUI, Y., OHASHI, F., OKAMOTO, S., TAKADA, S., SAKURAI, H. & IKEDA, M. 2008. Evaluation of biomarkers of occupational exposure to toluene at low levels. *International archives of occupational and environmental health*, 81, 253-262.

KAWAJIRI, K., NAKACHI, K., IMAI, K., YOSHII, A., SHINODAA, N., WATANABE, J. 1990. Identification of genetically high risk individuals to lung cancer by DNA polymorphisms of the cytochrome P450IA1 gene *FEBS Lett*. 263:131-133.

KELLER, W. 1842. On the Conversion of Benzoic Acid into Hippuric Acid. Provincial Medical *Journal and Retrospect of the Medical Sciences*, 4, 256-257.

KILLENBERG, P. G., DAVIDSON, E. D. & WEBSTER, L. T., JR. 1971. Evidence for a medium-chain fatty acid: coenzyme A ligase (adenosine monophosphate) that activates salicylate. *Mol Pharmacol*, 7, 260-8.

KIRA, S. 1977. Measurement by gas chromatography of urinary hippuric acid and methylhippuric acid as indices of toluene and xylene exposure. *British journal of industrial medicine*, 34, 305-309.

KNIGHTS, K. M. & DROGEMILLER, C. J. 2000. Xenobiotic-CoA Ligases: Kinetic and Molecular Characterization. *Current Drug Metabolism*, 1, 49-66.

KNIGHTS, K. M. 1998. Role of hepatic fatty acid: coenzyme A ligases in the metabolism of xenobiotic carboxylic acids. *Clininal and Experimental Pharmacology and Physiology*, 25, 776-782.

KNIGHTS, K.M., MINERS, J.O. 2012. Amino acid conjugation: A novel route of xenobiotic carboxylic acid metabolism in man. In: AV Lyubimov, ed. Encyclopedia of Drug Metabolism and Interactions. Somerset, USA: *John Wiley and Sons*, 595–610.

KNIGHTS, K.M., SYKES, M.J., MINERS, J.O., 2007 Amino acid conjugation: contribution to the metabolism and toxicity of xenobiotic carboxylic acids. *Expert Opin. Drug Metab. Toxicol*. 3 (2), 159-168.

KOCHAN, G., PILKA, E. S., VON DELFT, F., OPPERMANN, U. & YUE, W. W. 2009. Structural snapshots for the conformation-dependent catalysis by human medium-chain acyl-acyl coenzyme A synthetase ACSM2A. *Journal of molecular biology*, 388, 997-1008.

KOLVRAA, S. & GREGERSEN, N. 1986. Acyl-Acyl CoA:glycine N-acyltransferase: organelle localization and affinity toward straight- and branched-chained acyl-acyl CoA esters in rat liver. *Biochem.Med.Metab Biol*., 36**,** 98-105.

KONGTIP, P., VARARUSSAMI, J. & PRUKTHARATHIKUL, V. 2001. Modified method for determination of hippuric acid and methylhippuric acid in urine by gas chromatography. Journal of Chromatography B: *Biomedical Sciences and Applications*, 751, 199-203.

KRUSE, M., SCHULZE-BAHR, E., CORFIELD, V., BECKMANN, A., STALLMEYER, B., KURTBAY, G., OHMERT, I., SCHULZE-BAHR, E., BRINK, P. & PONGS, O. 2009. Impaired

endocytosis of the ion channel TRPM4 is associated with human progressive familial heart block type I. *J Clin Invest*, 119, 2737-44.

LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M. & FITZHUGH, W. 2001. Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921.

LEES, H. J., SWANN, J. R., WILSON, I. D., NICHOLSON, J. K. & HOLMES, E. 2013. Hippurate: The Natural History of a Mammalian–Microbial Cometabolite. *Journal of proteome research*, 12, 1527-1546.

LIN, J. H. & LU, A. Y. 1997. Role of pharmacokinetics and metabolism in drug discovery and development. *Pharmacological reviews*, 49, 403-449.

LINDNER, I., RUBIN, D., HELWIG, U., NITZ, I,, HAMPE, J., SCHREIBER, S., SCHREZENMEIR, J., DORING, F. 2006. The L513S polymorphism in medium-chain acyl - CoA synthetase 2 (MACS2) is associated with risk factors of the metabolic syndrome in a Caucasian study population. *Mol Nutr Food Res* 50:270–274.

LISKA, D. J. 1998. The detoxification enzyme systems. *Altern Med Rev*, 3, 187-98.

LISKA, D., LYON, M. & JONES, D. S. 2006. Detoxification and biotransformational imbalances. *The Journal of Science and Healing*, 2, 122-140.

LIU, L., LI, Y., LI, S., HU, N., HE, Y., PONG, R., LIN, D., LU, L. & LAW, M. 2012. Comparison of next-generation sequencing systems. *BioMed Research International*., 251364. *PMC. Web*. 18 Nov. 2015.

LOFTUS, B. J., KIM, U. J., SNEDDON, V. P., KALUSH, F., BRANDON, R., FUHRMANN, J., MASON, T., CROSBY, M. L., BARNSTEAD, M., CRONIN, L., DESLATTES MAYS, A., CAO, Y., XU, R. X., KANG, H. L., MITCHELL, S., EICHLER, E. E., HARRIS, P. C., VENTER, J. C. & ADAMS, M. D. 1999. Genome duplications and other features in 12 Mb of DNA sequence from human chromosome 16p and 16q. *Genomics*, 60, 295-308.

LOHR, J. W., WILLSKY, G. R. & ACARA, M. A. 1998. Renal drug metabolism. *Pharmacological reviews*, 50, 107-142.

LUPSKI, J. R., GONZAGA-JAUREGUI, C., YANG, Y., BAINBRIDGE, M. N., JHANGIANI, S., BUHAY, C. J., KOVAR, C. L., WANG, M., HAWES, A. C. & REID, J. G. 2013. Exome sequencing resolves apparent incidental findings and reveals further complexity of SH3TC2 variant alleles causing Charcot-Marie-Tooth neuropathy. *Genome Med*, 5, 57.

MAHLER, H. R., WAKIL, S. J. & BOCK, R. M. 1953. Studies on fatty acid oxidation. I. Enzymatic activation of fatty acids. *J Biol Chem*, 204, 453-68.

MANACH, C., SCALBERT, A., MORAND, C., REMESY, C. & JIMENEZ, L. 2004. Polyphenols: food sources and bioavailability. *Am J Clin Nutr*, 79, 727-47.

MANCINELLI, L., M. CRONIN and W. SADEE, 2000. Pharmacogenomics: the promise of personalized medicine. *AAPS PharmSci*, 2: E4.

MARDIS, E. R., 2008. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 9: 387-402.

MARGULIES, M., EGHOLM, M., ALTMAN, W.E., ATTIYA, S., BADER, J.S., BEMBEN, L.A., BERKA, J., BRAVERMAN, M.S., CHEN, Y., CHEN, Z. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057):376-380.

MARTIN, J., HAN, C., GORDON, L. A., TERRY, A., PRABHAKAR, S., SHE, X., XIE, G., HELLSTEN, U., CHAN, Y. M., ALTHERR, M., COURONNE, O., AERTS, A., BAJOREK, E., BLACK, S., BLUMER, H., BRANSCOMB, E., BROWN, N. C., BRUNO, W. J., BUCKINGHAM, J. M., CALLEN, D. F., *et al*. 2004. The sequence and analysis of duplication-rich human chromosome 16. *Nature*, 432, 988-94.

MCCARROLL, S. A., and D. M. ALTSHULER, 2007. Copy-number variation and association studies of human disease. *Nat Genet,* 39: S37-42.

MCLAREN, W., PRITCHARD, B., RIOS, D., CHEN, Y., FLICEK, P., CUNNINGHAM, F. 2010. Deriving the consequences of genomic variants with the ensembl API and SNP effect predictor. *Bioinformatics*, 26(16):2069-2070.

MELDRUM, C., DOYLE, M.A., TOTHILL, R.W. 2011. Next-generation sequencing for cancer diagnostics: A practical perspective. *The clinical biochemist.reviews / australian association of clinical biochemists*, 32(4):177-195.

MERRIMAN, B., TORRENT, I., ROTHBERG, J.M., R&D TEAM. 2012. Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*, 33(23):3397-3417.

MITCHELL, G. A., GAUTHIER, N., LESIMPLE, A., WANG, S. P., MAMER, O. & QURESHI, I. 2008. Hereditary and acquired diseases of acyl-acyl coenzyme A metabolism. *Mol.Genet.Metab*, 94, 4-15.

MURPHY, P. J. 2001. Xenobiotic metabolism: a look from the past to the future. *Drug Metabolism and Disposition*, 29, 779-780.

MURPHY, P. J. 2008. The development of drug metabolism research as expressed in the publications of ASPET: Part 1, 1909–1958. *Drug Metabolism and Disposition*, 36, 1-5.

Nair, B., 2001. Final report on the safety assessment of benzyl alcohol, benzoic acid, and sodium benzoate. *Int. J. Toxicol*. 20 (Suppl 3), 23–50.

NANDI, D. L., LUCAS, S. V. & WEBSTER, L. T., JR. 1979. Benzoyl-coenzyme A:glycine N-acyltransferase and phenylacetyl-coenzyme A:glycine N-acyltransferase from bovine liver mitochondria. Purification and characterization. *J.Biol.Chem*., 254, 7230-7237.

NEBERT, D., EISEN, H., NEGISHI, M., LANG, M., HJELMELAND, L. & OKEY, A. B. 1981. Genetic mechanisms controlling the induction of polysubstrate monooxygenase (P-450) activities. *Annual review of pharmacology and toxicology*, 21, 431-462.

NIEDRINGHAUS, T.P., MILANOVA, D., KERBY, M.B., SNYDER, M.P., BARRON, A.E. 2011. Landscape of next-generation sequencing technologies. *Analytical chemistry*, 83(12):4327-4341.

OMIECINSKI, C. J., HEUVEL, J. P. V., PERDEW, G. H. & PETERS, J. M. 2011. Xenobiotic metabolism, disposition, and regulation by receptors: from biochemical phenomenon to predictors of major toxicities. *Toxicological Sciences*, 120, S49-S75.

PABINGER, S., DANDER, A., FISCHER, M., SNAJDER, R., SPERK, M., EFREMOVA, M., KRABICHLER, B., SPEICHER, M.R., ZSCHOCKE, J., TRAJANOSKI, Z. 2014. A survey of tools for variant analysis of next-generation genome sequencing data. *Briefings in bioinformatics*, 15(2):256-278.

PARK, H.-M., LEE, S.-H., CHUNG, H., KWON, O.-H., YOO, K.-Y., KIM, H.-H., HEO, S.-C., PARK, J.-S. & TAE, G.-S. 2007. Immunochromatographic analysis of hippuric acid in urine. *Journal of analytical toxicology*, 31, 347-353.

PARKINSON A., Biotransformation of xenobiotics.1996. *In: Klassen C. D., Ed. Cassarett and Doull's Toxicology: The Basic Science of Poisons*, New York: McGraw-Hill, 5, 113–186.

PARKINSON, A., OGILVIE, B. 2008. Biotransformation of xenobiotics In Casarett & Doull's Toxicology: The Basic Science of Poisons. *McGraw -Hill* , New York, 7, 161-304.

PENNER, N., WOODWARD, C., PRAKASH, C. 2012. Metabolizing enzymes and biotransformation reactions. In:Zhang D, Surapaneni S, editors. ADME-Enabling technologies in drug design and development. Hoboken, *NJ: Wiley*; 545–565.

PIPER, P. W. 1999. Yeast superoxide dismutase mutants reveal a pro-oxidant action of weak organic acid food preservatives. *Free Radic Biol Med*, 27, 1219-27.

PRITCHARD, J. K., 2001. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet*, 69: 124-137.

QUICK, A. J. 1931. The conjugation of benzoic acid in man. *Journal of Biological Chemistry*, 92, 65-85.

REILLY, S.-J., O'SHEA, E.M., ANDERSSON. U., O'BYRNE, J., ALEXSON, S.E.H., HUNT, M.C. 2007. A peroxisomal acyltransferase in mouse identifies a novel pathway for taurine conjugation of fatty acids. *FASEB J*, 21:99–107.

RODRIGUEZ-MURILLO, L., XU, B., ROOS, J. L., ABECASIS, G. R., GOGOS, J. A. & KARAYIORGOU, M. 2014. Fine mapping on chromosome 13q32-34 and brain expression analysis implicates MYO16 in schizophrenia. *Neuropsychopharmacology*, 39, 934-43.

ROTHBERG, J.M., HINZ, W., REARICK, T.M., SCHULTZ, J., MILESKI, W., DAVEY, M., LEAMON, J.H., JOHNSON, K., MILGREW, M.J., EDWARDS, M. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475(7356):348-352.

RUSHMORE, T. H. and KONG, A. N. 2002. Pharmacogenomics, regulation and signaling pathways of phase I and II drug metabolizing enzymes. *Curr. Drug Metab*, 3, 481-490.

SAKAI, T., NINUMA, Y., YANAGIHARA, S. & USHIO, K. 1983. Simultaneous determination of hippuric acid and o-, m-and p-methylhippuric acids in urine by high-performance liquid chromatogaphy. Journal of Chromatography B: *Biomedical Sciences and Applications*, 276, 182-188.

SAKUMA, T. 1991. Alteration of urinary carnitine profile induced by benzoate administration. *Arch. Dis. Child*., 66, 873-875.

SAMUELS, D. C., HAN, L., LI, J., QUANGHU, S., CLARK, T. A., SHYR, Y. & GUO, Y. 2013. Finding the lost treasures in exome sequencing data. *Trends in Genetics*, 29, 593-599.

SARKADI, B., HOMOLYA, L., SZAKA, G. & VA`RADI, A. 2006. Human Multidrug Resistance ABCB and ABCG Transporters: Participation in a Chemoimmunity Defense System. *Physiol. Rev*. 86, 1179–1236.

SARKISSIAN, C. N., SCRIVER, C. R. & MAMER, O. A. 2007. Quantitation of phenylalanine and its trans-cinnamic, benzoic and hippuric acid metabolites in biological fluids in a single GC-MS analysis. *Journal of mass spectrometry*, 42, 811-817.

SHENDURE, J., PORRECA, G.J., REPPAS, N.B., LIN, X., MCCUTCHEON, J.P., ROSENBAUM, A.M., WANG, M.D., ZHANG, K., MITRA, R.D., CHURCH, G.M. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. S*cience*, 309(5741):1728-1732.

SMITH, R.L. and WILLIAMS, R.T. 1974. Comparative metabolism of drugs in man and monkeys. *J Med Primatol,* 3:138–152.

SPIEKERKOETTER, U. & WOOD, P. A. 2010. Mitochondrial fatty acid oxidation disorders: pathophysiological studies in mouse models. *Journal of inherited metabolic disease*, 33, 539-546.

STEINBERG, S. J., S. J. WANG, D. G. KIM, S. J. MIHALIK, and P. A. WATKINS. 1999. Human very-long-chain acyl -CoA synthetase: cloning, topography, and relevance to branched-chain fatty acid metabolism. *Biochem. Biophys. Res. Commun*. 257: 615–621.

STEVENTON, G. B., & HUTT, A. J. 2002. The amino acid conjugations. In C. Ioannides (Ed.), Handbook of enzyme systems that metabolise drugs and other xenobiotics. *John Wiley and Sons*, 501–520.

TENNESSEN, J. A., BIGHAM, A. W., O'CONNOR, T. D., FU, W., KENNY, E. E., GRAVEL, S., MCGEE, S., DO, R., LIU, X. & JUN, G. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. science, 337, 64-69.

TFOUNI, S. A. & TOLEDO, M. C. 2002. Estimates of the mean per capita daily intake of benzoic and sorbic acids in Brazil. *Food Addit Contam*, 19, 647-54.

TOMOKUNI, K. & OGATA, M. 1972. Direct colorimetric determination of hippuric acid in urine. *Clinical chemistry,* 18, 349-351.

TSAO, R. 2010. Chemistry and biochemistry of dietary polyphenols. *Nutrients*, 2, 1231-46.

URE, A. 1841. On gouty concretions; with a new method of treatment. *Pharm J Trans*, 1, 24.

VAN DER SLUIS, R., BADENHORST, C. P., ERASMUS, E., VAN DYK, E., VAN DER WESTHUIZEN, F. H. & VAN DIJK, A. A. 2015. Conservation of the coding regions of the glycine N-acyltransferase gene further suggests that glycine conjugation is an essential detoxification pathway. *Gene*, 571, 126-134.

VESSEY, D. A. & LAU, E. P. 1998. Determination of the Sequence of the Arylacetyl Acyl-Acyl CoA: Amino Acid N-Acyltransferase from Bovine Liver Mitochondria and Its Homology to the Aralkyl Acyl-Acyl CoA: Amino Acid N-Acyltransferase. *J Biochem Mol Toxic*, 12, 275-279.

VESSEY, D. A. 1997. Enzymes involved in the formation of amide bonds. Biotransformation, *Elsevier Science*, Oxford, 3. 455-475.

VESSEY, D. A. and Hu, J. 1995. Isolation from bovine liver mitochondria and characterization of three distinct carboxylic acid :CoA ligases with activity toward xenobiotics. *J. Biochem. Toxicol*. 10, 329-337.

VESSEY, D. A. and KELLEY, M. 1997. Purification and partial sequencing of the XL-I form of xenobiotic-metabolizing medium- chain fatty acid: CoA ligase from bovine liver mitochondria, and its homology with the essential hypertension protein. Biochim. *Biophys. Acta* 1346, 231-236.

VESSEY, D. A. and KELLEY, M. 1998. Characterization of the monovalent and divalent cation requirements for the xenobiotic carboxylic acid: CoA ligases of bovine liver mitochondria. *Bochiim. Biophys. Acta* 1382, 243-248.

VESSEY, D. A., HU, J. and KELLEY, M. 1996. Interaction of salicylate and ibuprofen with the carboxylic acid: CoA ligases from bovine liver mitochondria. *J. Biochem. Toxicol*. 11, 73-78.

VESSEY, D. A., KELLEY, M. & WARREN, R. S. 1999. Characterization of the CoA ligases of human liver mitochondria catalyzing the activation of short- and medium-chain fatty acids and xenobiotic carboxylic acids. *Biochimica et Biophysica Acta*, 1428, 455-462.

VESSEY, D. A., KELLEY, M., LAU, E. & ZHANG, S. Z. 2000. Monovalent cation effects on the activity of the xenobiotic/medium-chain fatty acid: CoA ligases are substrate specific. Journal of biochemical and molecular toxicology, 14, 162-168.

VESSEY, D. A., LAU, E., KELLEY, M. & WARREN, R. S. 2003. Isolation, Sequencing, and Expression of a cDNA for the HXM-A Form of Xenobiotic/Medium-Chain Fatty Acid:CoA Ligase from Human Liver Mitochondria. *J Biochem Molecular Toxicology*, 17, 1-6.

VESSEY, D.A., HU, J. 1995. Isolation from bovine liver mitochondria and characterization of three distinct carboxylic acid: coA ligases with activity toward xenobiotics. *J Biochem Toxicol*, 10(6):329-37.

VESSEY, D.A., KELLEY, M., WARREN, R.S. 1999. Characterization of the CoA ligases of human liver mitochondria catalyzing the activation of short- and medium-chain fatty acids and xenobiotic carboxylic acids. *Biochim Biophys Acta*, 1428:455-62.

VESSEY, D.A., LAU, E., KELLEY, M., WARREN, R.S. 2003. Isolation, Sequencing, and Expression of a cDNA for the HXM-A Form of Xenobiotic/Medium- Chain Fatty Acid:coA Ligase from Human Liver Mitochondria. *J Biochem Mol Toxicol*, 17(1):1-6.

VOELKERDING, K.V., DAMES, S.A., DURTSCHI, J.D. 2009. Next-generation sequencing: From basic research to diagnostics. *Clinical chemistry*, 55(4):641-658.

WATKINS, P. A., MAIGUEL, D., JIA, Z. & PEVSNER, J. 2007. Evidence for 26 distinct acyl-acyl coenzyme A synthetase genes in the human genome. *J Lipid Res*, 48, 2736-50.

WHITE, A. 1941. Growth-Inhibition Produced in Rats by the Oral Administration of Sodium Benzoate: Effects of Various Dietary Supplements. *Yale J Biol Med*, 13, 759-68.

WIKOFF, W. R., ANFORA, A. T., LIU, J., SCHULTZ, P. G., LESLEY, S. A., PETERS, E. C. & SIUZDAK, G. 2009. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc Natl Acad Sci U S A*, 106, 3698-703.

WILLIAMS, H. R. T., COX, I. J., D.G., W., COBBOLD, J. F. L., TAYLOR-ROBINSON, S. D., MARSHALL, S. E. & ORCHARD, T. R. 2010. Differences in gut microbial metabolism are responsible for reduced hippurate synthesis in Chron's disease. *BMC Gastroenterology*, 10(108).

WILLIAMS, R. E. & LOCK, E. A. 2005. Sodium benzoate attenuates D-serine induced nephrotoxicity in the rat. *Toxicology*, 207, 35-48.

WILLIAMS, R. T., 1978. Species variations in the pathways of drug metabolism. *Environ Health Perspect*, 22: 133-138.

WONG, L.C. 2013. Next- generation molecular diagnosis of mitochondrial disorders. *Mitochondrion*, 13(4):379-387.

XU, B., IONITA-LAZA, I., ROOS, J. L., BOONE, B., WOODRICK, S., SUN, Y., LEVY, S., GOGOS, J. A. & KARAYIORGOU, M. 2012. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat Genet*, 44, 1365-9.

XU, B., WOODROFFE, A., RODRIGUEZ-MURILLO, L., ROOS, J. L., VAN RENSBURG, E. J., ABECASIS, G. R., GOGOS, J. A. & KARAYIORGOU, M. 2009. Elucidating the genetic architecture of familial schizophrenia using rare copy number variant and linkage scans. *Proc Natl Acad Sci U S A*, 106, 16746-51.

XU, C., LI, C. Y.-T. & KONG, A.-N. T. 2005. Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Archives of pharmacal research*, 28, 249-268.

ZHANG, Y., CHOHNAN, S., VIRGA, K. G., STEVENS, R. D., ILKAYEVA, O. R., WENNER, B. R., BAIN, J. R., NEWGARD, C. B., LEE, R. E., ROCK, C. O. & JACKOWSKI, S. 2007. Chemical Knockout of Pantothenate Kinase Reveals the Metabolic and Genetic Program Responsible for Hepatic Coenzyme A Homeostasis. *Chemistry & Biology*, 14(3):291-302.