The establishment of a model system in which variants of the medium-chain ligase (*ACSM*) and glycine *N*-acyltransferase (*GLYAT*) can be co-expressed and analysed

P Lourens

orcid.org/0000-0002-0820-0299

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

Supervisor:	Prof R van der Sluis
Co-supervisor:	Prof JZ Lindeque

Graduation October 2022

27131157

Acknowledgements

Foremost, I would like to thank and express my sincere gratitude to my supervisor Prof. R. van der Sluis for the continuous support, advice, and patience throughout my research. Furthermore, I would like to thank her for her immense support during the writing of my dissertation.

In addition to my supervisor, I would like to thank and express my sincere gratitude to my cosupervisor Prof. J.Z. Lindeque for supporting me, aiding in my experimental difficulties with the GC–MS and contributing to the writing of my dissertation.

Most importantly, I would like to thank my family, especially my mother Alet Lourens, late father Petrus Lourens, sister Chantél Lourens, and stepfather Petrus van der Merwe, for their continuous support, patience, motivation and listening throughout my academic career.

Last, I would like to especially thank Mr. Strauss van Graan for his continuous support and laboratory advice throughout my experimental procedures and obstacles. Moreover, I would like to thank my laboratory partners for their advice and listening.

I would also like to thank the North-West University for the bursaries allocated to me in pursuit of my degree and the NRF fund for wholly supporting this study.

Abstract

Even though glycine conjugation pathway discovery dates back to the 18th century, it is still an important detoxification pathway that is severely neglected in research. Xenobiotics are omnipresent in the environment (as pollutants) and are mainly used as food preservatives (e.g., sodium benzoate). The biphasic glycine detoxification pathway is important for detoxifying these xenobiotics by conjugation to glycine and then excretion into the urine. The first phase of the glycine conjugation pathway activates the xenobiotic to form a xenobiotic-CoA molecule in the presence of ATP, whereas the second phase conjugates glycine to the activated xenobiotic-CoA and releases CoA as a result. The hydrophilicity of xenobiotics is increased by conjugation with glycine to be excreted in the urine. Phase I activation of the glycine conjugation pathway is performed by ACSM2B (medium-chain fatty acid ligase, E.C. 2.6.1.2), and phase II conjugation is performed by GLYAT (glycine *N*-acyltransferase, E.C. 2.3.1.13). An additional feature of the glycine conjugation pathway is specifically activating medium-chain fatty acids to be used in β-oxidation as an additional energy-producing pathway. However, other substrates, such as organic acids (e.g., isovaleryl-CoA), could accumulate due to metabolic defects. This accumulation of organic acids may also enter the second phase of the glycine conjugation pathway to be conjugated and excreted.

Previous studies on ACSM2B or GLYAT only focused on one variant, and some studies have already indicated detrimental effects in animal models and associated increased ingestion of xenobiotics with ADHD in children. These studies, however, did not focus on defective enzymes of the glycine conjugation pathway; therefore, research on the effect of genetic variation in the enzymes of the glycine conjugation pathway is still lacking. An expression model in which variants of *ACSM2B* and *GLYAT* can be expressed is, therefore, still a necessity in research of the glycine conjugation pathway. This establishment will also contribute to generating an enzymatic model of the glycine conjugation pathway.

This study attempted to establish an expression model in which variants of *ACSM2B*, *GLYAT*, and *GLYATL1* (one of the GLYAT paralogues) can be expressed and analysed. Two expression systems were analysed. The first was the coexpression system using the pETDuet-1_ACSM2B/GLYAT construct, and the second was the expression of pHis17_GLYAT and pET23a(+)_GLYATL1. Both systems used the OverExpress™ C41(DE3)pLysS expression host to express the highest haplotype of each gene. In the first expression system, soluble GLYAT expression was observed, and in the second expression system, soluble GLYAT and GLYATL1 expression were observed. ACSM2B could not be expressed, possibly due to its toxic effect on the expression host. The enzymes in the second

expression system were purified, concentrated, and stored in Tris-HCl (pH 8.0) buffer after optimisation of the expression conditions.

The matrices were analysed for the soluble lysate and purified, concentrated GLYAT and GLYATL1 before GC–MS analysis. Different enzyme reactions were set up using the C41(DE3)pLysS soluble lysate. The purified, concentrated GLYAT and GLYATL1 could not be used to set up enzyme reactions due to severe peak distortions. The substrates for each reaction varied between benzoyl-CoA, phenylacetyl-CoA, isovaleryl-CoA, glycine, and glutamine. Some of the reactions contained mixtures of the substrates. The reactions were then prepared for GC–MS analysis and run on an established GC–MS system. The GC–MS system measured the decrease (if possible) of the substrates (such as benzoyl-CoA) and the consequent increase of the end-products (such as hippuric acid). The extracted ion chromatograms indicated that the expressed soluble GLYAT (in combination with glycine) could effectively convert benzoyl-CoA to hippuric acid in both expression systems. The first system was only used to detect glycine conjugation to benzoyl-CoA. This is because it was the only xenobiotic substrate available at that point in the experimental phase. The soluble GLYAT in the second system could, however, convert some phenylacetyl-CoA in the presence of glycine, but no conversion was seen of phenylacetyl-CoA by GLYAT in the presence of glutamine. The expressed soluble GLYATL1 was not able to convert benzoyl-CoA. Soluble GLYATL1 was, however, able to successfully convert phenylacetyl-CoA in the presence of either glycine or glutamine. Isovaleryl-CoA and its products could not be measured, but the addition of isovaleryl-CoA to the GLYAT reaction, which contained benzoyl-CoA, indicated an inhibitory effect on the elimination of benzoyl-CoA. The overall results agreed with reported literature that two distinctive acyltransferase enzymes exist to detoxify benzoate (GLYAT + glycine) and phenylacetyl-CoA (GLYATL1 + glycine/glutamine). Furthermore, the results indicated that two successful expression systems were set up for GLYAT and one for GLYATL1, in which variants of these genes may be expressed and analysed.

Further characterisation of all the enzymes involved in the glycine conjugation pathway and their variants still needs to be determined. This will contribute to the effect of genetic variation on the glycine conjugation pathway and lead to a more characterised glycine conjugation pathway to establish an enzymatic model of this biphasic pathway. The more characterised glycine conjugation becomes a better understanding of the glycine conjugation pathway and its role in human health. Furthermore, it will lead to better treatments for individuals who suffer from metabolic defects (such as isovaleric acidemia) or disorders (such as ADHD in children).

Keywords: Glycine conjugation, medium-chain fatty acid: CoA ligase (ACSM), glycine *N*-acyltransferase (GLYAT), glycine *N*-acyltransferase like 1 (GLYATL1), GC–MS, benzoyl-CoA, phenylacetyl-CoA, hippuric acid, detoxification, xenobiotics

Table of Contents

Abstract.....
List of abbreviations.....
List of figures.....
List of tables.....
Dissertation outline.....
Chapter 1: Literature review.....
Chapter 2: Establishment of a co-expression model for ACSM2B and GLYAT.....
Chapter 3: Establishment of an expression system to express soluble GLYAT and GLYATL1.....
Chapter 4: Analysis of the fluctuation of selected metabolites detoxified by the glycine conjugation pathway using an established GC-MS method.....
Chapter 5: Conclusion, future prospects, study limitations, and study outcomes.....
References and appendices.....
Chapter 1: Literature review.....1
1.1 Introduction.....1
1.2 Phase I reactions of the glycine conjugation pathway.....5
1.2.1 Genetic and enzyme expression information of ACSMs.....6
1.2.2 Enzyme activity of ACSMs.....6
1.3 Phase II reaction of the glycine conjugation pathway.....7
1.3.1 Genetic information of GLYAT.....7
1.3.2 Enzyme activity of GLYAT.....8
1.3.3 Genetic information about GLYATL1.....13
1.3.4 Enzyme activity of GLYATL1.....14
1.4 Metabolite changes that might occur if the enzymes in Phases I and II are not effective.....14
1.4.1 Isovaleric acidaemia and the contribution to the detoxification load of the Phase II glycine conjugation pathway.....17
1.5 Different types of expression models that can be used.....19
1.5.1 Prokaryotic models.....19
1.5.2 Eukaryotic models.....23
1.5.3 Viral models.....25
1.5.4 Choosing the most applicable expression model.....26
1.5.5 Cell-free models.....26
1.6 Techniques that can be used for enzyme activity evaluation.....27
1.6.1 Enzyme assays.....28
1.6.2 Mass spectrometry.....31
1.7 Problem statement.....34

1.8 Research aim and objectives 36
1.8.1 Aim 36
1.8.2 Objectives 36
1.9 Ethical considerations 37
Chapter 2: Establishment of a co-expression model for ACSM2B and GLYAT 38
2.1 Introduction 38
Simplified workflow of Chapter 2 40
2.2 Materials and methods 41
2.2.1 Materials 41
2.2.2 Methods 44
2.3 Results and discussion of the pETDuet-1 expression system 63
2.3.1 Confirmation of the sizes of the vectors used in this study 63
2.3.2 Sequencing of the MCS-1 and MCS-2 of the pETDuet-1 vector 64
2.3.3 pETDuet-1 vector PCR amplification 65
2.3.4 PCR temperature optimisation for the amplification of *ACSM2B* and *GLYAT* 66
2.3.5 Creating the pETDuet-1_ACSM2B/GLYAT vector using InFusion® 68
2.3.6 Analysis of the soluble expression of ACSM2B and GLYAT 74
2.4 Summary 79
Chapter 3: Establishment of an expression system to express soluble GLYAT and GLYATL1 80
3.1 Introduction 80
Simplified workflow of Chapter 3 81
3.2 Materials and methods 82
3.2.1 Materials 82
3.2.2 Methods 83
3.3 Results and discussion 87
3.3.1 Sequence determination after pHis17_GLYAT and pET23a(+)_GLYATL1 was transformed into C41(DE3)pLysS cells 87
3.3.2 Determining the optimal IPTG concentration to express GLYAT and GLYATL1 88
3.3.3 Determining the optimal induction time to express soluble GLYAT and GLYATL1 90
3.3.4 Purification of GLYAT and GLYATL1 from the cleared lysate of the C41(DE3)pLysS cells 92
3.3.5 Concentration and buffer exchange of the purified GLYAT and GLYATL1 94
3.3.6 Protein quantification results from the Qubit™ 2.0 fluorometer 97
3.4 Section summary 97
Chapter 4: Analysis of the fluctuation of selected metabolites detoxified by the glycine conjugation pathway using an established GC-MS method 99
4.1 Introduction 99

4.2 Materials and methods 100
4.2.1 Materials 100
4.2.2 Methods for the GC-MS analysis of the conversions reactions in the soluble lysate containing either GLYAT or GLYATL1 101
4.2.3 Methods for the GC-MS analysis of the conversion reactions using the concentrated GLYAT or GLYATL1 107
4.3 Results and discussion of the GC-MS analysis of the enzyme reactions using the soluble lysate 108
4.3.1 GC-MS analysis of the pure substrate and some products to determine retention times 108
4.3.2 Results of GC-MS analysis of GLYAT from the pETDuet-1 expression system cell lysate using the 105 m/z ion for metabolite change detection 110
4.3.3 GC-MS analyses of expressed GLYAT using the pHis17 expression system 112
4.3.4 GC-MS analysis of the biotransformation of phenylacetyl-CoA by GLYAT and GLYATL1 in the soluble lysate 115
4.4 The results and discussion of the GC-MS analysis of enzyme reactions containing the purified, concentrated GLYAT or GLYATL1 120
4.4 Section summary 121
Chapter 5: Conclusion and future prospects 124
5.1 Introduction 124
5.2 Conclusion 125
5.3 Future prospects 127
5.4 Study limitations 129
5.5 Study output 129
Appendices 130
Appendix A – Reagents, equipment and software used in this study 130
Appendix B – Quick reference of the conversion of µM to ppm 135
References 136

List of abbreviations

x g – Gravitational force

α – Alpha

β – Beta

% - Percent

°C – Degrees Celsius

µL – microlitres

µM – micromolar

22-G – 22 Gauge

6 x His tag – Histidine tag

A. tumefaciens – *Agrobacterium tumefaciens*

A_{230} – Absorbance at 230 nm wavelength

A_{260} – Absorbance at 260 nm wavelength

A_{280} – Absorbance at 280 nm wavelength

A_{600} – Absorbance (at 600 nm)

ACS – Acyl-CoA synthetases

ACSM – Medium-chain ligases

ADHD – Attention deficit hyperactivity disorder

AEC – 3-amino-9-ethylcarbazole

AMP – adenosine monophosphate

AmpR – Ampicillin resistance gene

amu – Atomic mass unit

ATP – adenosine triphosphate

B. subtilis – *Bacillus subtilis*

bp – Base pairs

BR – Broad range

BRENDA – Braunschweig Enzyme Database

BSA – Bovine serum albumin

CAF – Central Analytic Facility

CASTOR disorders – Disorders associated with CoA

CFPS – Cell-free protein system

cm – Centimetre

Co-A/Co-ASH – Coenzyme

Da – Dalton

ddH_2O – Double distilled water (Milli-Q water)

DMSO – Dimethyl sulfoxide

DNA – deoxyribonucleic acids

DTNB – 5,5-dithio-bis-(2-nitrobenzoic acid)

E. coli – *Escherichia coli*

e.g., – Exempli Gratia

EDTA – Ethylenediaminetetraacetic acid

ELISA – Enzyme-linked immunosorbent assay

EI^+ - Electron impact

EIC – Extracted ion chromatogram

ESI – Electron spray ionisation

EtBr – ethidium bromide

EtBr – ethidium bromide

eV – Electron volts

F1 ori – F1 origin of replication

For. – Forward

g – grams

GC – Gas chromatography

GC – Gas chromatography

GLYAT – Glycine *N*-acyltransferase

GLYAT-L1 – GLYAT-like 1

GLYAT-L2 – GLYAT-like 2

HF – High fidelity

HPLC–ESI–MS/MS – High-performance LC-Electron Spray Ionisation-MS/MS

HRP – horseradish peroxidase

HT TOFMS – High-throughput Time-of-flight MS

HXM-A – Former name of ACSM2B

IPTG – Isopropyl β-D-1-thiogalactopyranoside

Kb – Kilobases

K_{cat} – Catalytic activity

KCl – Potassium chloride

kDa – Kilodalton

KH_2PO_4 – Potassium phosphate

K_i – Inhibition constant

K_m – Michaelis constant

K_m^{app} - Apparent K_m for the reaction

kU – Kilounits

L – Litres

LacI – Regulator gene of the lac operon

LB – Luria-Bertani

LC – Liquid chromatography

LCFA – Long-chain fatty acid

M – Molar

MCAD – Medium-chain acyl-CoA dehydrogenase deficiency

MCFA – Medium-chain fatty acid

MCS – Multiple cloning site

mg – Milligrams

min – minutes

mL – Millilitres

mM – Millimolar

mM – Millimolar

mol – Moles

MOX – Methyloxiamine hydrochloride

MRM – Multiple reaction monitoring

mRNA – messenger ribonucleic acids

MS – Mass spectrometry

MWCO – Molecular weight cut off

Na_2HPO_4 – Sodium phosphate

NaCl – sodium chloride

nd – Not detectable

ng – Nanogram

Ni^{2+} - nickel(II) ion

NIST – National Institute of Standards and Technology

nm – Nanometres

nmol – nanomoles

OA – Organic acid

OD_{600} – Optical density at 600 nm wavelength

ori – Origin of replication

P. aeruginosa – *Pseudomonas aeruginosa*

PAGE – Polyacrylamide gel electrophoresis

PCR – Polymerase chain reaction

PEP – Phosphoenolpyruvate

pmol – Picomoles

PPE – Personal protective equipment

PP_i – Inorganic pyrophosphate

R. leguminosarum – *Rhizobium leguminosarum*

RBS – Ribosome binding site

Rev. – Reverse

rop – Repressor of primer

Rpm – Rotations per minute

RT – Retention time

SASBMB – South African Society for Molecular Biology and Biochemistry

SCFA – Short-chain fatty acid

SDS – Sodium dodecyl sulfate

sec – seconds

SOC – Super optimal broth with catabolite repression

SOP – Standard operating procedure

Sp. – Species

S-tag – Purification tag

T7RNAP – T7 RNA polymerase

TAE – Tris-acetate EDTA

TBS – Tris-buffered saline

TEMED – Tetramethylethylenediamine

TGS – Tris, glycine and SDS buffer

T_m – Melting temperature

Trx-tag – thioredoxin tag

TSS – Transformation and storage buffer

U – Enzyme unit

V – Volts

v/v – Volume per volume

VLCFA – Very long-chain fatty acid

V_{max} – Maximum velocity of the reaction taking place

w/v – Weight per volume

WGE – Wheat germ extract

List of figures

Figure 1: Depiction of the catabolism of dietary polyphenols to produce benzoate and hippurate (from Badenhorst *et al*., 2014). 2
Figure 2: Depiction of the two phases of the glycine conjugation pathway detoxifying a xenobiotic. 4
Figure 3: Consequences of xenobiotic administration in a hyperglycinemic individual. 15
Figure 4: Consequences of a defective GLYAT in a hyperglycinemic individual. 16
Figure 5: Depiction of defective leucine metabolism. 17
Figure 6: Consequences of isovaleryl-CoA conjugation and hydrolysation reactions due to accumulating isovaleryl-CoA. 18
Figure 7: A simplified depiction of plasmid-mediated recombinant protein expression. 20
Figure 8: Concept of a bacterial cell surviving when expressing proteins of interest. 21
Figure 9: Depiction of the process of viral-plant infection to express recombinant proteins (from Gleba *et al*., 2007). 25
Figure 10: Adapted simplification of the cell-free expression system (from Rosenblum & Cooperman, 2014). 27
Figure 11: Simplified activity states of an allosteric enzyme. 30
Figure 12: Indicates the global consumption rates of benzoic acid/benzoate in 2017 (from IHSMarkit (2017)). 36
Figure 13: Principle of InFusion® cloning reactions. 39
Figure 14: Simplified workflow for the methods of Chapter 2. 40
Figure 15: Plasmid map of the pETDuet-1 vector that was used for InFusion® cloning of *ACSM2B* and *GLYAT*. 42
Figure 16: The pHis17 vector containing *ACSM2B*. 43
Figure 17: The pHis17 vector containing *GLYAT*. 43
Figure 18: A simplified depiction of both InFusion® reactions to insert *ACSM2B* then *GLYAT* into the pETDuet-1 vector. 50
Figure 19: Simplified workflow for the experimental procedure in sections 2.2.2.10-2.2.2.13. 59
Figure 20: Image of the agarose gel with all plasmid vectors used in the pETDuet-1 system. 64
Figure 21: MCS-1 alignment of the T7 terminator sequence with the pETDuet-1 vector reference. 65
Figure 22: MCS-2 alignment of the T7 terminator sequence with the pETDuet-1 vector reference. 65
Figure 23: Amplification of the linearised pETDuet-1 vector. 66

Figure 24: Temperature optimisation for ACSM2B and GLYAT. 67
Figure 25: Amplification of *ACSM2B* from the pETDuet-1_ACSM2B vector. 69
Figure 26: Sequence alignments of *ACSM2B* from the pETDuet-1_ACSM2B vector. 70
Figure 27: Image of the agarose gel that confirms the presence of both *ACSM2B* and *GLYAT* genes. 71
Figure 28: Sequence alignments of *GLYAT* from the four sequencing primers used. 72
Figure 29: Sequence alignments of *ACSM2B* from the four sequencing primers used. 73
Figure 30: Image of the Coomassie-stained gels to determine the presence of ACSM2B and GLYAT. 75
Figure 31: Images of the western blot confirmation of the presence of ACSM2B and GLYAT. 76
Figure 32: SDS-PAGE and western blot analysis of no *ACSM2B* expression. 78
Figure 33: Simplified workflow of methods in Chapter 3. 81
Figure 34: The pET23a(+)_GLYATL1 vector with *GLYATL1*. 82
Figure 35: *GLYATL1* alignment to reference sequence. 87
Figure 36: *GLYAT* alignment to the reference sequence. 87
Figure 37: IPTG gradient for GLYAT expression. 89
Figure 38: IPTG gradient GLYATL1. 90
Figure 39: Time gradient of GLYAT expression. 91
Figure 40: Time gradient of GLYATL1. 92
Figure 41: The nitrocellulose membrane of the purification of GLYAT from the C41(DE3)pLysS after induction at the optimal expression conditions. 93
Figure 42: The Coomassie-stained gel and the nitrocellulose membrane of the purification of GLYATL1 from the C41(DE3)pLysS after induction at the optimal expression conditions. 94
Figure 43: Concentration of GLYAT from the purified first elution fraction. 96
Figure 44: Concentration of GLYATL1 from the first elution fraction of the purification. 96
Figure 45: Reaction set-up of the soluble lysate or concentrated proteins with the xenobiotic-CoA, OA-CoA, and amino acid substrates. 100
Figure 46: EIC of benzoic acid (hydrolysed form of benzoyl-CoA). 108
Figure 47: EIC of isovalerylglycine. 109
Figure 48: EIC of hippuric acid (benzoylglycine). 109
Figure 49: EIC indicating multiple peaks due to metabolites in the cleared lysate. 110
Figure 50: GC-MS results of benzoyl-CoA in the pETDuet-1 expression system. 111
Figure 51: GLYAT and only Benzoyl-CoA added in the pHis17_GLYAT expression system. 113
Figure 52: GLYAT with Benzoyl-CoA added and Isovaleryl-CoA added in the pHis17_GLYAT expression system. 115

Figure 53: Phenylacetyl-Co biotransformation with GLYAT and glycine or glutamine. 116
Figure 54: Phenylacetyl-CoA in GLYATL1 with glutamine and presence of benzoyl-CoA. 117
Figure 55:No hippuric acid formation in lysate with GLYATL1. 118
Figure 56: EIC for glycine concentrations in reactions with phenylacetyl-CoA as substrate. 119
Figure 57: Analysis of Tris-HCl (pH 8.0) as good reaction and storage buffer. 121
Figure 58: Conversion of µM to ppm. 135

List of tables

Table 1: Different genetic information of the *ACSM2*. 6
Table 2: Different genetic information of *GLYAT*. 7
Table 3: Different enzyme kinetic values obtained for benzoate and glycine as substrates for GLYAT from different studies (adapted from Rohwer *et al.* (2021)). 9
Table 4: Enzyme kinetics for other preferred GLYAT substrates with Benzoyl-CoA for comparison purposes. 10
Table 5: Amino acid preference for GLYAT (from van der Westhuizen *et al*., 2000). 12
Table 6: Different genetic information of *GLYATL1.* 13
Table 7: Common advantages and disadvantages of protein expression models. 26
Table 8: Advantages and disadvantages of metabolite analysis techniques. 33
Table 9: The primers used for sequencing, and fragment and vector amplification. 51
Table 10: The restriction enzymes and conditions used for linearising the listed vectors. 52
Table 11: Basic PCR setup using the CloneAmp™ HiFi PCR Premix. 52
Table 12: Thermocycler conditions for amplifying the linearised pETDuet-1 vector backbone. 52
Table 13: PCR setup for the amplification of *ACSM2B*. 53
Table 14: Thermocycler conditions for the amplification of *ACSM2B*. 53
Table 15: PCR setup for the amplification of *GLYAT*. 54
Table 16: Thermocycler conditions for the amplification of *GLYAT*. 54
Table 17: Infusion reactions set-up. 55
Table 18: The reaction setup and conditions for linearising the pETDuet-1_ACSM2B vector. 56
Table 19: PCR setup for the amplification of pETDuet-1_ACSM2B vector. 56
Table 20: Thermocycler conditions for the amplification of pETDuet-1_ACSM2B vector. 56
Table 21: The sequence determination primers for *ACSM2B* and *GLYAT*. 58
Table 22: Reagent set-up for casting 1 SDS-PAGE resolving gel. 60
Table 23: Reagent set-up for casting 1 SDS-PAGE stacking gel. 61
Table 24: The primers used for sequencing determination of *GLYAT* and *GLYATL1* from pHis17_GLYAT and pET23a(+)_GLYATL1, respectively. 83
Table 25: Quantification results from the Qubit™ 2.0 Fluorometer. 97
Table 26: The enzyme reaction set up for GLYAT and GLYATL1 with the soluble lysate. . 103
Table 27: The reactions and reaction combinations setup when using the soluble lysate of the induced cells and the soluble lysate of the cells that did not contain GLYAT or GLYATL1. 104
Table 28: Storage and reaction set-up of concentrated GLYAT and GLYATL1 in Tris-HCl (pH 8.0). 107

Table 29: Summary of the results obtained from GC-MS analysis of GLYAT and GLYATL1. 123
Table 30: List of chemicals, reagents and equipment needed for the expression protocols. 130

Dissertation outline

Chapter 1: Literature review

This chapter thoroughly introduces the glycine conjugation pathway to the reader, an overview of different expression models, and enzyme activity measurement techniques. The problem statement, aims, and objectives are included at the end of this chapter.

Chapter 2: Establishment of a coexpression model for ACSM2B and GLYAT

This chapter describes the subcloning method to create the pETDuet-1_ACSM2B/GLYAT construct. It also contains the expression and SDS–PAGE analysis techniques to evaluate the success of the expression of soluble ACSM2B and GLYAT. The results of this chapter were combined with a discussion of the observed results and finished with a chapter summary.

Chapter 3: Establishment of an expression system to express soluble GLYAT and GLYATL1

This chapter describes the techniques used for expressing and optimising expression conditions for GLYAT and GLYATL1. Furthermore, it explains the techniques used for the purification, concentration, and buffer exchange for the soluble expressed GLYAT and GLYATL1. Last, the results were combined with the discussion of the observed results and ended with a chapter summary.

Chapter 4: Analysis of the fluctuation of selected metabolites detoxified by the glycine conjugation pathway using an established GC–MS method

Both matrices were tested on the GC–MS method. These matrices were the soluble lysate and Tris-HCl (pH 8.0). Reactions were then set up with the soluble lysate with different substrates. This chapter explains the technique used for reaction preparation before GC–MS analysis was performed. A general organic acid detection GC–MS method was used to detect the substrates and products of the glycine conjugation pathway and the substrates and products of GLYATL1. The EICs were analysed to detect a decrease in substrates and an increase in products. Last, the results were combined with a discussion of the observed results from the EICs, and it ended with a chapter summary.

Chapter 5: Conclusion, future prospects, study limitations, and study outcomes

This final chapter contains the conclusion of this study and some future prospects. Last, some study limitations are included in this chapter with the study outcomes.

References and appendices

The references used in this study are listed here. Some additional data regarding this study are also presented in the appendices.

Chapter 1: Literature review

1.1 Introduction

The biotransformation of benzoic acid in humans dates back to the 18th century, when it was used as a respiratory stimulant and tonic (Conti & Bickel, 1977). In 1841, Alexander Ure found that a few hours after the ingestion of benzoic acid, hippurate was excreted in his urine (Ure, 1841). After that, William Keller consumed half a drachm of benzoic acid (approximately 1.94 g) in his food, and the findings resulted in a similar conclusion to that of Alexander Ure (Keller, 1842). In 1845, a French chemist established the structure of hippurate by boiling it in inorganic acids, which led to a split into two structural components, benzoic acid and glycine. This led to the elucidation of the detoxification pathway associated with glycine (as reviewed by Conti & Bickel, 1977). At present, the pathway is known to detoxify xenobiotics and utilise medium-chain fatty acids (MCFAs) as a source for β-oxidation to produce ATP (Badenhorst *et al.*, 2014; Goetzman *et al.*, 2020; Gregersen *et al.*, 1986; Houten & Wanders, 2010; Watkins, 1997).

Xenobiotics are molecules/chemicals (containing a carboxylic acid group) to which an individual may be exposed that do not usually form part of that individual's metabolism (Knights *et al.*, 2007). These chemicals are omnipresent in environmental pollutants, anticorrosives (antifreeze), medicine (acetylsalicylate), pesticides, preservatives (such as sodium benzoate), and in some cases, are found in other natural plant and animal products, such as milk, berries and nuts (Beezhold *et al.*, 2014; Del Olmo *et al.*, 2017; Knights & Miners, 2012; Knights *et al.*, 2007; Wibbertmann *et al.*, 2005). The popularity of sodium benzoate in most commercial products is due to its antimicrobial properties (Wibbertmann *et al.*, 2005), but xenobiotics (such as benzoate) can also be, in some cases, used to treat nonketotic hyperglycinemia and urea cycle deficiencies such as hyperammonemia (Lennerz *et al.*, 2015; Matsuo *et al.*, 2012).

Some natural substrates can be introduced into this pathway through the microbial metabolism of dietary polyphenols. Benzoate and benzoyl-CoA serve as substrates for ACSMs and GLYAT, respectively, and can also originate from the catabolism of dietary polyphenols by the human gut microbiome (Figure 1) (Badenhorst *et al.*, 2014; Knights & Miners, 2012; Rechner *et al.*, 2002; van der Sluis, 2018). The catabolism of these polyphenols to simpler aromatic acids (phenylpropionate, cinnamate and benzoate) can be seen in Figure 1, where these aromatic acids are transported and absorbed into the liver to be activated by ATP-dependent ACSMs. Phenylpropionate is activated to form phenylpropionyl-CoA (step 2), which is further converted to cinnamoyl-CoA by medium-chain acyl-CoA dehydrogenase (MCAD) (step 3).

Cinnamoyl-CoA is further oxidised to benzoyl-CoA and acetyl-CoA (step 4). Benzoate (directly catabolized from dietary polyphenols) is also activated to form benzoyl-CoA. Benzoyl-CoA, from both sources, is conjugated with glycine to form hippurate, which is excreted in the urine (step 5). In MCAD deficiency, phenylpropionyl-CoA is converted to phenylpropionylglycine since the conversion of phenylpropionyl-CoA to cinnamoyl-CoA is exclusively performed by MCAD (Flath *et al.*, 1997).

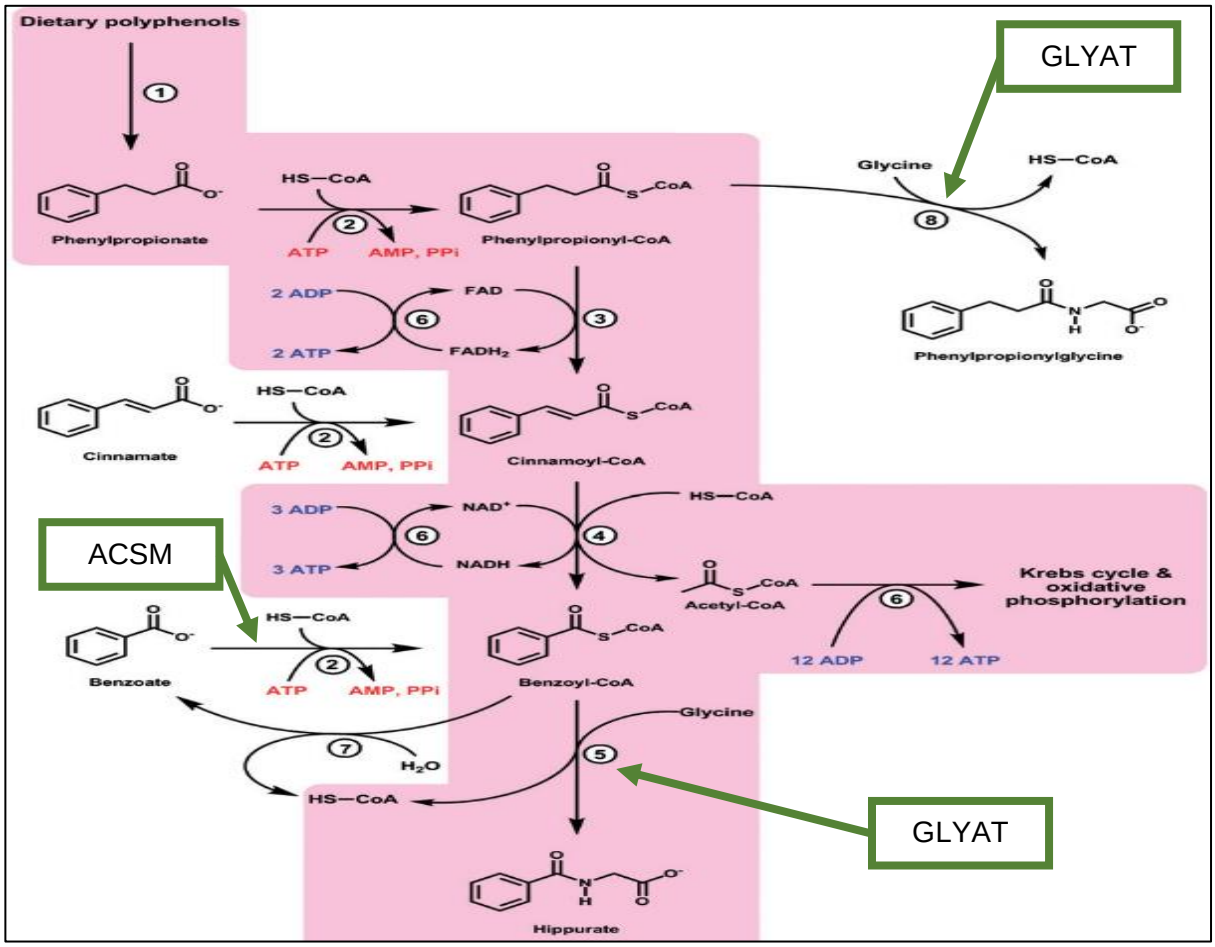


Figure 1: Depiction of the catabolism of dietary polyphenols to produce benzoate and hippurate (from Badenhorst *et al*., 2014).
The applicable enzyme from this study is marked with green rectangles. These were ACSM (Phase I) and GLYAT (Phase II). ATP- Adenosine triphosphate, AMP – Adenosine monophosphate, PP_i – Inorganic phosphate, HS-CoA – Coenzyme A, ADP – Adenosine diphosphate, NAD/NADH – Nicotinamide adenine dinucleotide, FAD/FADH – Flavin adenine dinucleotide.

Apart from the xenobiotic and MCFA substrates entering the glycine conjugation pathway, other substrates (such as organic acid (OA)-CoA molecules) can also be introduced into the glycine conjugation pathway. These molecules are formed in the human body but are usually eliminated through other metabolic pathways. A problem arises when these eliminating enzymes are defective, and OA-CoA molecules accumulate in the human body. An excellent example of a load increasing defect is leucine catabolism, which accumulates isovaleryl-CoA that enters the glycine conjugation pathway (de Sousa *et al.*, 1986; Krieger & Tanaka, 1976;

Tanaka *et al.*, 1966). A graphical depiction of all the mentioned substrate entry points can be seen in Figure 2 (numbers 1, 2 and 3). These are essential alternative substrate origins that should be considered due to their effect on increasing or relieving the substrate load on the glycine conjugation pathway.

Detoxification of xenobiotics that enter the human body from various sources occurs in a biphasic conjugation pathway within the liver or kidneys coupled to an excretory step (Cardenas *et al.*, 2010; Dempsey *et al.*, 2014; Piper & Piper, 2017; Watanabe *et al.*, 2020; Yamamoto *et al.*, 2009). Phase I reactions are essential for increasing the xenobiotic compound's hydrophilicity and creating conjugation sites for Phase II detoxification reactions (Dempsey *et al.*, 2014; Yamamoto *et al.*, 2009). Some of the known products formed in the Phase I reaction in the glycine conjugation pathway are molecules such as acyl-CoA esters, isovaleryl-, 2-methyl-butyryl-, butyryl-, isobutyryl-, hexanoyl-, octanoyl-, decanoyl-CoA (Gregersen *et al.*, 1986), salicyl-CoA and benzoyl-CoA (Lennerz *et al.*, 2015; van der Sluis, 2018). MCFA-CoA molecules are mainly used for ATP production in mitochondrial β-oxidation (Badenhorst *et al.*, 2014; Goetzman *et al.*, 2020; Gregersen *et al.*, 1986; Houten & Wanders, 2010; Watkins, 1997), whereas xenobiotic-CoAs (benzoyl-CoA and salicyl-CoA) are removed by conjugation to glycine. A depiction of the detoxification phases can be seen in Figure 2, where first, the xenobiotic reacts with ATP and CoASH (catalysed by ACSM) to yield a high energy xenobiotic-CoA, pyrophosphate and AMP (Phase I) and last, GLYAT conjugates the mitochondrial xenobiotic-CoA to the *N*-terminal of the amino acid glycine (Phase II) (Badenhorst *et al.*, 2013; Dempsey *et al.*, 2014; Knights & Miners, 2012; Knights *et al.*, 2007; Yamamoto *et al.*, 2009). In Figure 2 (Phase II), it can be seen that the release of CoASH in the Phase II reaction ensures that adequate levels of CoASH are present within the mitochondria (Badenhorst *et al.*, 2013; Knights & Miners, 2012; Knights *et al.*, 2007; Mawal *et al.*, 1997).

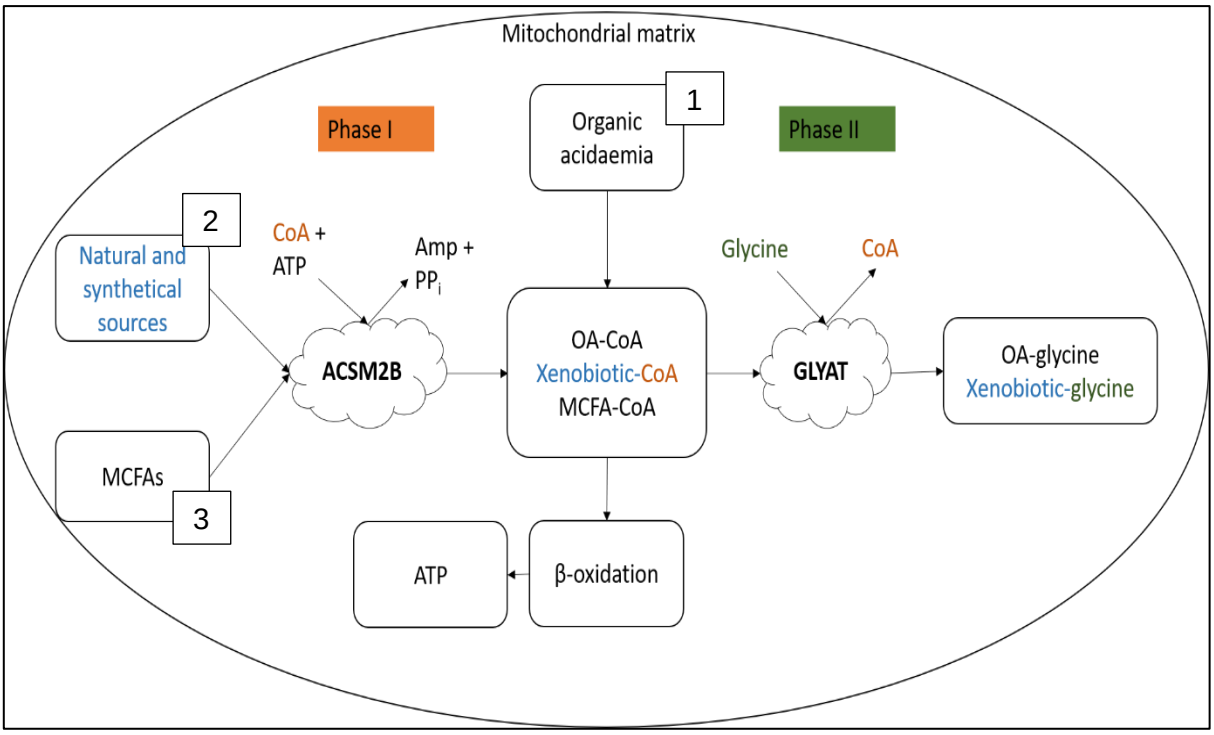


Figure 2: Depiction of the two phases of the glycine conjugation pathway detoxifying a xenobiotic. Numbers 1-3 depict the substrate entry sites into the glycine conjugation pathway. Phase I is marked with an orange rectangle, whereas Phase II is marked with a green rectangle. The substrate, intermediate and product marked in blue, blue-orange and blue-green respectively, depict a common substrate conversion known to be done by the glycine conjugation pathway from benzoate to hippurate. The enzyme co-factors are marked in orange. Glycine is marked with green in Phase II as an amino acid substrate.

The glycine conjugation pathway is crucial in activating medium-chain fatty acids (MCFAs) and removing xenobiotics from multiple species, especially humans. Considering this, it is surprising that both phases of this detoxification pathway are not well characterised (Cardenas *et al.*, 2010). The effectivity of the pathway may be altered when genetic polymorphisms cause alterations in the kinetic parameters of medium-chain ligases (ACSM (E.C. 2.6.1.2)) and/or glycine *N*-acyltransferase (GLYAT (E.C. 2.3.1.13)) (van der Sluis *et al.*, 2013). Previous studies have found that genetic variation in *GLYAT* influences enzyme activity. Although the N156S variant increases enzyme activity (van der Sluis *et al.*, 2013), other studies have linked some metabolic deficiencies with decreased enzyme activities present in the glycine conjugation pathway. For example, when these unfavourable genetic variants occur and express a defective enzyme with decreased enzyme activity or no enzyme activity, it may ultimately lead to numerous metabolic deficiencies. These are metabolic deficiencies such as carcinogenesis, hepatocellular carcinoma, nonketotic hyperglycinemia neoplasms and porphyrias (BRENDA, 2019), CASTOR disorders (disorders associated with CoA) (Mitchell *et al.*, 2008), the accumulation of toxic levels of xenobiotics (Badenhorst *et al.*, 2013), organic acidurias (Gregersen *et al.*, 1986; Knights *et al.*, 2007), Reye's syndrome (Glasgow, 2006; Piper & Piper, 2017), carnitine deficiency (Gregus *et al.*, 1992), hepatic ATP depletion, altered

glucose homeostasis (Lennerz *et al.*, 2015), attention deficit hyperactivity disorder (ADHD) (Beezhold *et al.*, 2014; Kilic *et al.*, 2006), and can be related to various kidney diseases or an alteration in kidney function (Watanabe *et al.*, 2020).

1.2 Phase I reactions of the glycine conjugation pathway

Mitochondrial ACSMs in phase I reactions are responsible for activating xenobiotics to yield high-energy xenobiotic-CoA complexes with AMP and inorganic pyrophosphate as by-products (Watkins *et al.*, 2007). A simple depiction of the activation step can be seen in Figure 2 (Phase I), where the xenobiotic reacts with CoASH to form the respective benzoyl-CoA and enters Phase II. It should be noted that ACSM2B not only activates xenobiotics but also fatty acids to form MCFA-CoA molecules (Figure 2) and that both of these activation reactions (the formation of MCFA-CoA and xenobiotic-CoA) are ATP dependent (Knights *et al.*, 2007).

Fatty acids are ubiquitously expressed in many species and are classified into four main classes: short-chain fatty acids (SCFAs), MCFAs, long-chain fatty acids (LCFAs) and very-long-chain fatty acids (VLCFAs), depending on the length of the carbon chain, which can range from 2 to >30. Furthermore, the structure of these fatty acid chains may be different from one another because some of them may be saturated, monounsaturated, or polyunsaturated or may contain methyl branches at one or more sites (Watkins *et al.*, 2007).

The selectivity of fatty acids to serve as substrates for fatty acid ligase (also known as acyl-CoA synthetase (ACS)) enzymes depends on the fatty acid chain length. The enzyme's name will then be followed by the length of the fatty acid substrate. For example, ACS is the root name for different families of ACS enzymes, followed by “M” if MCFAs are the optimal substrates to be finally known as ACSMs. Thus, the optimal substrate for ACSMs will be MCFAs with carbon lengths of 4 ~ 10. Examples include pentanoic acid (valeric acid), hexanoic acid (caproic acid), heptanoic acid (enanthic acid), octanoic acid (caprylic acid) and, to some extent, decanoic acid (capric acid) (Knights & Miners, 2012; Knights *et al.*, 2007; Watkins *et al.*, 2007). When MCFAs are used as an energy source in the absence of glucose or low levels thereof, they enter the mitochondria, where MCFA-CoA molecules are formed through ACSMs (van der Sluis, 2018). β-Oxidation thus occurs as an alternative energy-producing metabolic pathway for organs such as the liver, heart, skeletal muscle, and kidneys (Houten & Wanders, 2010; Watanabe *et al.*, 2020). It is important to note that MCFAs do not follow the carnitine transport pathway like LCFAs and VLCFAs to enter the mitochondrion but diffuse through the mitochondrial membrane to be activated by ACSM2B within the mitochondrial matrix to undergo β-oxidation (Bremer, 1983; Gillingham *et al.*, 2017; Goetzman *et al.*, 2020; Nakajima *et al.*, 1997; Vessey *et al.*, 1999).

The xenobiotic-CoA and OA-CoA molecules from the phase I reaction do not enter mitochondrial β-oxidation but are further processed in this pathway's phase II reaction.

1.2.1 Genetic and enzyme expression information of ACSMs

Several different isoforms of *ACSM* currently exist, namely, ACSM 1, 2A, 2B, 3, 4, 5 and 6 and are expressed in various tissues at different levels (Boomgaarden *et al.*, 2009). Furthermore, ACSM2A and ACSM2B were thought to have the highest expression levels in the human liver, and the genes are both highly conserved in global populations (van der Sluis, 2018). However, a recent study indicated that the expression of ACSM2 enzymes is localised in the kidneys and, more specifically, the proximal renal tubules (Watanabe *et al.*, 2020). Different genetic characteristics of both enzymes can be seen in Table 1, which were retrieved from Ensembl version 99 (March 2022) (Cunningham *et al.*, 2019). This table only contains genetic information for *ACSM2A* and *ACSM2B* due to the significance in the glycine conjugation pathway.

Enzyme	ACSM2A	ACSM2B
Gene	NM_001308172.2	NM_001105069.2
Exons	14 (13 coding)	14 (13 coding)
Transcript length (bp)	2894	2935
Total variants	13474	16046
Missense variants	590	604
Strand	Forward	Reverse
Location	Chromosome 16p12.3	Chromosome 16p12.3
Nucleotide sequence similarity	98.8%	
Amino acids residues	577	
Amino acid similarity	97.1%	

Table 1: Different genetic information of *ACSM2*.

1.2.2 Enzyme activity of ACSMs

Although the crystalline structure of ACSM2A was determined and suggested that the ligand sites are able to bind MCFAs and ATP (Kochan *et al.*, 2009) and that ACSM2A mainly catalyses indolepropionic acid (Koshiba *et al.*, 2020), there is still no kinetic activity information on ACSM2A, and it has yet to be investigated (van der Sluis, 2018; Watanabe *et al.*, 2020).

Analyses of liver mitochondrial lysates eluded the presence of two distinct ligases that are capable of activating both xenobiotics and MCFAs termed HXM-A (now ACSM2B) and HXM-

B (now ACSM1) at a more basic pH (van der Sluis & Erasmus, 2016; Vessey *et al.*, 1999). ACSM2B was recombinantly expressed in COS-7 cells after being amplified from human cDNA and cloned into a p3xFLAG-CMV-7 vector. The cells that contained the HXM-A/FLAG fusion protein were analysed in the presence of benzoate and phenylacetate. The cells containing the vector indicated activity towards benzoate (0.91 pmol/min/mg) and phenylacetate (0.71 pmol/min/mg) (Vessey *et al.*, 2003).

Since then, no further studies have been conducted to accurately determine the kinetic parameters and efficiency of human ACSMs (Knights & Miners, 2012; van der Sluis, 2018). The products of the enzymatic conversion by ACSM2B are further processed by conjugation to glycine in the second phase of the glycine conjugation pathway.

1.3 Phase II reaction of the glycine conjugation pathway

1.3.1 Genetic information of GLYAT

GLYAT has been successfully isolated from the liver and kidney mitochondria of multiple species, such as ovine, bovine, rhesus monkeys, rabbits and humans (Cardenas *et al.*, 2010; Dempsey *et al.*, 2014; Gregersen *et al.*, 1986; Knights & Miners, 2012; Knights *et al.*, 2007), and is different from the bile acid CoA amino acid *N*-acyltransferases that conjugate bile acids with taurine or glycine (Knights *et al.*, 2007). Additional genetic information for human *GLYAT* (h*GLYAT)* can be seen in Table 2 and was retrieved from Ensembl version 99 (March 2022) (Cunningham *et al.*, 2019).

Table 2: Different genetic information of *GLYAT*.

Enzyme	GLYAT
Gene	ENST00000344743.8 (NM_201648.3)
Exons	6 (5 coding)
Location	Chromosome 11q12.1
Transcript length (bp)	2024
Amino acid residues	296
Total Variants	9200
Missense variant	274
Strand	Reverse strand
Paralogues	*GLYATL1* and *GLYATL2*

By phylogenetically analysing the conservation of *GLYAT* in various species, Rohwer *et al.* (2021) found a clear separation between the clades for hominids and the other species used in their study. The other species were *Mus musculus*, *Rattus norvegicus*, *Pteropus vampyrus*, *Loxodonta africana*, *Bos taurus*, and *Tursiops truncates*. The segregation of the hominid clade and that of the other species indicated minor changes between *GLYAT* during the human and chimpanzee split, further suggesting that it remained relatively conserved across all population groups. The expressed protein is approximately 34 kDa in size (Badenhorst *et al.*, 2013; Cardenas *et al.*, 2010; Dempsey *et al.*, 2014; Knights *et al.*, 2007) and has an affinity for conjugating xenobiotic-CoAs with glycine (Dempsey *et al.*, 2014; Knights & Miners, 2012; Nandi *et al.*, 1979; van der Westhuizen *et al.*, 2000).

Rohwer *et al.* (2021) also analysed allelic variation and haplotype diversity in a worldwide human population. This analysis found that only two variants had the highest allele frequency of >0.5%. These were Asn156Ser (94.87%) and Ser17Thr (19.67%), whereas the rest had allele frequencies between 0.2-0.0004%. Furthermore, their analysis indicated that the Asn156Ser missense variant also had the highest homozygous genotype frequency of 90.3%, followed by Ser17Thr of 4.5% compared to six other homozygotes. These were Arg131His (East Asian-0.005%), Arg131Cys (South Asian-0.002%), Met65Thr (Latino-0.001%), Thr73Ile (East Asian-0.001%), His101Tyr (South Asian-0.002%) and Thr244Met (Latino-0.001%).

It has been indicated that polymorphisms eliciting a negative effect on the enzyme efficiency of GLYAT seldom occur within the worldwide population (van der Sluis *et al.*, 2013). For example, the C_{199} variant of *GLYAT* that only has an activity of 5% compared to the wild type is less frequently encountered. The expression of this variant may lead to adverse health effects if too many xenobiotics are consumed (van der Sluis, 2018).

1.3.2 Enzyme activity of GLYAT

Because xenobiotics have various structures and points of origin, it is essential to consider the effect of genetic variations (polymorphisms) on the enzyme activity, substrate specificity, and expression level of ACSMs and GLYAT since these polymorphisms can affect detoxification efficiency and substrate specificity, which is crucial for efficient xenobiotic detoxification (Cardenas *et al.*, 2010; Knights & Miners, 2012).

Previous studies have reported varied kinetic parameters for GLYAT, which may be explained by discrepancies in the methods used to measure the enzyme activity. The DTNB method, which is commonly used in multiple studies (Bartlett & Gompertz, 1974; Dempsey *et al.*, 2014; Kelley & Vessey, 1994; Matsuo *et al.*, 2012; Nandi *et al.*, 1979; van der Sluis *et al.*, 2013; van

der Sluis *et al.*, 2017), was not effective in measuring amino acid conjugation at low levels (van der Westhuizen *et al.*, 2000). The K_{Mapp} values for benzoyl-CoA ranged from 13 – 57300 µM, whereas the K_{mapp} for glycine ranged from 6.4 – 26.6 mM. The V_{max} values ranged from 543 ± 21 – 17100 nmol/min/mg (Table 3) (Kelley & Vessey, 1994; Matsuo *et al.*, 2012; Mawal & Qureshi, 1994; van der Sluis *et al.*, 2013; van der Westhuizen *et al.*, 2000). The reaction was thought to have been a sequential mechanism. Xenobiotic-CoA first associates with the enzyme and then associates with glycine before releasing CoASH and the glycine conjugated molecule (Nandi *et al.*, 1979). However, this was refuted in a recent study that indicated that a sigmoidal kinetic mechanism is the best model for this reaction mechanism with positive cooperativity (van der Sluis *et al.*, 2017). Rohwer *et al.* (2021) explained this variation in the kinetics parameters and proposed an enzyme mechanism and final kinetic parameters for GLYAT.

Table 3: Different enzyme kinetic values obtained for benzoate and glycine as substrates for GLYAT from different studies (adapted from Rohwer *et al.* (2021)).

Benzoyl-CoA (K_m) (µM)	Glycine (K_m) (mM)	V_{max} (nmol/min/mg protein)	Variant	Purification method	Reference
9	N/A	10400	Unknown	Cellular fractioning	(Bartlett & Gompertz, 1974)
67±5	6.5±1	N/A	Unknown	Cellular fractioning and ion-exchange chromatography	(Kelley & Vessey, 1994)
57900	N/A	17100	Unknown	Cellular fractioning and chromatography	(Mawal & Qureshi, 1994)
13	6.4	543±21	Unknown	Cellular fractioning and chromatography	(van der Westhuizen *et al.*, 2000)
209	26.6	807	Unknown	Nickel affinity chromatography	(Matsuo *et al.*, 2012)
38±4	N/A	1230±60	Recombinant S_{156}	Nickel affinity chromatography	(van der Sluis *et al.*, 2013)
49±13	20±4	157±22	Recombinant S_{156} variant	Nickel affinity chromatography	(van der Sluis *et al.*, 2017)

53±8	21±2	161±15	Recombinant S_{156}	Nickel affinity chromatography	(van der Sluis *et al.*, 2017)
106700 ± 3600	39.9 ± 1.8	951 ± 50.5	Recombinant S_{156}	Nickel affinity chromatography	(Schutte *et al.*, 2019)
97 ± 3	23 ± 2	850 ± 60	Recombinant S_{156}	Nickel affinity chromatography	(Rohwer *et al.*, 2021)
118 ± 7	29 ± 3	620 ± 20	Recombinant $S_{17,}$ S_{156}	Nickel affinity chromatography	
61 ± 3	30 ± 3	830 ± 5	Recombinant S_{156}, S_{199}	Nickel affinity chromatography	

Since benzoyl-CoA can act as a substrate for GLYAT, it is crucial to conduct studies that investigate the enzyme kinetics for other alternative substrates. The results of previously determined enzyme kinetics for GLYAT are summarised in Table 4. The K_m values were converted to µM from mM for ease of comparison. The reaction velocities are low with high K_m values, suggesting a lower affinity of GLYAT towards these substances than benzoyl-CoA, tiglyl-CoA, and isovaleryl-CoA. The enzyme kinetics for GLYAT in the presence of benzoyl-CoA, tiglyl-CoA and isovaleryl-CoA had high reaction velocities and low K_m values, which further suggests that GLYAT has a stronger affinity as well as a strong substrate specificity for benzoyl-CoA (Dempsey *et al.*, 2014; Matsuo *et al.*, 2012), tiglyl-CoA and isovaleryl-CoA as substrates (Bartlett & Gompertz, 1974).

Table 4: Enzyme kinetics for other preferred GLYAT substrates with benzoyl-CoA for comparison purposes.

Substrate	**K_m (µM)**	**Glycine (mM)**	**V_{max} (nmol /min/ mg)**	**Variant**	**Purification method**	**Reference**
Octanoyl-CoA	198000	N/A	3300	Unknown	Cellular fractioning and chromatography	Mawal and Qureshi (1994)
Salicyl-CoA	83700	N/A	10100		Cellular fractioning and chromatography	
Isovaleryl-CoA	12400	N/A	7640		Cellular fractioning and chromatography	
Hexanoyl-CoA	2680±770	1150±210	26.5± 5.8		Cellular fractioning	

Decanoyl-CoA	2408±887	690±210	20.4±3.3		Cellular fractioning	Gregersen *et al.* (1986)
Butyryl-CoA	2400±880	970±210	29.5±3.7		Cellular fractioning	
Isovaleryl-CoA	672±164	523±206	13.4±4.4		Cellular fractioning	
Octanoyl-CoA	322±87	770±110	7.7±0.5		Cellular fractioning	
Acetyl-CoA	210	N/A	2600		Cellular fractioning	Bartlett and Gompertz (1974)
Isovaleryl-CoA	180	N/A	12300		Cellular fractioning	
Propionyl-CoA	180	N/A	4400		Cellular fractioning	
Tiglyl-CoA	110	N/A	33300		Cellular fractioning	
2-Methylbutaryl-CoA	110	N/A	8300		Cellular fractioning	
Benzoyl-CoA	97±3	23±2	850±60	Recombinant S_{156}	Nickel affinity chromatography	Rohwer *et al.* (2021)
3-Methylcrotonyl-CoA	14	N/A	5700	Unknown	Cellular fractioning	Bartlett and Gompertz (1974)

The active site of h*GLYAT* was suggested to be between the amino acid 128 – 178 regions. This is an important site for binding glycine and acyl residue transfer (Matsuo *et al.*, 2012). Other amino acids may also be bound to the xenobiotic-CoA molecules being detoxified. However, it binds to a much lesser extent to form the amino acid conjugate than glycine. These amino acids are glutamine, alanine, glutamic acid, serine and asparagine (Nandi *et al.*, 1979; van der Westhuizen *et al.*, 2000). The amino acid preference of GLYAT is summarised in

Table 5.

Table 5: Amino acid preference for GLYAT (from van der Westhuizen *et al*., 2000).
*nd – could not be determined

Amino acid preference	K_m (Amino acid) μM	Conjugation rate (nmol/min/mg protein)
Glycine	6400	543±21
Alanine	997000	27±4
Glutamic acid	nd	4.8±0.7

Inhibition of GLYAT is inevitable since there exists an extensive range of xenobiotics and OA-CoA molecules. GLYAT was shown to be inhibited by oleoyl-CoA (Dempsey *et al.*, 2014), ethanol (on benzoyl-CoA conjugation) (Levy, 1979), phenylacetyl-CoA, indoleacetyl-CoA (Nandi *et al.*, 1979; Webster *et al.*, 1976), potassium chloride (KCl) (Kelley & Vessey, 1994), CoASH and hippurate (Schachter & Taggart, 1954). Due to the affinity being greater for some xenobiotics, it may inhibit the detoxification and elimination rates of others. Salicylate detoxification and elimination are inhibited when benzoate is present (Amsel & Levy, 1969).

When dismissing the specificity of the enzymes, it is clear that the availability of glycine and CoA are the limiting factors for the second phase reaction to take place (Marsh *et al.*, 2005) since glycine availability becomes the rate-limiting step when benzoate levels exceed 2 g (Amsel & Levy, 1969). CoA is released in Figure 2 (Phase II) when GLYAT conjugates glycine to the xenobiotic. This ensures that CoASH sequestration does not occur and that adequate levels of CoA are available for subsequent reactions with xenobiotics (van der Sluis *et al.*, 2017). CoA is a vital molecule to be maintained in the human body due to its intensive role in mitochondrial β-oxidation and other metabolic processes (Badenhorst *et al.*, 2013).

On the other hand, glycine has important neurotransmitter properties, acts as a cytoprotective agent (Gundersen *et al.*, 2005; López-Corcuera *et al.*, 2009) and is a major component of collagen and elastin (the most abundant proteins in the human body) (Wang *et al.*, 2013). Apart from these functions, glycine plays a vital role in the survival of animals and humans. It has a role in immunity, growth, and development and can be synthesized from serine, threonine, choline, and glyoxylate (Wang *et al.*, 2013). CoA and glycine are two essential molecules involved in maintaining the elimination rate of xenobiotics (Gregus *et al.*, 1992). Orally administered glycine influenced benzoate detoxification rates (Marsh *et al.*, 2005), whereas the administration of pantothenic acid increased the levels of acetyl-CoA, a precursor for CoA (Vadali *et al.*, 2004).

Considering the essential and intensive role that GLYAT plays in the human body, it is crucial to consider the presence of similar enzymes that may aid in detoxifying toxic xenobiotic- or OA-CoA molecules if the load on the glycine conjugation pathway increases. One such enzyme is GLYATL1.

1.3.3 Genetic information about GLYATL1

GLYATL1 was found to have stronger expression in the liver and kidneys after being compared between 18 different human tissues (Zhang *et al.*, 2007). Additional genetic information for human *GLYATL1* was retrieved from Ensembl version 99 (March 2022) (Cunningham *et al.*, 2019) in Table 6.

Table 6: Different genetic information of *GLYATL1.*

Enzyme	GLYATL1
Gene	NM_001389712.2
Exons	5
Location	Chromosome 11q12.1
Transcript length (bp)	909
Amino acid resides	302
Total Variants	4926
Missense variants	282
SNPs	2167
Strand	Forward
Paralogues	*GLYAT* and *GLYATL2*

When the amino acid sequences of 8 species were aligned by Zhang *et al.* (2007), it was observed from their study that human GLYATL1 had a 97.7% similarity to troglodytes, 91.4% similarity to monkeys, 39.6% similarity to cattle, 39.1% similarity to mice, 38.9% similarity to chimpanzees, 38.2% similarity to rats, and 37.4% similarity to dogs, which suggested that the protein remained conserved during evolution. Even though both enzymes GLYAT and GLYATL1 indicated acyltransferase ability (Matsuo *et al.*, 2012; Webster *et al.*, 1976), *GLYATL1* only had a 39.0% similarity to *GLYAT* and 49.0% to *GLYATL2* (Matsuo *et al.*, 2012). Furthermore, the expressed *GLYATL1* yields a 35.1 kDa protein localised in the endoplasmic reticulum as well as the the cytosolic region of the liver and kidney (Matsuo *et al.*, 2012; Zhang *et al.*, 2007).

1.3.4 Enzyme activity of GLYATL1

The difference in enzyme activity is that GLYAT has benzoyl-CoA glycine *N*-acyltransferase abilities, whereas GLYATL1 has phenylacetyl-CoA glutamine *N*-acyltransferase activity. GLYATL2 did not indicate activity under experimental conditions (Matsuo *et al.*, 2012). Matsuo *et al.* (2012) found that glutamine is preferably conjugated to phenylacetyl-CoA by GLYATL1. This confirms the findings of Webster *et al.* (1976). They predicted two possible enzymes in the liver mitochondria that either utilise glycine (in the case of hGLYAT) or glutamine as a primary conjugating enzyme (in the case of GLYATL1). Although the prediction of the two enzymes was accurate, more recent studies indicated that GLYATL1 is not located in the mitochondria, but rather the endoplasmic reticulum and the cytosolic region (Matsuo *et al.*, 2012; Zhang *et al.*, 2007).

Matsuo *et al.* (2012) suggested that GLYAT is primarily an aralkyl transferase, in contrast to GLYATL1, which is considered an arylacetyl transferase. The observation for the preference of glycine as an amino acid acceptor is supported by the results of van der Westhuizen *et al.* (2000) for GLYAT. The K_m value for glutamine could not be determined in that study to support the observation of glutamine preference of GLYAT. Furthermore, the assumption made by Matsuo *et al.* (2012) was that glutamine is mostly the preferred amino acid acceptor for GLYATL1. However, their study was self-contradicting regarding K_m values for GLYAT, and no K_m value data are available for GLYATL1 and phenylacetyl-CoA.

Considering that the glycine conjugation pathway can influence the abundance of specific metabolites, it is essential to know which metabolites are mainly affected and the extent of these complications in the glycine conjugation pathway.

1.4 Metabolite changes that might occur if the enzymes in Phases I and II are not effective

After phase I and II reactions, glycine conjugates are excreted in the urine or bile in individuals with a normal glycine conjugation pathway. The lipophilicity of the xenobiotic is decreased during glycine conjugation. Therefore, the possible reason for hippurate being excreted in the urine is the inability of hippurate to diffuse back into the mitochondrial matrix (Badenhorst *et al.*, 2014).

However, because metabolism is a series of integrated and interdependent substrate-enzyme processes taking place, ACSMs and GLYAT affect not only the activity of a single metabolic pathway but also various others when defective. For instance, when the intake of xenobiotics and MCFAs increases and the efficiency of ACSM2B, GLYAT or both decreases, it may cause substrate imbalances, xenobiotic and OA accumulation, sequestration of some essential

cofactors (such as CoA (Gregus *et al.*, 1992)) (Badenhorst *et al.*, 2013; Dempsey *et al.*, 2014; van der Sluis, 2018; van der Sluis *et al.*, 2017), and a decrease in β-oxidation rates (Figure 3 and Figure 4) (Badenhorst *et al.*, 2014; Gregersen *et al.*, 1986; Houten & Wanders, 2010; Watkins, 1997). The important substrates, cofactors, and intermediates in Figure 3 and Figure 4 are marked with red arrows representing some changes. These changes may consequently affect an individual's overall metabolism to cause various metabolic defects even when a single amino acid substitution was made in *ACSM2B* or *GLYAT* (Cardenas *et al.*, 2010; Yamamoto *et al.*, 2009).

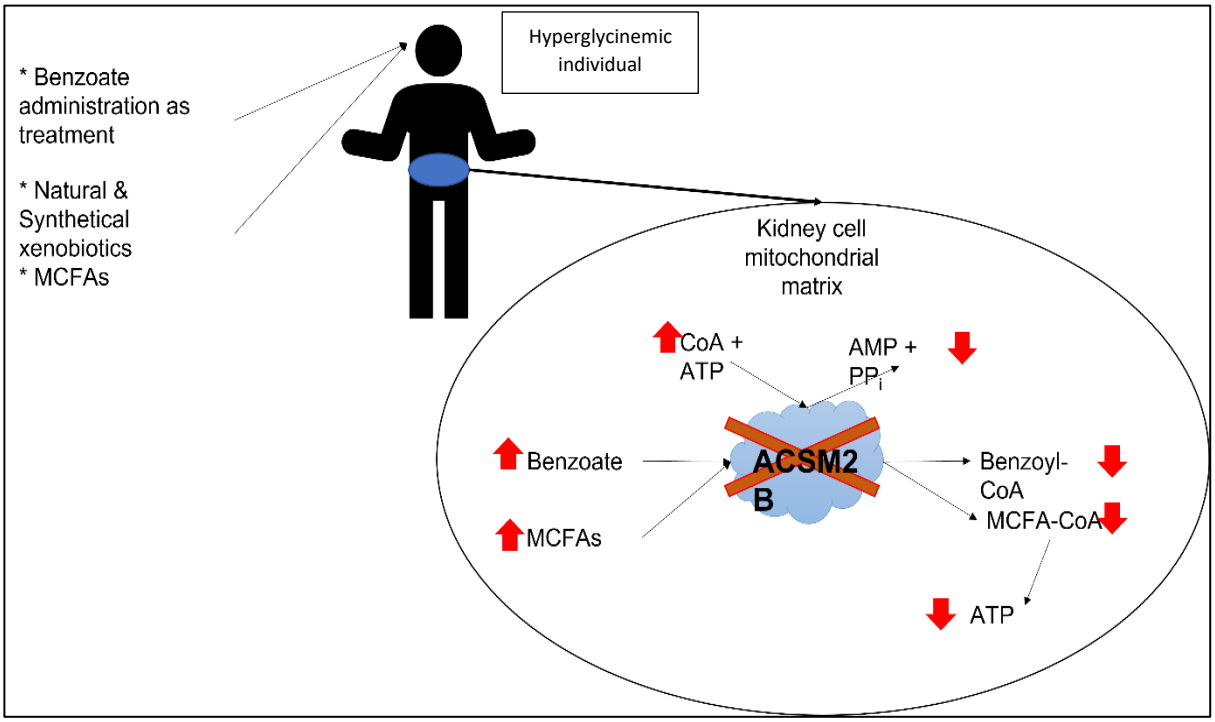


Figure 3: Consequences of xenobiotic administration in a hyperglycinemic individual.
The substrate origins are marked with stars. The red arrows indicate which substrates and products may increase or decrease due to a defective first phase of the glycine conjugation pathway. The defective enzyme for Phase I is ACSM2B, marked with the red cross through it.

Luckily, some metabolic deficiencies can be treated by manipulating detoxification pathways by administering substrates or cofactors of the glycine conjugation pathway (Lennerz *et al.*, 2015; Mitchell *et al.*, 2008; Piper & Piper, 2017). These are deficiencies such as hyperglycinemia and urea cycle disorders. Both are treated with intravenous benzoate administration (Lennerz et al., 2015; Van Hove et al., 2005). For example, in a normal individual, the excretion of glycine conjugates prevents the accumulation of toxic levels of the neurotransmitter glycine, according to the glycine deportation hypothesis (Badenhorst *et al.*, 2014; Beyoglu & Idle, 2012; Beyoglu *et al.*, 2012). However, if an individual suffers from hyperglycinemia and has a defective ACSM2B enzyme, an increase in benzoate will be observed after the external administration of benzoate. Defective ACSM2B can also lead to a

potential rise in free MCFAs since ACSM2B activates it to produce ATP in the β-oxidation pathway. These metabolite alterations can be seen in Figure 3, indicated by the red arrows.

If ACSM2B is not defective but GLYAT fails to conjugate glycine to benzoyl-CoA, the xenobiotics used as a treatment or the cofactors administered may cause more harm than treating the defects. The harm caused is a result of accumulating intermediates (such as benzoyl-CoA and isovaleryl-CoA) and substrates (such as glycine) and sequestration of other cofactors (such as CoA) in this pathway. These alterations can be seen in Figure 4, indicated by red arrows. Furthermore, suppose the defect in GLYAT remains undetected during treatment with benzoate. In that case, it may lead to various other metabolic defects, such as CASTOR disorders that involve CoASH sequestration when xenobiotic-CoA and MCFA-CoA molecules are formed, but further metabolism is stalled or ineffective (Badenhorst *et al.*, 2013; Badenhorst *et al.*, 2014; Knights *et al.*, 2007; Mitchell *et al.*, 2008).

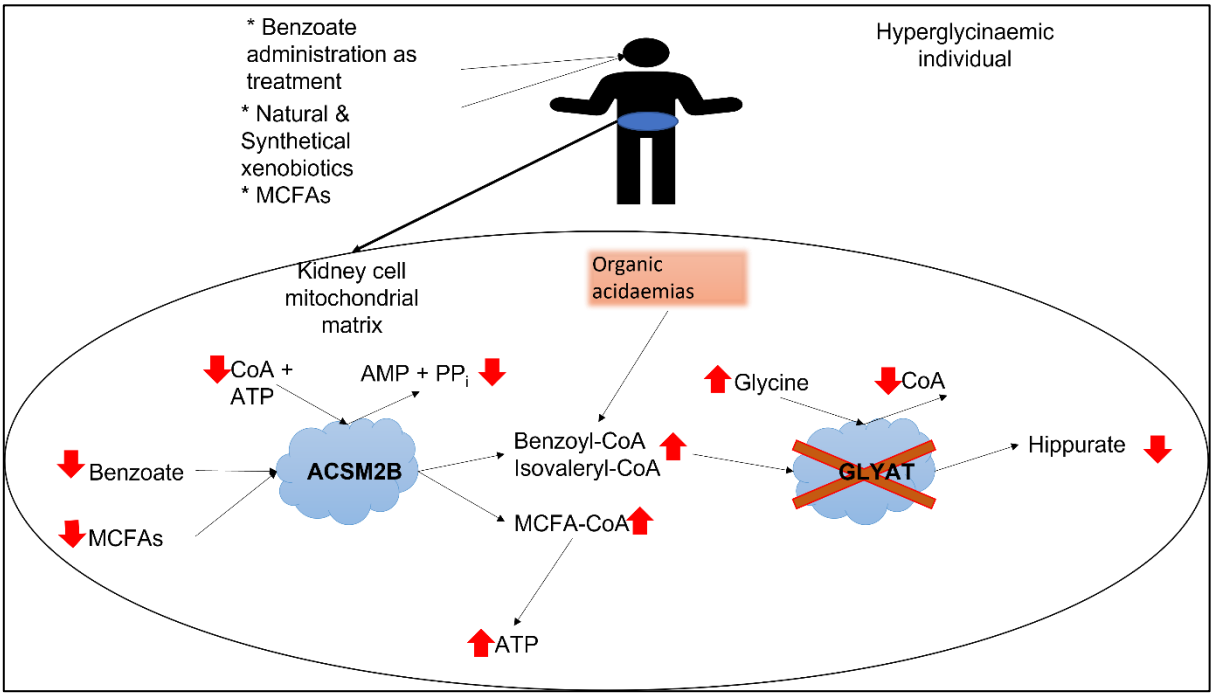


Figure 4: Consequences of a defective GLYAT in a hyperglycinemic individual.
The origins of the substrates are marked with stars. The red arrows indicate substrate and product levels changes when GLYAT is defective. It also indicates whether the change is increasing or decreasing. The orange rectangle indicates another important substrate origin (OA-CoA metabolites) site that should be considered since it may influence the detoxification efficiency of the glycine conjugation pathway and, more specifically, GLYAT. The defective GLYAT enzyme is marked with a red cross through it. CoA – Coenzyme A, MCFA – Medium chain fatty acids, ATP – Adenosine triphosphate, AMP – Adenosine monophosphate, PPi – Inorganic phosphate, OA – Organic acid.

When the activities of ACSM2B and GLYAT are simultaneously reduced, the problems explained in Figure 3 and Figure 4 are further complicated when organic acidemias are considered. This complication is due to the additional strain being added onto the glycine

conjugation pathway, especially when the excretion of the OA-CoA molecules is glycine conjugation dependent. One such organic acidaemia is isovaleric acidaemia, which is prevalent in the South African population (Dercksen *et al.*, 2012).

1.4.1 Isovaleric acidaemia and the contribution to the detoxification load of the phase II glycine conjugation pathway

Isovaleric acidaemia is caused by a defect in the leucine catabolic pathway that leads to isovaleryl-CoA accumulation. The irreversible decarboxylation of α-ketoisocaproic acid and the inability to convert isovaleryl-CoA to β-methylcrotonyl-CoA are two main drivers for the accumulation of isovaleryl-CoA (Tanaka *et al.*, 1966). The failure to convert isovaleryl-CoA to β-methylcrotonyl-CoA is due to a defect in isovaleryl-CoA dehydrogenase (E.C. 1.3.8.4) (Figure 5) (de Sousa *et al.*, 1986). Isovaleryl-CoA cannot be converted back to α-ketoisocaproic acid since the reaction is irreversible. Therefore, the bioaccumulation of isovaleryl-CoA is imminent.

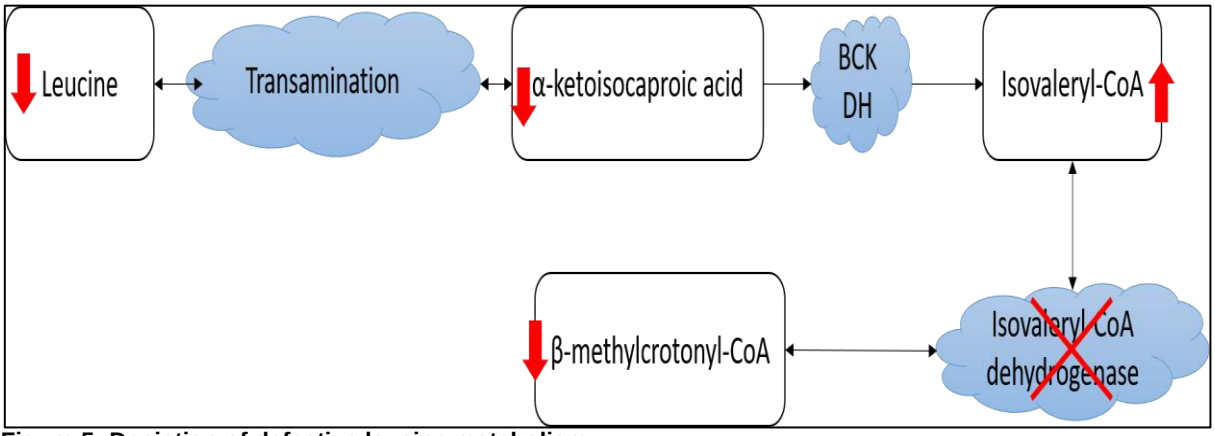


Figure 5: Depiction of defective leucine metabolism.
The figure indicates the effect of a defective isovaleryl-CoA dehydrogenase. The red arrows indicate the changes that may occur during this defect that ultimately leads to isovaleric acidemia. *BCKDH – Branched Chain Keto α-acid Dehydrogenase.

Usually, isovaleryl-CoA is converted to its glycine conjugate (isovalerylglycine) by glycine acylase, which is then excreted by the urinary tract. This changes when the glycine conjugation system is overloaded with CoA substrates, in which case isovaleryl-CoA is then hydrolysed, resulting in the accumulation of isovaleric acid that can lead to severe ketoacidosis, as depicted in Figure 6 (Krieger & Tanaka, 1976; Tanaka *et al.*, 1966). Furthermore, the additional load on the glycine conjugation pathway caused by isovaleryl-CoA accumulation influences the availability of glycine in other glycine-dependent pathways (Krieger & Tanaka, 1976). Apart from the bioaccumulation of isovaleryl-CoA in individuals with isovaleric acidemia, some other metabolites were also present in the urine of these individuals, including 3-hydroxyisovaleric acid, 4-hydroxyisovaleric acid (Lehnert & Niederhoff, 1981),

methylsuccinic acid, mesaconic acid, isovalerylglycine and isovalerylglutamate (Ensenauer *et al.*, 2004; Wysocki *et al.*, 1983).

The detrimental changes that might occur due to defective enzymes in the glycine conjugation pathway and the consequences that it may have further emphasise the need for accurately characterising this detoxification pathway.

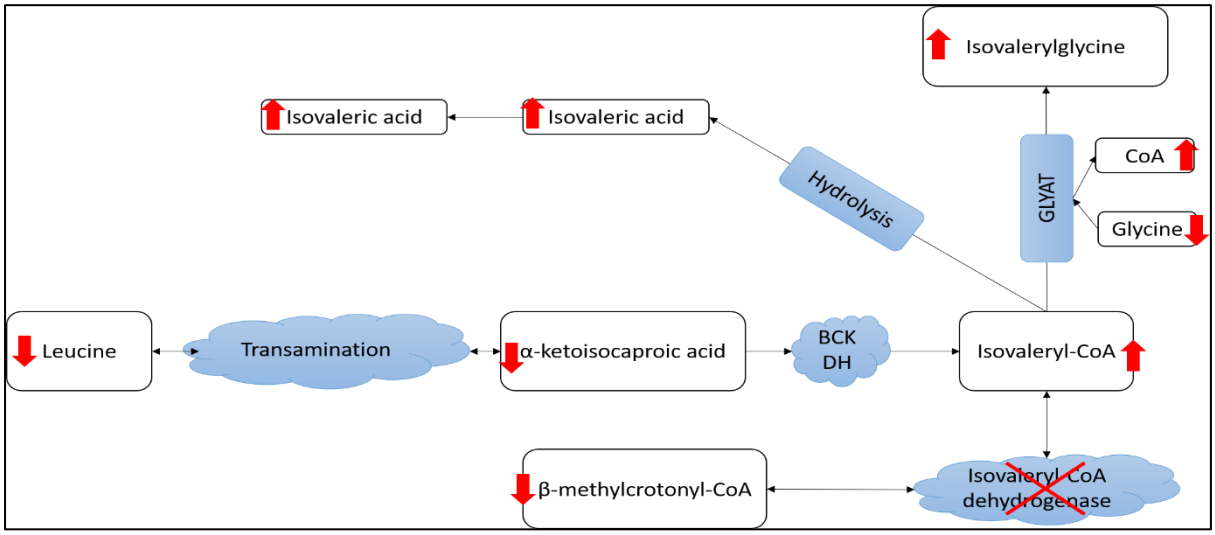


Figure 6: Consequences of isovaleryl-CoA conjugation and hydrolysation reactions due to accumulating isovaleryl-CoA.
The red arrows indicate the changes of metabolites that may occur during leucine metabolism and the effect it may have on the glycine conjugation pathway. If the glycine conjugation pathway is overwhelmed, the isovaleryl-CoA is hydrolysed back to isovaleric acid. The defective isovaleryl-CoA dehydrogenase is marked with a cross. BCKDH – Branched-chain alpha-keto acid dehydrogenase, CoA – Coenzyme A.

Previous studies (Gregersen *et al.*, 1986; Nandi *et al.*, 1979; van der Sluis *et al.*, 2013; van der Sluis *et al.*, 2017; van der Westhuizen *et al.*, 2000; Vessey *et al.*, 2003) only focused on characterising either ACSM2B (in crude liver extracts) or GLYAT separately as a one-step pathway. This study, however, will be the first to attempt to characterise the glycine conjugation pathway as a whole. The separation between the analysis of the two phases may be mainly due to the difficulty of measuring intermediates of this pathway, such as xenobiotic-CoA or OA-CoA.

An appropriate expression model should be established before the efficiency of ACSMs and GLYAT can be measured. Establishing a proper expression model for ACSMs and GLYAT will aid in investigating the effect of various substrates on the glycine conjugation pathway as a whole. Furthermore, it will also assist in studying the impact of different variants and the extent thereof in the glycine conjugation pathway, which will ultimately contribute to characterising the glycine conjugation pathway more extensively as a two-step pathway.

1.5 Different types of expression models that can be used

Finding an appropriate expression model can be difficult since every model has its advantages, disadvantages, and limitations. An array of expression models currently exist, and deciding which expression model to use is dictated by the characteristics of the protein of interest (Puetz & Wurm, 2019).

These models are organism dependent, such as prokaryotic (mainly using the bacteria *Escherichia coli* or *Bacillus subtilis*) or eukaryotic expression models (mainly using different yeast species or mammalian/plant/insect cells) (Puetz & Wurm, 2019; Yin *et al.*, 2007). Some systems use viruses or viral particles to infect a target cell, which actively produces the protein of interest (Gleba *et al.*, 2007).

Although traditional systems effectively express proteins of interest, advancements in techniques have provided the *in vitro* cell-independent method with various advantages over traditional techniques (Katzen *et al.*, 2005).

Each of these models will be briefly described along with its advantages and disadvantages to aid in selecting an appropriate expression system.

1.5.1 Prokaryotic models

Prokaryotic models have advantages, such as convenient protocols and cost-effectiveness when expressing foreign proteins of interest (Owczarek *et al.*, 2019). The expression of recombinant proteins in prokaryotes is usually performed by following a set of consecutive steps (Rosenblum & Cooperman, 2014) that can be seen in Figure 7. It starts with amplification of the gene of interest with PCR (1), inserting the gene into an appropriate cloning vector (2), the newly constructed plasmid is transformed into a bacterial cell (3), the cells are grown on media containing specific antibiotics that select the cells which contain the plasmid with the corresponding antibiotic resistance gene (4), after selection, the plasmids are extracted from the cloning host and transformed into an expression host (5), the cells are cultured and protein expression occurs (via the preferred induction method) (6), after expression, the proteins of interest can be retrieved by cell lysis and protein purification (7). Dual expression systems can also be implemented to avoid using two different plasmids for the expression of multiple proteins in a single host.

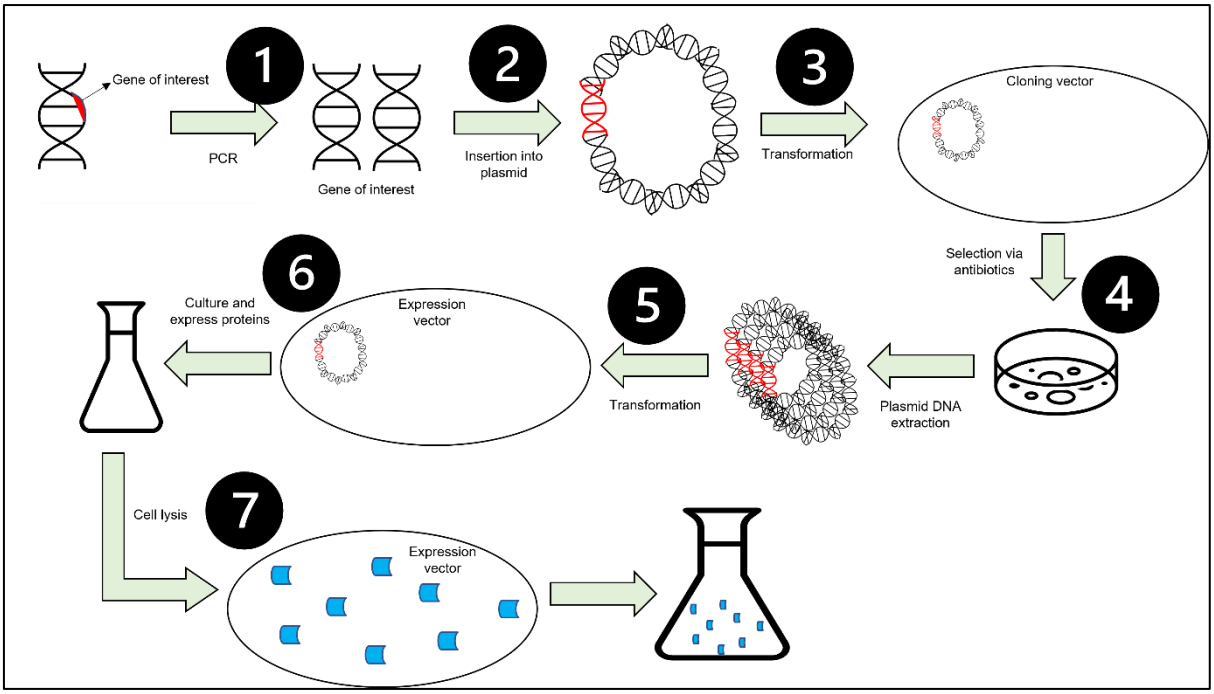


Figure 7: A simplified depiction of plasmid-mediated recombinant protein expression.
This figure depicts 7 common steps in recombinant protein expression using a cell-dependent method. The number in the figure indicates the steps preformed.

Apart from these common advantages of prokaryotic models, the use of some strains of *E. coli* or *B. subtilis* has more specific benefits in terms of protein yield, cell competency, excretion and growth rate (Huang *et al.*, 2012; McKenzie & Abbott, 2018; Rosano & Ceccarelli, 2014; Tripathi & Shrivastava, 2019; Yin *et al.*, 2007). For example, *B. subtilis* can readily be transformed by using either viruses or recombinant plasmids. Furthermore, it can excrete the expressed proteins directly into a simple and inexpensive growth medium while maintaining high cell densities, as depicted in Figure 8 (Yin et al., 2007), eliminating step 7 (Figure 7). Growth media can, in theory, be used to evaluate the expressed protein without requiring timely cell lysis techniques to retrieve the expressed proteins.

Because bacteria can be easily manipulated to express proteins of interest, a prokaryotic expression model may thus be very effective in analysing the biotransformation of benzoate (a bactericidal preservative) when ACSM and GLYAT are expressed within bacterial cells. This effect may be similar to that of metabolism-based plasmid-dependent systems. The plasmid provides an additional advantage toward the survivability of the cells that are usually affected when not expressing the proteins of interest. The metabolism-based plasmid-dependent systems can be divided into anabolic or catabolic systems. The catabolic system is more applicable in the case of the glycine conjugation detoxification pathway, where host growth and survival may be impaired if plasmids containing beneficial genes/genes are not present within the host organism (Kroll *et al.*, 2010).

Sodium benzoate is extensively used as a preservative due to its antimicrobial effects (Stanojevic, 2009). For example, when *Bacillus subtilis* is exposed to benzoic acid, it induces starvation and inhibits the transport of amino acids, which is lethal to the microbe and consequently causes mortality (Davidson *et al.*, 2005). However, should the glycine conjugation pathway (both the ACSM and GLYAT) be present within the microorganisms used, it ought to survive since it acquires the ability to detoxify the xenobiotics efficiently. This concept can be seen in Figure 8.

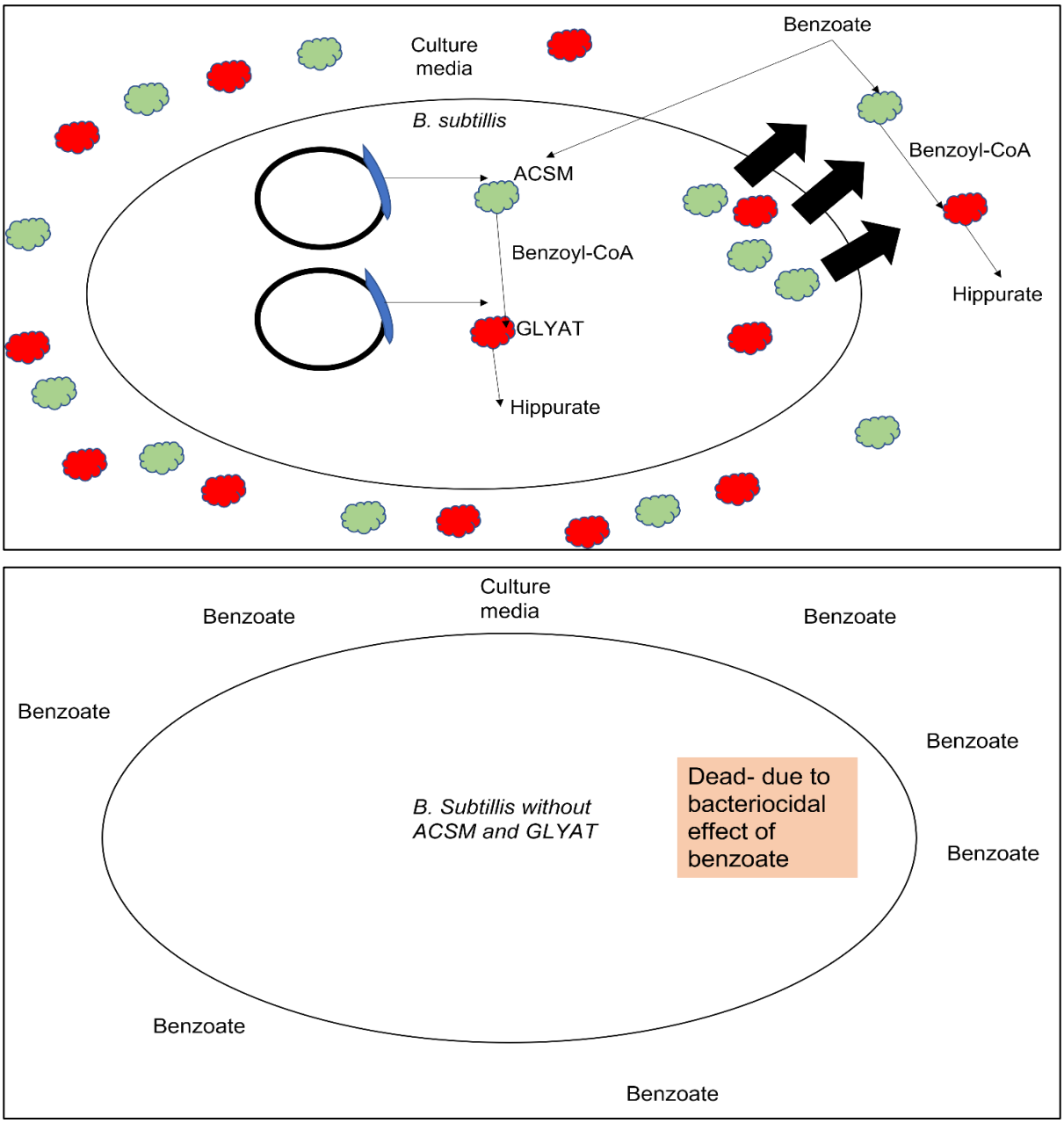


Figure 8: Concept of a bacterial cell surviving when expressing proteins of interest.
The concept depicts *B. subtillis* surviving in a benzoate-rich media. ACSM2B is indicated by the green bubbles in and outside the cell, whereas GLYAT is depicted with red bubbles in and outside the cell. The black arrows indicate the ability of *B. subtillis* to secrete the recombinant proteins into the growth media. It further depicts that the cells without the recombinant proteins die.

It should also be noted when using this approach to force the expression of an enzyme that some bacteria have natural strategies for survival in the presence of xenobiotics such as benzoate. These include three alternative metabolic pathways that use benzoate as a substrate, as described by Chen *et al.* (2014) and Davidson *et al.* (2005). The first pathway occurs under anaerobic conditions that convert benzoate to benzoyl-CoA, which is reduced to cyclohex-1,5-diene-1-carbonyl-CoA to be finally cleaved into acetyl-CoA. The second reaction occurs under aerobic conditions that oxidise benzoate through mono-oxygenases to dihydroxylated species. These are then cleaved and linearised by dioxygenases. The third set of reactions is called the epoxide pathway, where benzoate is converted to benzoyl-CoA through benzoate-CoA ligases. Benzoyl-CoA is then converted to epoxide, which undergoes a set of subsequent reactions to finally be involved in reactions similar to those of β-oxidation and the β-ketoadipate pathways and has succinyl-CoA and acetyl-CoA as products (Chen *et al.*, 2014). These pathways are exclusive to certain bacteria that degrade benzoate and/or different forms of benzoate. These are 2-aminobenzoate and benzoate degradation in *A. evanssii* (Altenschmidt *et al.*, 1993; Zaar *et al.*, 2001), benzoate degradation in *B. stearothermophilus* (Zaar *et al.*, 2001) and *P. putida* (Kaldalu et al., 2000) and the degradation of 4-hydroxybenzoate in *P. aeruginosa, R. leguminosarum, A. tumefaciens* and *Acinetobacter sp.* (Diaz *et al.*, 2000; Diaz *et al.*, 2001).

Even though prokaryotic models have advantages, some disadvantages should be considered since they can critically impact the expressed protein. When using *E. coli* for expressing the protein, the recombinant protein yield is high, but truncated species may be produced if the proteins of interest are larger than 70 kDa. Furthermore, the formation of inclusion bodies and codon bias is a common phenomenon that must be considered (McKenzie & Abbott, 2018; Tripathi & Shrivastava, 2019). If prokaryotic expression models are ineffective, the wheat germ extract (WGE) eukaryotic model may be better since the integrity is better than that of prokaryotic systems (Gagoski *et al.*, 2016).

Posttranslational modifications for eukaryotic expressed proteins are critical to ensure effective protein formation and may affect interaction with other molecules, the stability of the protein, cellular localisation thereof and enzymatic activity (Cruz *et al.*, 2019; ThermoFisher, 2020). Bacterial expression models can cause inappropriate protein folding because they do not contain appropriate chaperones or folding machinery when eukaryotic proteins are expressed (Assur *et al.*, 2012; Huang *et al.*, 2012; McKenzie & Abbott, 2018; Scopes, 2002; Yin *et al.*, 2007). When *E. coli* is used to express foreign proteins, it is important to note that some posttranslational problems may arise in fatty acid acylation, N- and O-linked glycosylation, phosphorylation and disulfide bridge formation, all of which are needed to form

an active protein (Owczarek *et al.*, 2019; Yin *et al.*, 2007). To avoid the issue of posttranslational modification, plasmids expressing chaperones may be cotransformed into the bacterial strain to aid in the proper folding of the expressed enzymes.

Another problem that should be considered is the occurrence of rare codons found in eukaryotic cells in contrast to the lack thereof in bacterial cells and that bacterial cells can be susceptible to metabolic stress, which in turn negatively affects the solubility of the expressed proteins (Owczarek *et al.*, 2019; Tripathi & Shrivastava, 2019).

Apart from disadvantages influencing protein expression, laboratory restrictions may influence the study decisions and outcomes. For example, working with specific bacterial species or yeast is not allowed in some laboratories. This limitation only allows some bacterial strains to be considered and may negatively affect the output of the results.

To overcome some of these protein-specific disadvantages (such as protein folding and posttranslational modification), eukaryotic models seem to be a better fit to express proteins of interest that are very sensitive to specific protein folding.

1.5.2 Eukaryotic models

Eukaryotic models include yeasts, mammals, plants, insect cells, microalgae, and filamentous fungi (Owczarek *et al.*, 2019; Puetz & Wurm, 2019; Tripathi & Shrivastava, 2019). Some benefits of using eukaryotic protein expression models are that they are scalable, do not have special media requirements (specific to yeasts and microalgae), and include eukaryotic posttranslational processes and machinery such as the endoplasmic reticulum, Golgi apparatus and nucleus (Owczarek *et al.*, 2019; Yin *et al.*, 2007). Therefore, eukaryotic expression systems do not need extra attention to posttranslational processes compared to prokaryotic models. The expression of proteins of interest can be achieved in multiple ways, including transient transfection, viral infection, or stable integration into the host genome (Assur *et al.*, 2012).

Yeasts and filamentous fungi as protein expression models have high commercial value since they share similar characteristics to higher eukaryotes (Ahmad *et al.*, 2014). Furthermore, this type of model has all the advantages of prokaryotic expression models with the added benefit of intracellular produced proteins being excreted into the extracellular environment, ease of culture in bioreactors, being less expensive than mammalian cells (Owczarek *et al.*, 2019; Puetz & Wurm, 2019; Tripathi & Shrivastava, 2019; Yin *et al.*, 2007) and having

posttranslational modification processes (Tripathi & Shrivastava, 2019; Wu, 2008). The expression of recombinant proteins is similar to that of prokaryotes when using yeasts (Figure 7), which means that most advantages and disadvantages are shared. However, yeast proteases and metabolites can affect the expressed protein of interest to produce a defective protein (Owczarek *et al.*, 2019).

Plants can express foreign proteins of interest, which is achieved by establishing transgenic plants. However, using this technique is a timely and expensive process. Other disadvantages that can affect the protein of interest are contamination by insecticides and other plant metabolites when proteins have to be retrieved and purified. There are, however, some advantages, such as efficiently expressing heterologous proteins using viral infection for inserting the gene of interest (Gleba *et al.*, 2007), upscaling possibilities (Tripathi & Shrivastava, 2019) and posttranslational modification of the expressed proteins (Merlin *et al.*, 2014).

Unicellular plants, such as microalgae, are economical, easily controlled, and can reach high cell densities in a short time, much like bacterial models. A unique characteristic of microalgae is that recombinant proteins can be expressed using either the chloroplast or nuclear genome (Owczarek *et al.*, 2019) and has the added advantage of posttranslational modification when expressing recombinant proteins (Merlin *et al.*, 2014; Tripathi & Shrivastava, 2019).

Mammalian and insect cells can be used and manipulated to express proteins of interest. These systems can also work cooperatively with viral vectors, much like transgenic plants. For example, the baculovirus expression vector system is commonly used to infect insect cells (Tripathi & Shrivastava, 2019). Other recombinant DNA particle insertion methods are electroporation or microinjections (Owczarek *et al.*, 2019).

The most important advantage is that these systems can produce large and complex proteins, with the increased possibility that the newly synthesised proteins will undergo posttranslational modifications similar to those that occur in human cells (Owczarek *et al.*, 2019; Tripathi & Shrivastava, 2019). Mammalian models have their disadvantages. Some disadvantages that should be considered are the increased susceptibility to infection by animal viruses and the complexity of the media used in mammalian cell growth and maintenance (Owczarek *et al.*, 2019; Puetz & Wurm, 2019). Recombinant gene expression in mammalian cells offers the closest expression conditions to humans than any other model since the cellular conditions are mimicked in this expression model (Assur *et al.*, 2012). However, eukaryotic models pose the problem of being expensive and time consuming. Therefore, prokaryotic models seem to be more cost-efficient and less timely to express recombinant proteins in higher yields (Owczarek *et al.*, 2019; Tripathi & Shrivastava, 2019).

1.5.3 Viral models

Viral models can be used with plants in either first-generation viral vectors (Figure 9A) or second-generation viral vectors (Figure 9B). First-generation viral vectors are introduced to plants using infectious nucleic acid copies or mature viral particles. In contrast, second-generation viral vectors use deconstructed viral particles as an infection mechanism after being introduced to the plant (Gleba *et al.*, 2007). This model is time-consuming, expensive, and very sensitive to environmental conditions when using multicellular plants. Nevertheless, it can be scaled up and yield a high quality and quantity of protein (Tripathi & Shrivastava, 2019). When unicellular plants, such as microalgae, are used, the process combines the advantages of prokaryotic models with the advantages of multicellular plant models. For example, microalgae are fast-growing and rapidly manipulatable but have the added advantage of posttranslational modifications.

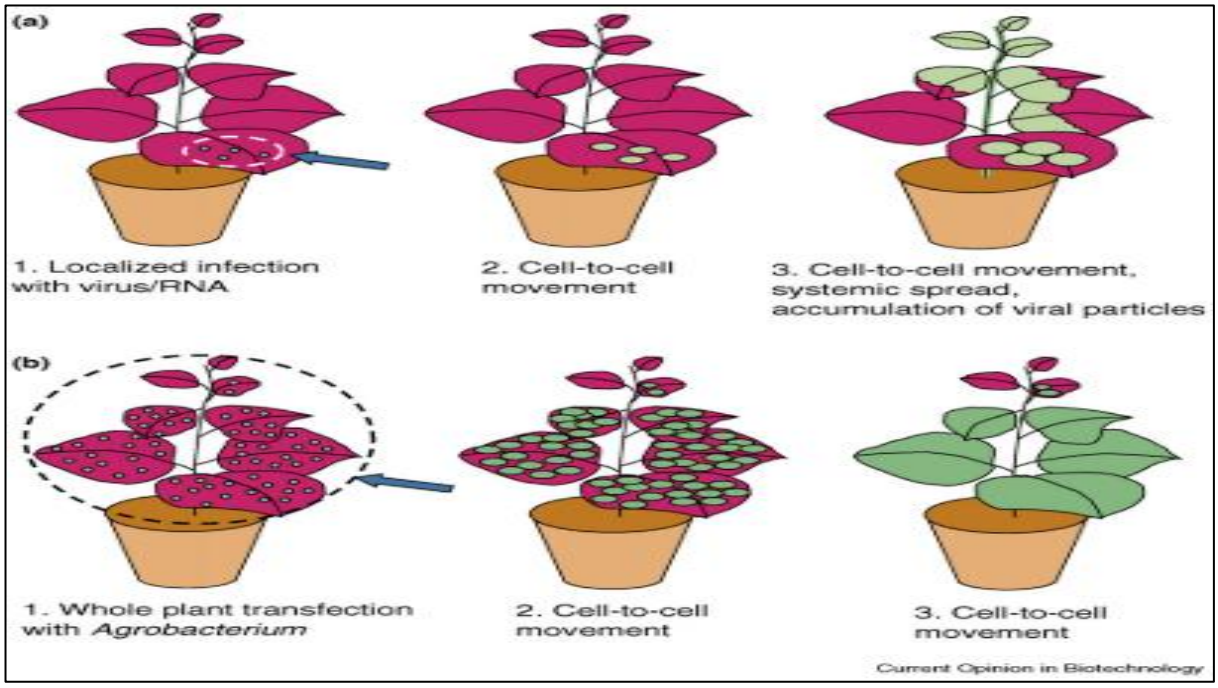


Figure 9: Depiction of the process of viral-plant infection to express recombinant proteins (from Gleba *et al*., 2007).
Part A depicts first-generation viral plant vectors, whereas part B depicts second-generation plant viral vectors.

Baculovirus infection is another alternative viral model that does not use plant tissue but rather insect cells. This system is highly specific to insect cells since it does not infect humans, livestock, or chickens. A prerequisite will be preparing the *baculovirus* by inserting the recombinant plasmid containing the gene of interest (McKenzie & Abbott, 2018; Tripathi & Shrivastava, 2019; Yin *et al.*, 2007). This model also combines some of the advantages of prokaryotic models with the added advantage of eukaryotic models. These advantages are the cost efficiency and growth rate of the cells combined with the required posttranslational modification of the protein.

1.5.4 Choosing the most applicable expression model

A table summarising the common advantages and disadvantages of the cell-dependent models is shown in **Error! Unknown switch argument.**. This table shows that bacterial or yeast models are usually preferred as a first choice in protein expression due to their abundant advantages.

Table 7: Common advantages and disadvantages of protein expression models.

Bacterial	Yeast	Plant	Mammalian or insect cells	Viral
Advantages	Advantages	Advantages	Advantages	Advantages
• Conveniency • Cost-effective • High protein yield • Cell competency • POI excretion • Growth rate	• High commercial value • High protein yield • POI excretion • Cost-effective • Growth rate • PTM	• Microalgae = shares bacterial advantages • PTM • Low contamination risk	• Produce large and complex proteins • PTM	• Shares eukaryotic models
Disadvantages	Disadvantages	Disadvantages	Disadvantages	Disadvantages
• Truncated proteins • Inclusion body formation • Codon bias • No PTM • No rare codons • Metabolic stress	• Metabolic stress • Codon bias • Truncated proteins • Inclusion body formation	• Time-consuming • Expensive • Contamination	• Susceptible to infection with viruses • Growth and maintenance • Complex protocols • Time-consuming	• Can be time consuming • Has the risk of contamination

A bacterial model will be more advantageous to use in this study. This is because of the cost-effectivity, growth rate and convenience of bacterial cells. However, special care will be given in selecting strains that cannot degrade benzoate. This will be done by avoiding the previously listed bacterial species as hosts and adding a control into the study to ensure there aren't additional benzoate degradative pathways present. Therefore, the control will be the cell lysate of the chosen strain that does not contain the recombinant proteins. When GC–MS then measures the biotransformation of benzoate, it will not be able to degrade the benzoate added and will therefore be a good host organism to analyse the glycine detoxification pathway.

1.5.5 Cell-free models

Cell-free protein expression is not limited by cell viability or membrane integrity and can be performed rapidly with more flexibility (Catherine *et al.*, 2013). Furthermore, it has higher translation efficiency and is easier to amend the *in vitro* environmental conditions and purify the expressed untagged proteins (Katzen *et al.*, 2005) than cell-dependent techniques. The reaction conditions do not pose a problem since they can also be easily configured for functional analyses under various reactions (Catherine *et al.*, 2013; Katzen *et al.*, 2005). Special reagents that aid in protein folding and formation can also be added to the reaction

mixture (Dopp & Reuel, 2018; Katzen *et al.*, 2005; Rosenblum & Cooperman, 2014; Schinn *et al.*, 2016). These reagents can be seen in Figure 10 and include chaperones, disulfide isomerase and translation machinery. However, if a system is optimised for the specific expression of proteins, it will not be as efficient in other systems that yield a different protein (Gagoski *et al.*, 2016).

Another pitfall in this method is that it is not entirely cell-free when cells are still needed to amplify the plasmid template (Figure 10B). To bypass the culture-based amplification of the template, PCR can be used (Figure 10A) to amplify the template gene. This method is not the most beneficial since linear DNA undergoes rapid degradation in extract-based cell-free synthesis due to high exonuclease activities (Catherine *et al.*, 2013; Schinn *et al.*, 2016). Some strategies can minimise this problem. Some include adding secondary stem–loop structures at the end of the transcribed mRNA to protect it from nuclease activity, inhibiting degrading enzymes by using Gam proteins, removing the genes that encode for degradative enzymes, and physically shielding the termini of the linear DNA (Catherine *et al.*, 2013).

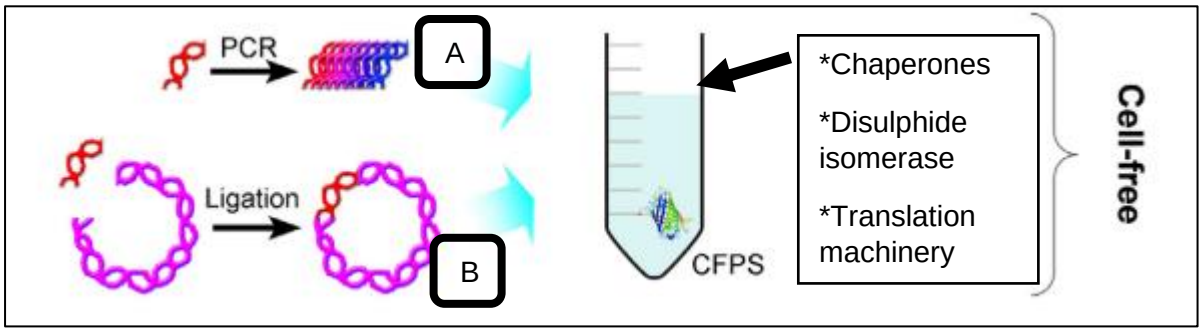


Figure 10: Adapted simplification of the cell-free expression system (from Rosenblum & Cooperman, 2014).
Part A of the figure utilises linear DNA fragments to synthesise the recombinant proteins, whereas part B uses a plasmid that contains the protein of interest to express the protein. The list next to the tube represents common reagents required to express and fold proteins from the DNA template. CFPS – Cell Free Protein System.

Different analytical methods are available to determine protein activity after selecting a fitting expression model to express the soluble protein of interest.

1.6 Techniques that can be used for enzyme activity evaluation

Changes in metabolite abundance during the phase I and phase II stages can be measured to compare the efficiency of differentially expressed variants of *ACSM2B* and *GLYAT*. The data obtained from measuring the changes in substrates and products can also be further used to compare substrate specificity and affinity when substrates are mixed. Several techniques can measure substrate clearance, including enzyme assays and chromatographic methods coupled to a mass spectrometer.

1.6.1 Enzyme assays

Enzyme assays serve different purposes, including determining the presence or absence of a specific enzyme or the amount of that specific enzyme in a sample. However, enzymes are susceptible to substances in the reaction mixture, and enzyme activity *in vitro* does not always resemble the actual *in vivo* enzyme activity, even when reaction conditions stay identical (Bisswanger, 2014; Scopes, 2002). Most physicochemical conditions that affect enzyme activity are inhibitors, pH, ionic strength, temperature and the nature of the salts present (Scopes, 2002). Another problem arises when the protocol must be adapted and performed on different instruments. Furthermore, when new enzymes are discovered, specific protocols must be first established according to the conditions for that enzyme (Bisswanger, 2014).

The most straightforward assays that can be performed measure the differences in absorbance values when one of the physicochemical parameters is altered (Scopes, 2002). These are photometric assays (Bisswanger, 2014; Del Olmo *et al.*, 2017; Matsuo *et al.*, 2012), which are less susceptible to disturbances but are sometimes not sensitive enough when measuring enzyme activity using low substrate concentrations. These are assays such as the DTNB assay (Gregersen *et al.*, 1986), which is problematic because it is not sensitive enough to detect xenobiotic-amino acid conjugation at low levels (van der Westhuizen *et al.*, 2000).

A continuous enzyme assay measures the reaction kinetics as it occurs (Scopes, 2002). Therefore, this type of assay may be a helpful tool in measuring the efficiency of both the Phase I and II velocities in two separate reactions. However, careful optimisation of the reaction to avoid the effects of inhibitors or other substances that may influence the actual efficiency of the enzymes must be done.

An enzyme assay that measures the release rates of pyrophosphate in the activation by ACSM2B in Figure 2 (Phase I) can be used to measure the efficiency of ACSMs. An example of this enzyme assay is the EnzChek® Phosphate assay kit, which quickly and sensitively measures the consequent release of pyrophosphate from enzymatic reactions in solution (ThermoFisher, 2002). This assay applies to the first phase of the glycine conjugation pathway, where pyrophosphate is released due to the occurring reaction. The amount of pyrophosphate released is then proportional to the ACSM2B activity. This assay, however, only measures one phase of the biphasic glycine conjugation pathway. This may be problematic since it only represents half of the pathway activity. An overall disadvantage of enzyme assays appears to be that they measure either substrate conversion or product formation but not both. Only one variable can be analysed at a time when using enzyme

assays. Therefore, this poses a problem when multiple variables need to be measured in a biphasic pathway (such as the glycine conjugation pathway).

Although there are many disadvantages to enzyme assays, there are some advantages and data that other methods, such as chromatography-mass spectrometry, do not provide. One crucial advantage of enzyme assays is, therefore, the accurate measurement of enzyme kinetics (such as V_{max}, K_m, K_i and K_{cat}). These data values can then be used to interpret important enzyme characteristics such as substrate affinity (K_m) and inhibition patterns (K_i). These values are important in the glycine conjugation pathway since they provide information about possible substrates and inhibitory substances. However, this can become more complicated when enzymes do not follow the normal Michaelis-Menten reactions. These are allosteric enzymes (such as GLYAT) that would rather have a sigmoidal enzyme activity response (Bisswanger, 2014; Scopes, 2002). For example, GLYAT was recently suggested to follow sigmoidal enzyme characteristics, which means that the enzyme has a ground/relaxed (little activity) and an activated state (more or maximal) activity Figure 11 (Rohwer *et al.*, 2021). This may influence the interpretation of the enzyme activity results since it is unknown in which state (ground or activated) the enzyme is being tested.

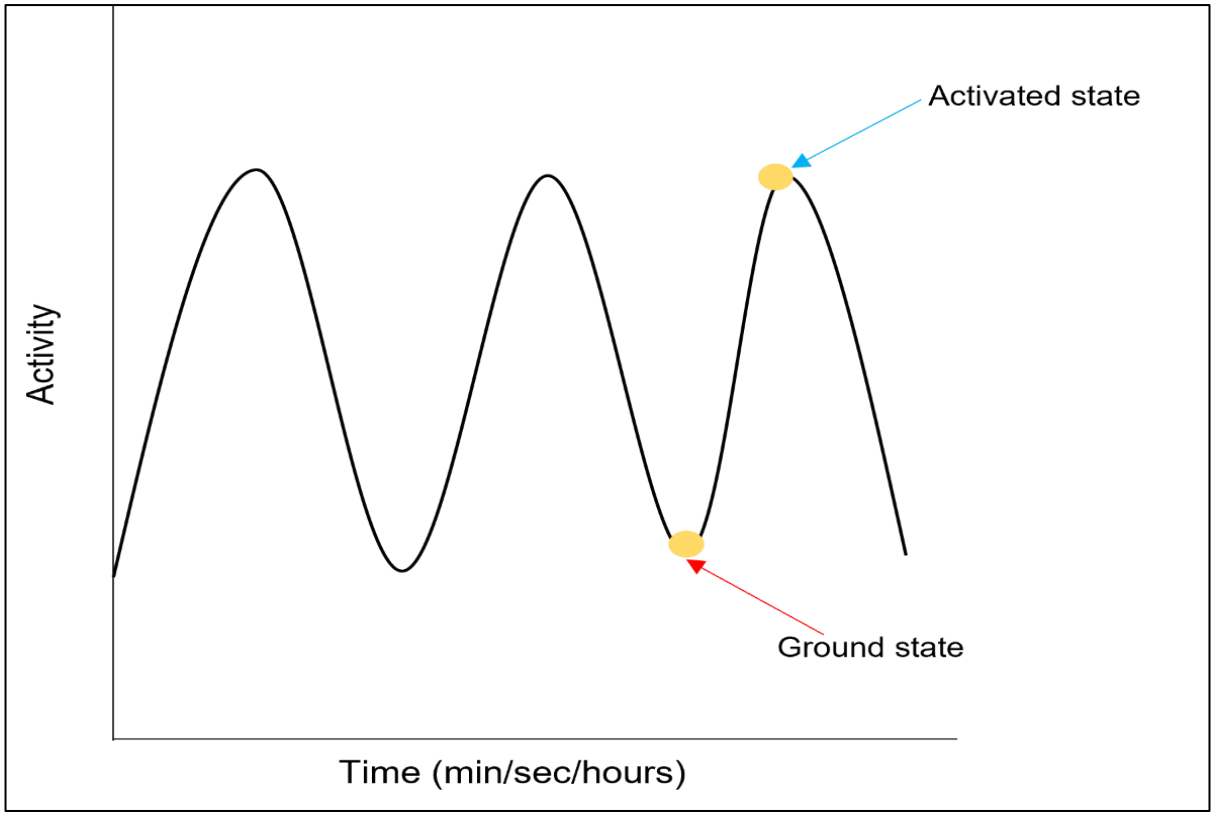


Figure 11: Simplified activity states of an allosteric enzyme.
The yellow dots indicate the respective states of an allosteric enzyme.

Another advantage is the precise measurement of colour intensity when an enzymatic conversion reaction occurs (if adequate substrate concentrations are used). Modern spectrophotometers even cover the UV spectra when the visible spectra cannot or are ineffective in measuring the colour change (Bisswanger, 2014). This will then be advantageous when using the DTNB method, which determines the glycine-dependent CoA release of the Phase II reaction of the glycine conjugation pathway at 412 nm (Gregersen *et al.*, 1986).

Because enzyme assays can only measure one specific variable at a time, it poses a significant limitation for this study. Therefore, to measure the efficiency of both ACSM and GLYAT simultaneously, MRM (multiple reaction monitoring) or a chromatography-MS (mass spectrometry) method may serve as a more efficient tool to detect changes in decreasing substrate levels (of Phase I) and increasing product levels (of Phase II). The production of products such as hippurate was measured similarly to that in van der Sluis *et al*. (2017). That study, however, only focussed on hippurate. Therefore, if both benzoic acid and hippurate levels were targeted, a more accurate representation of the glycine conjugation pathway could be observed.

1.6.2 Mass spectrometry

1.6.2.1 Gas chromatograph coupled to a mass spectrometer

This type of analysis uses a gas chromatography–mass spectrometry (GC–MS) system, which partitions compounds between gas and liquid phases at a given temperature. This then influences the movement of the molecules through the GC column. The mobile phase is gas-based in a GC–MS system. The compounds that move through the column at different stages (dependent on a temperature gradient) and those partitioned between a gas and liquid phase allow the compounds to be separated over time (Pereira Braga & Adamec, 2019).

The GC–MS technique is most helpful in evaluating volatile, semi volatile, and non-volatile compounds from various samples, even if it consists of a mixture of compounds (Villas-Bôas *et al.*, 2005; Zhang *et al.*, 2012). This technique provides multiple advantages, including a more precise and easy quantifiable approach, eliminating the use of hazardous, radioactive materials (Prabhu *et al.*, 2017). Furthermore, it is a robust, cost-effective analytical technique with high reproducibility (Villas-Bôas *et al.*, 2005; Zhang *et al.*, 2012), separation efficiency with low ion suppression and high detector response (Koek *et al.*, 2011; Mastrangelo *et al.*, 2015; Villas-Bôas *et al.*, 2005).

However, this method has some disadvantages. If non-volatile compounds are being analysed, an additional derivatization step to increase the boiling point of those compounds and increase the heat stability before being analysed must be included (Beale *et al.*, 2018; Mastrangelo *et al.*, 2015). This causes the procedure to become more time consuming and requires more care when handling the sample. Other pitfalls are that unknown derivatised compounds are challenging to analyse, heat liable compounds cannot be analysed, and GC–MS methods can only detect low molecular weight metabolites (Beale *et al.*, 2018; Mastrangelo *et al.*, 2015). Novel derivatives pose a problem due to chemical alterations (Villas-Bôas *et al.*, 2005). Furthermore, the more steps included in the sample preparation, the more experience is needed when analysing different metabolites in a sample (Mastrangelo *et al.*, 2015). A GC–MS system cannot measure larger molecules such as CoA (molecular weight – 767.535 g/mol) due to the detection limits (Vivekanandan-Giri *et al.*, 2011).

GC–MS analysis of the substrate and product changes of a metabolic pathway can, therefore, be used to determine metabolic pathway effectivity that has metabolites at very low concentrations (Del Olmo *et al.*, 2017). For example, by taking substrate presence (xenobiotics and glycine) and products synthesised (glycine conjugates) into account of the glycine conjugation pathway, quantifying these substrates and products will aid in elucidating

the effectivity of a specifically expressed variant. A similar approach was followed by Prabhu *et al.* (2017) to determine the efficiency of cholesterol synthesis enzymes. In this study, they measured the relative activity of cholesterol synthesis enzymes by determining substrate conversion rates using an established GC–MS system. This was done using labelled sterol intermediates as controls.

1.6.2.2 Liquid chromatograph coupled to a mass spectrometer

Liquid chromatography–mass spectrometry (LC–MS) is based on the separation of compounds based on the interaction with a stationary phase in a separating column before being detected by MS. In LC–MS, the molecules are moved through the column using a liquid mobile phase. The stationary phase is usually composed of spherical beads coated with functional groups that determine separation in the column. This technique can also be applied in the targeted and untargeted analysis of polar, nonpolar, and even neutral compounds. Depending on the experimental setup, different columns can differentiate between different compounds, even in a complex sample (Pereira Braga & Adamec, 2019).

When using an LC–MS system, various advantages include, but are not limited to, detecting compounds in low concentrations and high specificity towards specific compounds when using targeted detection. In contrast, the untargeted option can detect as many compounds as possible (Villas-Bôas *et al.*, 2005).

However, there are some disadvantages that should be considered when using an LC–MS system. These are the system's incompatibility regarding the liquid flow rate into the high vacuum of the MS. Furthermore, incompatibility of the system may also be caused when a non-volatile mobile phase is used in chromatographic separation. Ionisation is a critical step in analysing compounds in the LC–MS system. This may pose a problem when ionisation of non-volatile and thermally liable compounds occurs. It should, however, be noted that these disadvantages are circumvented by soft ionisation and modification of LC–MS techniques (Villas-Bôas *et al.*, 2005).

Two major disadvantages of the LC–MS system that may significantly impact this study are ion suppression and the need for expensive internal standards for accurate quantity comparison (Leung & Fong, 2013).

Considering the advantages and disadvantages listed in Table 8, it was determined that a GC–MS system would be better applicable to this study, specifically because it eliminates ion suppression and is not dependent on expensive internal standards.

Table 8: Advantages and disadvantages of metabolite analysis techniques.

Enzyme assays	GC–MS	LC–MS
Advantages	**Advantages**	**Advantages**
• Determines the presence and activity of enzymes • Measures enzyme kinetic data	• Can be targeted and untargeted • Exact and easily quantifiable approach • Do not use hazardous or radioactive materials • Robust and cost-effective • High reproducibility • High detector response • No ion suppression • Not dependent on expensive internal standards	• Can be targeted and untargeted • Detecting low concentrations of metabolites • Soft ionisation can circumvent some disadvantages

Disadvantages	Disadvantages	Disadvantages
• Enzymes are susceptible to a lot of physicochemical factors • *In vitro* activity does not always resemble *in vivo* activity • Instrument and protocol-specific • New protocol establishment for every new enzyme discovered • Not very sensitive to very low concentrations • Optimisations • Can only measure one variable at a time	• Nonvolatile metabolites need an extra derivatisation step • Can only detect low molecular weight metabolites • Cannot analyse heat liable metabolites • More experience is needed with complex methods • Cannot detect large molecules	• Incompatibility regarding liquid flow rate in high vacuum • The ionisation of nonvolatile and heat liable metabolites • Ion suppression • Dependent on expensive internal standards • Difficulty in detecting glycine (circumventable)

1.7 Problem statement

The glycine conjugation pathway is a conserved two-phase detoxification pathway responsible for the elimination (by urinary excretion) of different xenobiotics (such as benzoate) in the form of a glycine conjugate (hippurate) (van der Sluis, 2018; van der Sluis *et al.*, 2015; van der Westhuizen *et al.*, 2000). Currently, no studies have investigated the influence of genetic variation on the Phase I ACSM2B enzyme (van der Sluis, 2018). Only two studies determined the kinetic parameters of crude cell extracts, after which characterisation of the ACSM2B enzyme was severely neglected (Vessey *et al.*, 1999; Vessey *et al.*, 2003). However, multiple studies in the past decade have investigated the Phase II GLYAT counterpart of the glycine conjugation pathway (Badenhorst *et al.*, 2013; Badenhorst *et al.*, 2014; Dempsey *et al.*, 2014; Koshiba *et al.*, 2020; Matsuo *et al.*, 2012; Rohwer *et al.*, 2021; Schutte *et al.*, 2019; van der Sluis & Erasmus, 2016; van der Sluis *et al.*, 2013; van der Sluis *et al.*, 2017; van der Sluis *et*

al., 2015). These studies investigated the influence of genetic variation on *GLYAT* and enzyme activity and proposed an enzyme mechanism.

Sodium benzoate consumption was estimated to increase by approximately 5.5% annually during the 2017-2022 period. This increase may be due to improving living standards (regarding food security), extending shelf-life requirements and convenience, and increasing urbanisation (IHSMarkit, 2017). Global consumption during 2017 can be seen in Figure 12, where China, Western Europe and North America were the top consumers of benzoate. Although there are many socioeconomic advantages to using benzoate (usually in the form of sodium benzoate), the effect of increased consumption is neglected and not researched in depth.

There are, however, multiple animal studies indicating a detrimental effect when increased amounts of benzoate are ingested or present in the environment. Some of these detrimental effects were observed in the liver and kidney functions of rats (Oyewole *et al.*, 2012), motor impairment and anxiety in sodium benzoate-treated rats (Noorafshan *et al.*, 2014), ammonia toxicity in mice when benzoate treatment is performed (as reported by Nair (2001); Oyewole *et al.* (2012)), benzoate dose-dependent malformations in zebrafish embryos and hatching rate (Srinithi & Bhargava, 2018; Swenson, s.a.), and gastrointestinal tract lesions in zebrafish (Tsay *et al.*, 2007). Some adverse effects have also been observed in humans, e.g., ADHD (attention deficit hyperactivity disorder) in children (Beezhold *et al.*, 2014; Kilic *et al.*, 2006).

This knowledge gap of the neglected ACSM2B raises an important question of how the human body will adapt to the increased ingestion of xenobiotics and accumulation thereof if it has the ability to do so. There are no enzyme kinetic data available for ACSM2B except for one previous study (Vessey *et al.*, 2003). Since then, no other studies have been performed to accurately determine the kinetic parameters of ACSM2B (Knights & Miners, 2012; van der Sluis, 2018). Suppose the body cannot adapt and eliminate xenobiotics at a fast enough rate. This will lead to an accumulation of toxic levels of xenobiotics, xenobiotic-CoA, and OA-CoA, which may be toxic to the human body. Furthermore, it also raises the question of whether enzymes other than ACSM and GLYAT are involved in the glycine conjugation pathway, e.g., GLYATL1, to detoxify different xenobiotics and whether these enzymes can handle the overconsumption of xenobiotics or relieve the pressure placed on ACSM2B and GLYAT.

Therefore, the analysis of the efficiency of the biphasic glycine conjugation pathway plays a critical role in answering the abovementioned literature gaps. Furthermore, identifying other enzymes (such as GLYATL1 mentioned above) that also contribute to the detoxification

process of xenobiotics is crucial since it will aid in filling the bigger picture of the role of these enzymes in the glycine conjugation pathway.

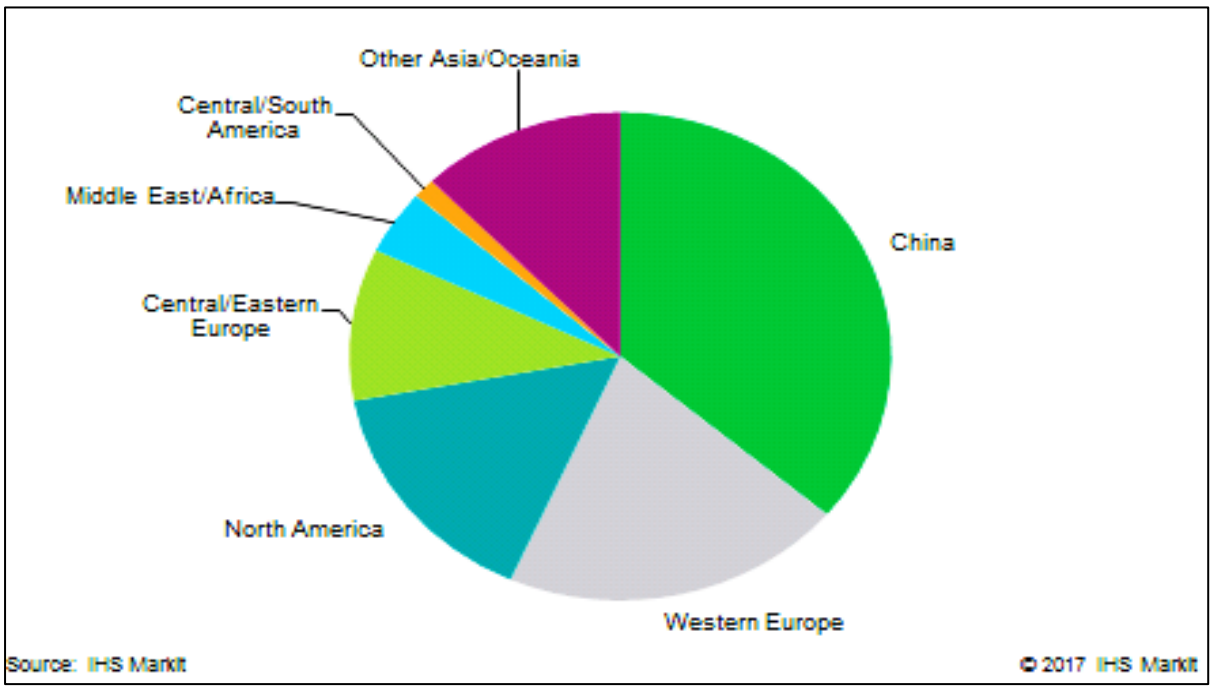


Figure 12: Indicates the global consumption rates of benzoic acid/benzoate in 2017 (from IHSMarkit (2017)).

Furthermore, by establishing a model system where ACSM2B and GLYAT can be characterised as a biphasic detoxification pathway, the effects of genetic variants of these enzymes on the detoxification ability of this pathway can then be measured in the future using this established expression system. For example, the $S_{156}C_{199}$ variant of *GLYAT* was shown to severely reduce the enzyme activity of GLYAT compared to the wild type (van der Sluis *et al.*, 2015). When GLYAT is defective, benzoyl-CoA might accumulate, and CoA might not be readily available in other CoA-dependent pathways.

1.8 Research aim and objectives

1.8.1 Aim

The main aim of this study was to establish a model system in which variants of *ACSM2B*, *GLYAT* and *GLYATL1* can be expressed to determine the xenobiotic biotransformation capabilities of the biphasic glycine conjugation pathway.

1.8.2 Objectives

- Identify an expression system that will result in the expression of biologically active recombinant ACMS2B and GLYAT.

- Evaluate the biotransformation of sodium benzoate in the cell culture model system expressing soluble ACSM2B and GLYAT using an established GC–MS method.

- Identify and optimise an expression system that will result in the expression of biologically active recombinant GLYAT and GLYATL1.

- Evaluate the biotransformation of benzoyl-CoA in the cell culture model system expressing soluble GLYAT and GLYATL1

1.9 Ethical considerations

This study is considered a no-risk study, which means no interaction with human participants or animals. No samples of human participants or animals will be used. This study will not affect the environment since established SOPs regarding waste removal and PPE will be followed. The ethical number assigned to this study is **NWU-00473-20-A1.**

Chapter 2: Establishment of a coexpression model for ACSM2B and GLYAT

2.1 Introduction

A coexpression system was analysed to express ACSM2B and GLYAT to characterise the individual enzymes of the glycine conjugation pathway in terms of substrate increase and product decrease. This means that the chosen expression model can be used in future studies to characterise the different components of the glycine conjugation pathway to allow for an enzymatic model to be established. Furthermore, it will also establish proof of concept that the glycine conjugation pathway is indeed needed for survival (van der Sluis *et al.*, 2015). For example, a bacterial cell that will survive benzoate toxicity when it contains both ACSM2B and GLYAT of the glycine conjugation pathway, in contrast to a bacterial cell that will die in the presence of benzoate.

The cloning and expression were performed using the amended workflow of Tolia and Joshua-Tor (2006) by using BL21(DE3)-derived OverExpress® C41(DE3)pLysS cells as the host. This host can express toxic proteins from various organisms, such as bacteria, plants, mammals, yeasts, and viruses. The host will be able to express toxic proteins due to containing a mutation that lowers T7 RNA polymerase activity, which lowers the expression of the toxic protein to prevent apoptosis. The pLysS strain of the OverExpress® C41(DE3) range contains a smaller second plasmid that expresses T7-lysozyme to aid in lowering the expression rates of the newly cloned genes (Saida *et al.*, 2006).

The steps in Figure 7 were followed to clone both *ACSM2B* and *GLYAT* into the pETDuet-1 vector, which can be seen in (Figure 15). *ACSM2B* was inserted into MCS-1 (Multiple cloning site-1) of the pETDuet-1 vector, followed by *GLYAT* into MCS-2 (Multiple cloning site-2). InFusion® cloning (TakaraBio) was used for both cloning steps. The basic principle of InFusion® reactions can be seen in Figure 13. Carbenicillin and chloramphenicol were used in subsequent growth media to ensure that only the cells containing the pETDuet-1_ACSM2B/GLYAT and pLysS plasmids were selected.

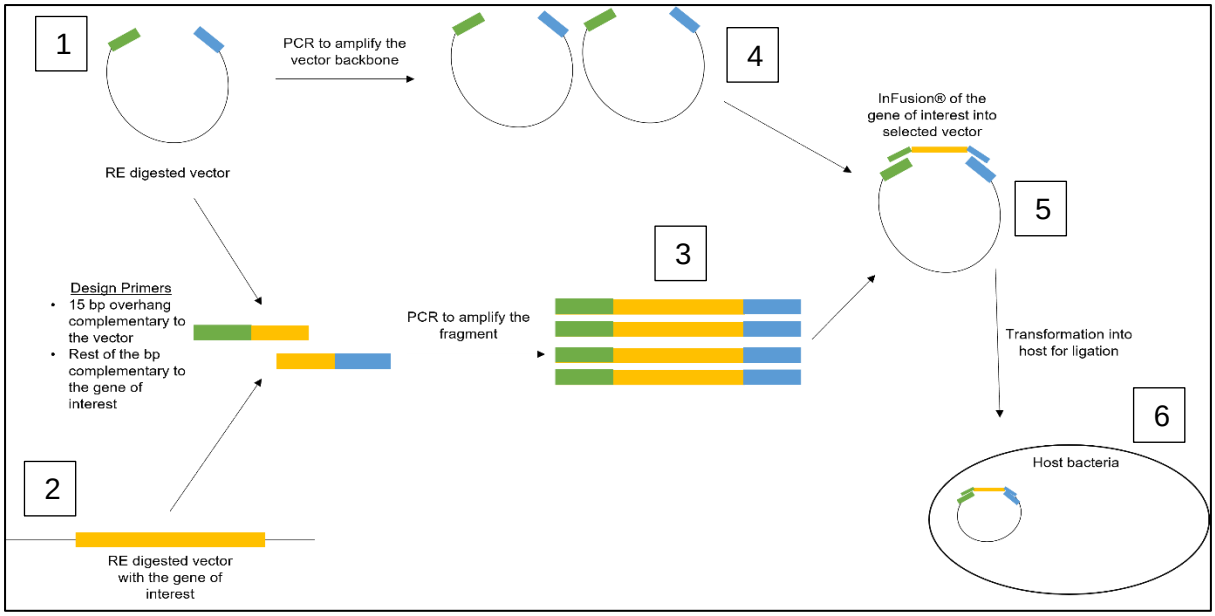


Figure 13: Principle of InFusion® cloning reactions.
The image indicated six simplified steps to explain the principle of InFusion® cloning. 1. Linearise the vector, 2. Linearise the vector with the fragment insert. 3. Amplify the target gene to match with the 15 bp ends of the linearised vector. 4. The linear vector is amplified to have the exact ends of the amplified fragment. 5. Both the linear vector and fragments are mixed in the InFusion® reaction tube. 6. The newly created vector with the insert is cloned into a cloning host.

A simplified workflow for this chapter can be seen in Figure 14. The figure only depicts the main steps of the experimental procedure, followed by a more detailed description of the methods used in this chapter.

Simplified workflow of Chapter 2

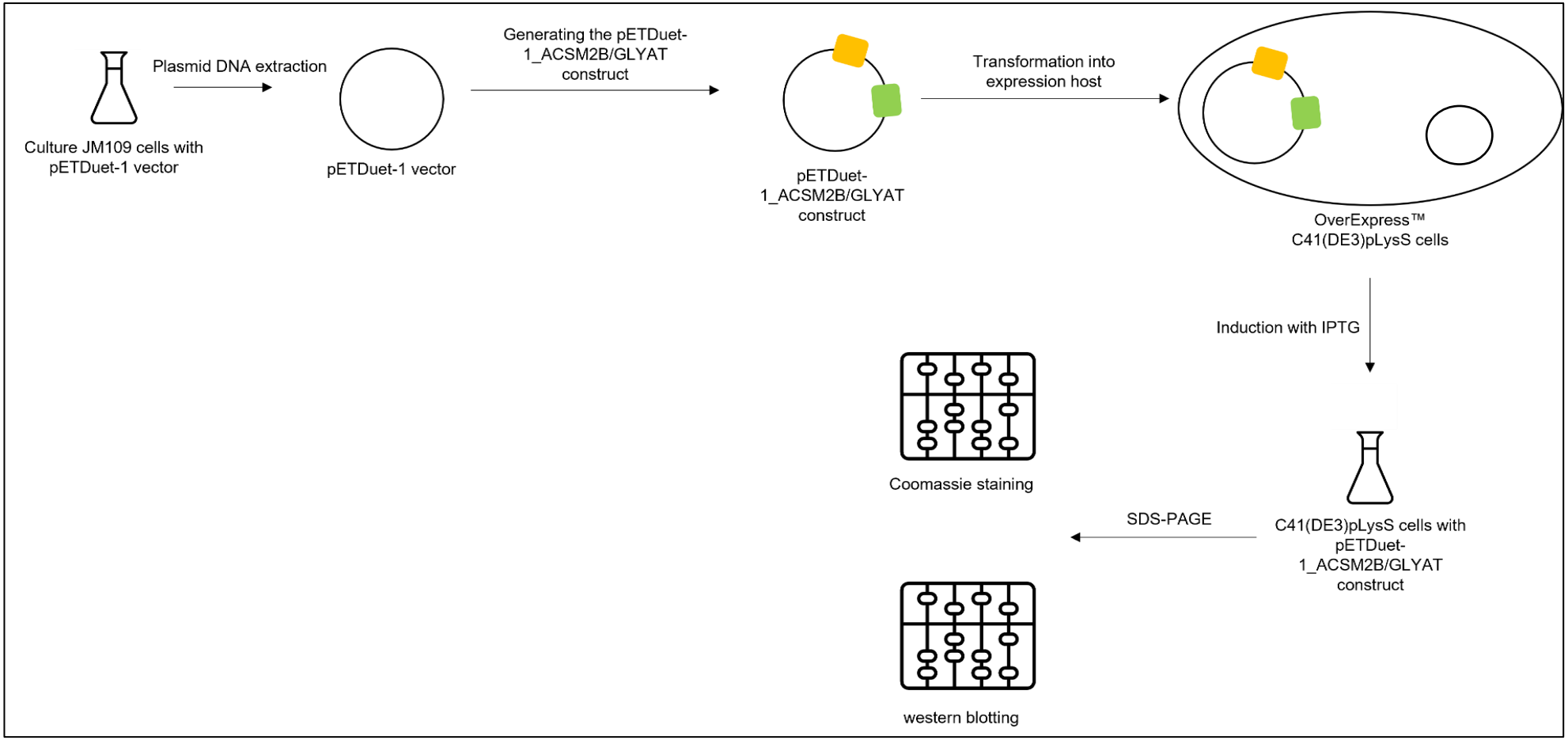


Figure 14: Simplified workflow for the methods of Chapter 2.
The orange and green rectangles on the circular plasmids represent *ACSM2B* and *GLYAT*, respectively. This image only represents the main workflow concepts of used in Chapter 2. The smaller black circle in the cell represents the pLysS plasmid.

2.2 Materials and methods

2.2.1 Materials

A table containing all the reagents needed and their manufacturers is added in Appendix A.

The pETDuet-1 vector in JM109 cells (Figure 15), pHis17_ACSM2B in DH5α cells (Figure 16), pHis17_GLYAT (Figure 17) in DH5α cells, and C41(DE3)pLysS cells were available in our laboratory as glycerol stocks stored at -80°C. Some of the general features of the plasmids in Figure 15, Figure 16, and Figure 17 were the ori (origin of replication) site, the AmpR (ampicillin resistance) promoter and gene, the T7 promoter and terminator, the 6xHis (histidine) fusion tags, and the RBS (ribosome binding site). Moreover, the unique features of Figure 15 were the MCS-1 and MCS-2 sites, the lacI promoter operator and gene, the f1 ori site, and the rop gene. These genes are important to select the organisms with the target gene as well as aid in expressing the target gene to obtain the target protein. In this case the target genes were *ACSM2B* and *GLYAT*.

The MCS-1 and MCS-2 sites are important sequences to allow for the cloning of a target gene. Because the pETDuet-1 vector contains two MCS sites, two target proteins could be expressed. The antibiotic resistant gene encodes for resistance against ampicillin or carbenicillin and is important during the selection stage after transformation. The cells that do not contain the transformed plasmid do not survive during selection with the respective antibiotic, in this case carbenicillin. Furthermore, the coding region for the proteins of interest is flanked by the bacteriophage derived T7 promoter and terminator to regulate expression of the gene of interest. The promoter further contains a ribosome binding site to allow for RNA polymerase to synthesise the respective mRNA for the target protein. The T7 promoter is also repressed by a bound repressor that will only activate gene expression during induction with an appropriate inducer molecule, in this case IPTG. When the IPTG is added to the media containing the selected cells, the inducer molecule binds to the repressor and gene transcription of the target gene can commence (Sambrook & Russell, 2001). The protein of interest is then usually expressed with either an N- or C- terminal fusion tags that are important for purification after protein expression. The fusion tags used for this study were the His-tag and S-tag. Moreover, the ori and f1 ori are the replication start sites on the plasmid and is important during plasmid replication after transformation into the host organism. The rop gene maintains the plasmid at a high copy number within the host.

Stock solutions (100 µM) for each primer were prepared by adding a specified volume of nuclease-free ddH_2O according to the manufacturer's specifications and were stored at -20°C. From the stock solution, a 10-fold dilution was prepared as a working solution at a final

concentration of 10 µM and used for subsequent PCRs. All the working solutions were stored at -20°C.

Glycerol stocks of the Stellar™ cells with the ACSM2B_pETDuet-1 vector and the C41(DE3)pLysS cells with the pETDuet-1_ACSM2B/GLYAT vector were prepared as follows: molecular biology grade water was used to prepare an 80% (v/v) glycerol working solution. The solution was stored at room temperature and used to prepare glycerol stocks for long-term storage by mixing 800 µL of cells with 200 µL 80% (v/v) glycerol working solution. The glycerol stocks were then flash-frozen with liquid nitrogen and stored at -80°C.

Stellar™ chemically competent cells for cloning (TakaraBio) and OverExpress™ C41(DE3)pLysS chemically competent cells (Sigma–Aldrich) were available in our laboratory. The Stellar™ chemically competent cells for cloning (TakaraBio) and JM109 chemically competent cells were used for cloning, and the OverExpress™ C41(DE3)pLysS chemically competent cells (Sigma–Aldrich) were used for protein expression. All competent cells were stored at -80°C.

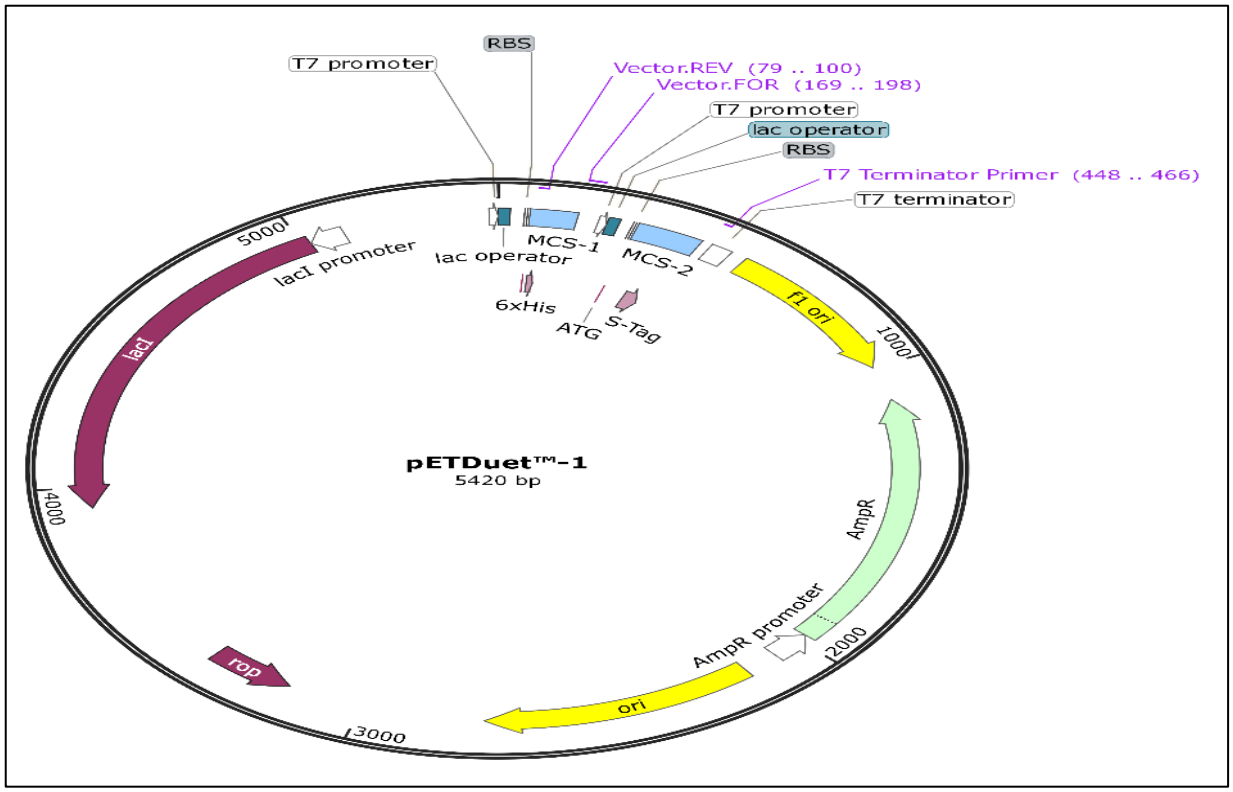


Figure 15: Plasmid map of the pETDuet-1 vector that was used for InFusion® cloning of *ACSM2B* and *GLYAT*.
The relevant features of the plasmid are the two MCS, the T7 terminator primer, the Vector.Rev and Vector_For primers, the AmpR gene, the 6xHis tag and S-tag, the bacterial ori site, and the LacI gene. This figure was compiled on Snapgene v.3.2.1.

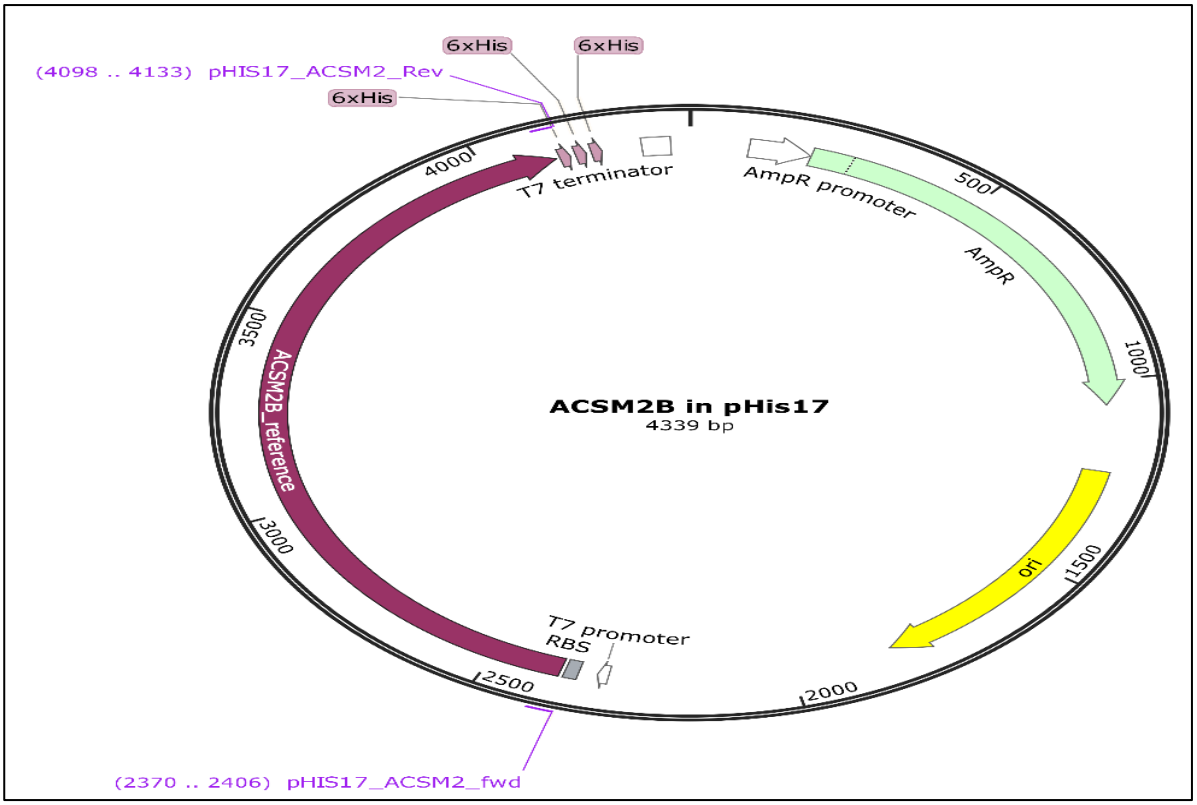


Figure 16: The pHis17 vector containing *ACSM2B*.
The relevant features were the *ACSM2B* fragment, forward and reverse primers, AmpR gene, 6xHis tags, ori site, and T7 promoter and terminator. This figure was compiled on Snapgene v. 3.2.1.

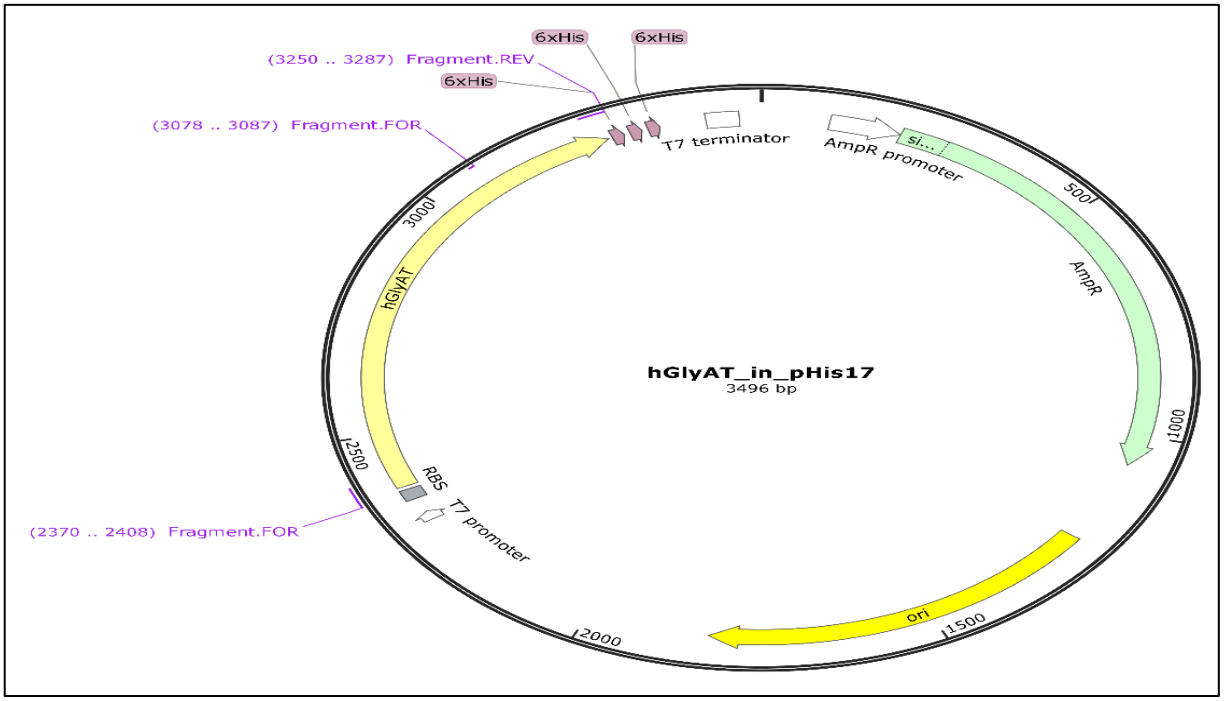


Figure 17: The pHis17 vector containing *GLYAT*.
The relevant features were the h*GLYAT*, the AmpR, the bacterial ori site, the T7 promoter and terminator, and the Fragment. Forward and Fragment.Reverse primers, and the 6xHis tags. Note that there are two primer binding sites for the Forward primer. The Forward primer at the 3078-3087, however, has a tail that does not bind 100 % to the fragment. This figure was compiled on snapgene v. 3.2.1.

2.2.2 Methods

2.2.2.1 Plasmid extraction

Plasmid DNA was either extracted using a mini preparation protocol, namely, the NucleoSpin® Plasmid DNA (NoLid) (Macherey-Nagel), or a midi preparation protocol, namely, the GenElute™ Plasmid Midiprep kit (Sigma–Aldrich). The mini and midi preparations were used to purify the plasmid DNA from JM109, Stellar™ and C41(DE3)pLysS cells containing the plasmid vectors.

Mini preparation

The NucleoSpin® Plasmid DNA (No-Lid) (Macherey-Nagel) Miniprep kit was used to purify plasmid DNA from Stellar™ and C41(DE3)pLysS cells. This kit uses silica-binding technology combined with spin columns.

Inoculations from the glycerol stocks were done into 25 mL LB-broth that contained 100 ng/µL carbenicillin and grown overnight at 37°C, shaking at 200 rpm. Chloramphenicol (34 ng/µL) was added to the LB broth if the C41(DE3)pLysS cells were used.

Ten milliliters of the cells was centrifuged at 11 000 x g for 30 seconds. After that, the supernatant was discarded. The cells were then resuspended in 500 µL Buffer A1. The pellet was resuspended by pipetting up and down. Then, 500 µL of Buffer A2 was added and mixed by inverting the tube 6-8 times and left at room temperature for 4 minutes.

Then, 600 µL of buffer A3 was added and mixed by inverting the tube 6-8 times until the sample turned colourless. The sample was then centrifuged at 11 000 x g for 10 minutes at room temperature.

While the samples were centrifuging, a NucleoSpin® Plasmid (NoLid) Column was placed into a collection tube. The samples were removed from the centrifuge, and 700 µL of the supernatant was loaded onto the prepared column. The column was then centrifuged at 11 000 x g for 1 minute. The flow-through was discarded. This step was repeated twice.

After that, the membrane was washed with 500 µL Buffer AW, preheated to 50°C, and centrifuged at 11 000 x g for 1 minute at room temperature. The flow-through was discarded. Then, 600 µL of Buffer A4 was added to the membrane and centrifuged at 11 000 x g for 1 minute. The flow-through was discarded. The column was then dried by centrifuging the tube with the column at 11 000 x g for 2 minutes. The collection tube and flow-through were discarded.

The column was placed in a new 1.5 mL microcentrifuge tube, and the plasmid DNA was eluted with 50 µL, preheated to 70°C, nuclease-free water by centrifugation at 11 000 x g for 2 minutes. The plasmid DNA samples were stored at -20°C.

Midi preparation

The GenElute™ Plasmid Midiprep kit (Sigma–Aldrich) was used to extract the pETDuet-1 plasmid from JM109 cells. This system combines silica-binding technology with spin columns. The pETDuet-1 vector DNA was extracted with this system.

Inoculation from the already present glycerol stock was done into 50 mL LB-broth containing 100 ng/µL carbenicillin and grown overnight at 37°C, shaking at 200 rpm. Forty millilitres of the cells were harvested by centrifugation at 5000 x g for 10 minutes at room temperature. The supernatant was decanted.

The pellet was resuspended in 1.2 mL of resuspension solution by pipetting up and down. Then, 1.2 mL of lysis solution was added to the resuspended cells, mixed with gentle tube inversion approximately 6-8 times and left at room temperature for 4 minutes. After the waiting period, 1.6 mL of neutralisation/binding buffer was added to the tube and inverted 6-8 times. The tube was then centrifuged at 15000 x g for 15 minutes at room temperature.

During the previous cell pelleting step, the column was prepared by adding 3 mL of Column Preparation Solution to the GenElute Midiprep DNA Binding column and centrifuging it with a swinging bucket rotor at 4000 x g for 2 minutes. The flow-through was discarded.

After the cell debris was pelleted, all the cleared lysate was added to the column and centrifuged with a swinging bucket rotor at 4000 x g for 2 minutes. The flow-through was discarded. The optional wash step was skipped.

The column was then washed by adding 3 mL of diluted wash solution to the column and centrifuged at 4000 x g for 5 minutes. The flow-through was discarded.

The GenElute Midiprep DNA Binding column was placed in a new 15 mL centrifuge tube, and 1 mL of preheated nuclease-free water was added to the column. The column was centrifuged at 4000 x g for 5 minutes. The DNA was stored at -20°C until needed.

Some of the pETDuet-1 vector DNA was sent for sequence determination as discussed in section 2.2.2.9.

2.2.2.2 Determining DNA concentration and purity

DNA concentration and purity were measured using a NanoDrop One (Inqaba™ Biotech). This technique was used for the concentration and purity measurements of the extracted

plasmid DNA, and the gel extracted linearised fragments and linearised vectors before InFusion® was performed.

This method uses the Beer–Lambert law to calculate the concentration of DNA by measuring how much light the sample absorbs. For dsDNA, a difference of 1 absorbance unit, measured at a wavelength of 260 nm, relates to 50 µg/mL DNA concentration in the sample.

The device determines each loaded sample's A_{260}/A_{280} and $A_{230}/_{A260}$ ratios. An accepted A_{260}/A_{280} value is between 1.7-1.9 for pure DNA, and the A_{230}/A_{260} value is closest to 2. An A_{260}/A_{280} value of > 1.7 indicates possible contamination with proteins, whereas a value > 2 indicates contamination with RNA.

2.2.2.3 Agarose gel electrophoresis

DNA samples were separated and visualised using agarose gel electrophoresis. A 1% (w/v) agarose (Melford, Sigma–Aldrich) gel with 1x TAE running buffer (40 mM Tris, 20 mM acetic acid, and 1 mM EDTA, pH 8.5) was used to separate the DNA samples. The electrophoresis system was set to a constant 80 V for 45-60 minutes, depending on the size of the DNA samples loaded. The O'Gene Ruler™ 1 kb DNA ladder was loaded into every first well of the gel, which served as a DNA size marker, and the Gel Loading Dye, Purple (6x) (New England Biolabs) was used as a loading buffer.

Ethidium bromide (EtBr) was used to visualise the DNA fragments. The ChemiDoc™ Imaging system (Bio-Rad) was used to image the agarose gel. ImageLab (v.5.2.1.) and ethidium bromide settings were used.

2.2.2.4 Restriction enzyme digestion of plasmids

Restriction enzyme digestion was first performed to ensure that the vectors were the correct size. Agarose electrophoresis was performed to confirm the vector sizes as discussed in section 2.2.2.3.

Before amplification and Infusion® of the fragments of interest into the vector can be done, restriction enzyme digestion is used to linearize the vectors that contain the fragments of interest and the vector in which the fragments were to be inserted.

All restriction enzyme reactions were set up according to the recommended protocol of the restriction enzyme manufacturer. Agarose gel electrophoresis was performed after restriction enzyme digestion as discussed in section 2.2.2.3 to purify the linearised fragments and linearised vectors as discussed in section 2.2.2.5.

2.2.2.5 Gel extraction and purification

Agarose gel electrophoresis was used as discussed in section 2.2.2.3 to separate the linearised fragments and vectors on the gel from the undigested/partially digested plasmids as described in section 2.2.2.4. The linearised fragments (*ACSM2B* and *GLYAT*) and the linearised vectors (pETDuet-1 and pETDuet-1_ACSM2B) can then be purified from the gel. The fragments can then be inserted into the vectors using InFusion®.

The bands were excised from the gels with a sterile blade and placed into preweighed 2 mL microcentrifuge tubes. The linearised DNA was then recovered by a silica-binding technique combined with spin columns.

The NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel) was used to extract the linear DNA from the gel pieces according to the manufacturer's protocol. The gel pieces were weighed, and 200 µL of NTI buffer was added per 100 mg gel. The samples were incubated at 50°C for 10 minutes with a brief vortex step every 2 minutes.

A NucleoSpin® Gel and PCR Clean-up column and collection tube were assembled, and 700 µL of the sample was loaded onto the membrane. The sample was then centrifuged at 11 000 x g for 1 minute. The flow-through was discarded and placed back into the collection tube. This step was repeated if there were more sample volumes left.

The silica membrane was washed by adding 700 µL Buffer NT3 onto the column and centrifuging it at 11 000 x g for 30 seconds. The flow-through was discarded. After that, the silica membrane was dried by centrifuging the tube at 11 000 x g for 1.5 minutes.

The linearized DNA was then eluted by adding 30 µL nuclease-free water (preheated to 70°C) directly onto the silica membrane. The membrane was incubated at room temperature for 5 minutes and then centrifuged at 11 000 x g for 1.5 minutes.

Agarose gel electrophoresis was performed as discussed in section 2.2.2.3 with some of the extracted plasmid DNA to confirm the extraction of the fragments and linearised vectors from the agarose gels. The rest of the sample was used for quality and concentration determination after confirmation. The quality and concentration of the DNA fragments and vectors were determined as discussed in section 2.2.2.2. All linear DNA samples were stored at -20°C until needed.

2.2.2.6 Preparation of competent cells

Competent C41(DE3)pLysS cells were prepared using the DMSO method (Chung *et al.*, 1989).

C41(DE3)pLysS cells were streaked on an LB (Luria-Bertani)-agar plate (1% tryptone (w/v), 0.5% yeast extract (w/v), 1% NaCl (w/v), with 15 agar (Sigma–Aldrich) capsules per 1 L) without antibiotics from a glycerol stock and grown overnight at 37°C. The following day, 5 mL of LB broth (1% tryptone (w/v), 0.5% yeast extract (w/v), 1% NaCl (w/v), per 1 L distilled water) (without antibiotics) was inoculated with a single colony and grown overnight at 37°C with shaking at 225 rpm.

Fifty millilitres of LB broth (without antibiotics) was inoculated with 1.2 mL of the 5 mL overnight culture and grown to an OD_{600} ~ 0.5. The culture was then centrifuged at 4000 rpm at 4°C for 10 minutes. The supernatant was decanted, and 5 mL ice-cold TSS (Transformation and storage solution) (LB-broth with 10% (w/v) PEG (molecular weight 3350 or 8000), 5% (v/v) DMSO, and 20-50 mM Mg^{2+} at a final pH of 6.5) was added.

The cells were then incubated on ice for at least 2 hours. Glycerol was added to a final concentration of 15% to make a glycerol stock of the competent cells. The competent cells were then aliquoted as 200 µL competent C41(DE3)pLysS cells in 2 mL microcentrifuge tubes, snap-frozen with liquid nitrogen, and stored at -80°C until needed.

2.2.2.7 Transformation of chemically competent Stellar™ and C41(DE3)pLysS cells

The transformation protocol followed was that of the Stellar™ chemically competent cells for cloning (TakaraBio). The transformation protocol was the same for the C41(DE3)pLysS cells.

The competent cells were removed from the -80°C storage and thawed on ice for approximately 5-10 minutes. While the cells were thawing, the SOC media (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4.7H_2O$, and 20 mM glucose) was placed in the 37°C incubator to heat up. Fifty microliters of the thawed competent cells were then transferred to a 14-mL round-bottom tube and mixed with > 5 ng vector DNA. The mixture was gently stirred after mixing the competent cells and vector DNA.

The tubes were placed on ice for 30 minutes and heat-shocked for 45 seconds at 42°C. After heat-shocking the cells, the tubes were placed on ice for 1-2 minutes. SOC (Super optimal broth with catabolite repression) media was added to the tube to a final volume of 500 µL and incubated at 37°C with shaking at 225 rpm.

LB agar plates that contained 100 ng/µL carbenicillin were used to make spread plates. One hundred microliters of the transformation mixture was spread onto each plate and grown overnight at 37°C.

From the LB-agar plates that contain the selective antibiotics, 6 colonies were chosen at random to verify the presence of the inserts into the vector. Single colonies were selected from

the plates and were then inoculated into 50 mL LB broth containing 100 ng/µL carbenicillin for the Stellar® cells containing the ACSM2B_pETDuet-1 vector and into 50 mL LB broth containing 100 ng/µL carbenicillin and 34 ng/µL chloramphenicol for the C41(DE3)pLysS cells with pETDuet-1_ACSM2B/GLYAT. The cells were grown overnight at 37°C with shaking at 200 rpm. For long-term storage, glycerol stocks were made, as discussed in section 2.2.1.

2.2.2.8 Generating the pETDuet-1_ACSM2B/GLYAT construct

The pETDuet-1_ACSM2B/GLYAT vector was not available in our lab. This vector had to be created for this study). The InFusion® method requires specially designed primers with 15 bp overhangs complementary to the pETDuet-1 vector. *ACSM2B*, *GLYAT* and the pETDuet-1 vector were then amplified, and the fragments were inserted one at a time into the pETDuet-1 vector using InFusion®. Figure 18 indicates the simplified workflow that was followed to construct the pETDuet-1_ACSM2B/GLYAT construct.

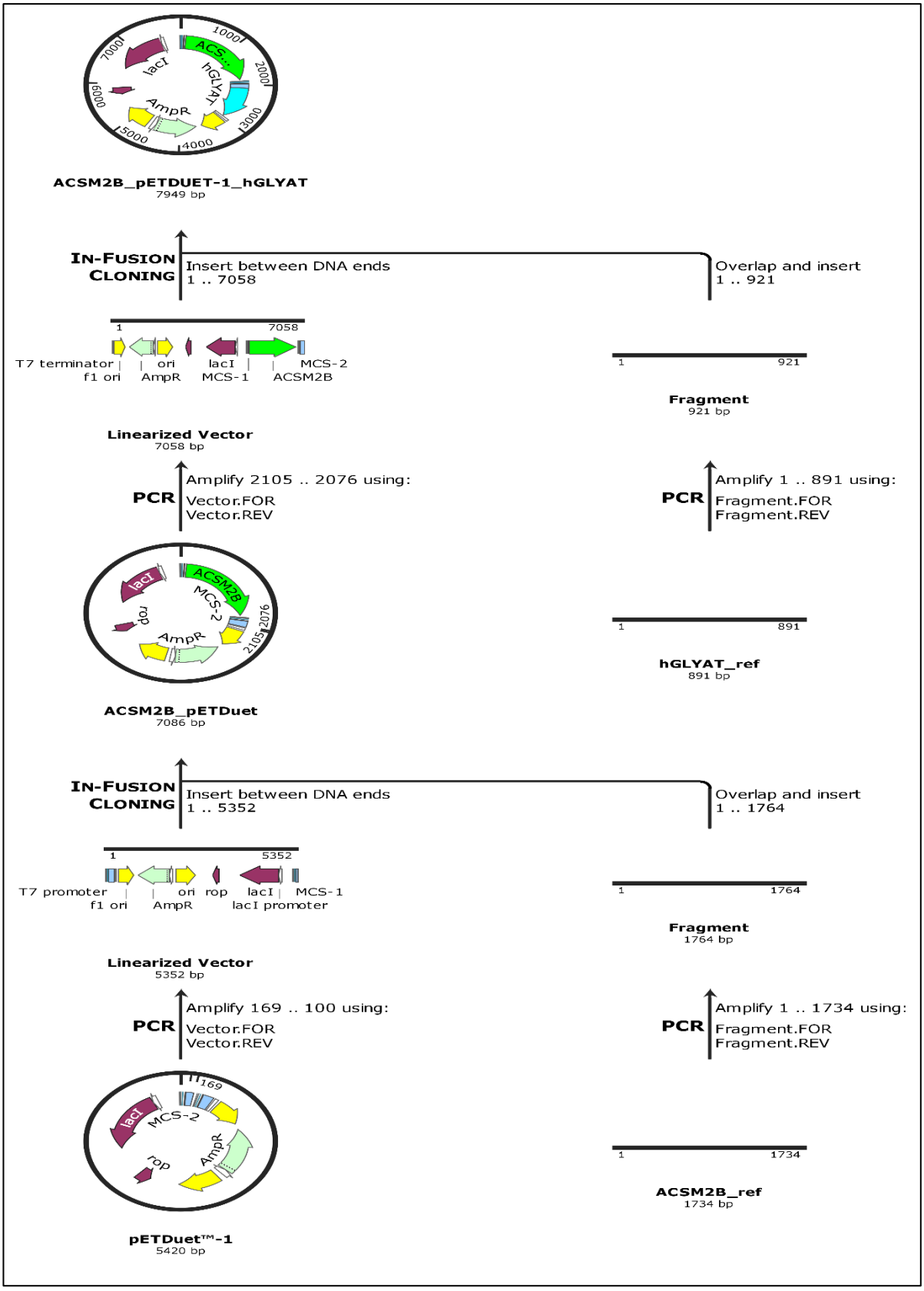


Figure 18: A simplified depiction of both InFusion® reactions to insert *ACSM2B* then *GLYAT* into the pETDuet-1 vector.
The image represents all the steps explained in section 2.2.2.8 read from the bottom to the top. The starting vector is the pETDuet-1 vector and the last plasmid is the generated pETDuet-1_ACSM2B/GLYAT construct. The dark green arrow represents *ACSM2B,* whereas the light blue arrow represents *GLYAT*.

2.2.2.8.1 Primer design for InFusion® cloning

Primers were designed with 15 bp overhangs complementary to the pETDuet-1 vector to amplify the fragments for InFusion® reactions. An additional primer was also designed to allow for sequencing of the construct to confirm that the *ACSM2B* gene was correctly inserted into the pETDuet-1 vector. This primer was the pETDuet upstream primer (Inqaba biotec™). All but one of the primers were ordered from Integrated DNA Technologies®.

The InFusion® cloning setting on Snapgene® (v.3.2.1) was used to simulate the insertion of *ACSM2B* and *GLYAT* into pETDuet-1 to design the primers (Table 9). The primer names are denoted as *binding strand*.*direction.target gene*. The underlined part represents the 15 bp overhang complementary to the vector to facilitate fragment insertion.

Table 9: The primers used for sequencing and fragment and vector amplification.

Primer name	Primer sequence
Fragment.For.*ACSM*2B	5'-CACCATCATCACCACATGCATTGGCTGCGAAAAGTTCAG - 3'
Fragment.Rev.*ACSM*2B	5'-TTACTTTCTGTTCGATCACTGCGCACGGGCTTTTC-3'
Vector.For.*ACSM*2B	5'-TCGAACAGAAAGTAATCGTATTGT-3'
Vector.Rev.*ACSM*2B	5'-GTGGTGATGATGGTGATGGC-3'
Fragment.For.h*GLYAT*	5'-CAGCACATGGACTCGATGATGTTACCATTGCAAGGTGC-3'
Fragment.Rev.h*GLYAT*	5'-CAGCGGTGGCAGCAGTCACAGAGGTACACAGTTCC-3'
Vector.For.h*GLYAT*	5'-CTGCTGCCACCGCTGAGC-3'
Vector.Rev.h*GLYAT*	5'-CGAGTCCATGTGCTGGCG-3'
Vector.For.pETDuet_upstream	5'-ATGCGTCCGGCGTAGA-3'

2.2.2.8.2 Restriction enzyme digestion of pETDuet-1

The circular pETduet-1 vector needed to be linearized to amplify the pETDuet-1 vector and create the correct 5' and 3' ends for downstream InFusion® cloning with *ACSM2B*. pHis17_GLYAT and pHis17_ACSM2B also needed to be linearized to amplify *GLYAT* and *ACSM2B* from these vectors with specific primers designed for InFusion® to create 15 bp overhangs complementary to the pETDuet-1 vector.

Restriction enzyme digestion was performed to linearize the pETDuet-1, pHis17_GLYAT and pHis17_ACSM2B vectors according to the conditions in Table 10.

Table 10: The restriction enzymes and conditions used for linearizing the listed vectors.

Target plasmid	Restriction enzyme	Restriction site	Buffer	Incubation temperature (°C)	Inactivation temperature (°C)	Incubation time (Hours)
pETDuet-1	BamHI-HF	G/GATCC	Cutsmart®	37	-	24
pHis17_GLYAT	NdeI	CA/TATG	Buffer 2.1	37	65	24
pHis17_ACSM2B	NdeI	CA/TATG	Buffer 2.1	37	65	24

2.2.2.8.3 PCR amplification of the linearized pETDuet-1 vector

A PCR was set up according to the CloneAmp™ HiFi PCR Premix protocol (TakaraBio). The linearised DNA, created with restriction enzyme digestion, as discussed in section 2.2.2.8.2, was used as the template for the PCRs.

The reagents and concentrations used for pETDuet-1 amplification can be seen in Table 11, whereas the thermocycler conditions can be seen in Table 12.

Table 11: Basic PCR setup using CloneAmp™ HiFi PCR Premix.

Reagent	Concentration	Volume (µL)
Hifi CloneAMP premix	1X	12.5
Vector.For.*ACSM2B*	5-7.5 pmol (0.2-0.3 µM)	1
Vector.Rev.*ACSM2B*	5-7.5 pmol (0.2-0.3 µM)	1
ddH_2O (nuclease-free)	-	9.5
DNA template	< 100 ng/µl	1
Total	-	25

Table 12: Thermocycler conditions for amplifying the linearised pETDuet-1 vector backbone.

Step	Cycles	Time (sec)	Temperature (°C)
Initial denaturation	1	60	95
Denaturation	30	30	98
Annealing		15	60
Extension		28	72
Final extension	1	300	72
Storage	-	∞	4

Agarose gel electrophoresis was performed as discussed in section 2.2.2.3 to verify the amplification of the linearised pETDuet-1 vector. Gel extraction was performed to retrieve the linear pETDuet-1 vector as discussed in section 2.2.2.5. After that, the concentration and purity of the linear pETDuet-1 vector were measured as discussed in section 2.2.2.2. The linear pETDuet-1 vector was stored at -20°C until needed.

2.2.2.8.4 PCR optimisation of the *ACSM2B* and *GLYAT* fragments

A temperature gradient was performed to amplify *ACSM2B* from the linear pHis17_*ACSM2B* and *GLYAT* from the linear pHis17_GLYAT created in section 2.2.2.8.2. The temperature gradient ranged between 55-57°C for *ACSM2B* and 52-65°C for *GLYAT*. The reagents used for the PCR amplification of *ACSM2B* are shown in Table 13, and the thermocycler setup is shown in Table 14. In contrast, the reagents used for *GLYAT* amplification can be seen in Table 15, and the thermocycler conditions can be seen in Table 16.

Table 13: PCR setup for the amplification of *ACSM2B*.

Reagent	Concentration	Volume (µL)
Hifi CloneAMP premix	1X	12.5
Fragment.For.*ACSM2B*	5-7.5 pmol (0.2-0.3 µM)	1
Fragment.Rev.*ACSM2B*	5-7.5 pmol (0.2-0.3 µM)	1
ddH_2O (nuclease-free)	-	9.5
DNA template	< 100 ng/µl	1
Total	-	25

Table 14: Thermocycler conditions for the amplification of *ACSM2B*.

Step	Cycles	Time (sec)	Temperature (°C)
Initial denaturation	1	60	95
Denaturation	30	30	98
Annealing		5	55-57
Extension		6	72
Final extension	1	300	72
Storage	-	∞	4

Table 15: PCR setup for the amplification of *GLYAT*.

Reagent	Concentration	Volume (µL)
Hifi CloneAMP premix	1X	12.5
Fragment.For.*hGLYAT*	5-7.5 pmol (0.2-0.3 µM)	1
Fragment.Rev.*hGLYAT*	5-7.5 pmol (0.2-0.3 µM)	1
ddH_2O (nuclease-free)	-	9.5
DNA template	< 100 ng/µl	1
Total	-	25

Table 16: Thermocycler conditions for the amplification of *GLYAT*.

Step	Cycles	Time (sec)	Temperature (°C)
Initial denaturation	1	60	95
Denaturation	30	30	98
Annealing		5	52-65
Extension		5	72
Final extension	1	300	72
Storage	-	∞	4

After the gradients were performed, agarose gel electrophoresis was used to verify the *ACSM2B* and *GLYAT* amplification, as discussed in section 2.2.2.3. The best temperature to use was thus 55.1°C for *ACSM2B* and 52°C for *GLYAT*. The amplified *ACSM2B* and *GLYAT* fragments were extracted from the agarose gel as discussed in section 2.2.2.5, and the concentration and purity were determined as discussed in section 2.2.2.2.

2.2.2.8.5 InFusion® of *the ACSM2B* fragment into the linearized pETDuet-1 vector

Infusion® cloning of the *ACSM2B* fragment into the linearised pETDuet-1 vector (MCS-1) was performed according to the instructions of the InFusion® HD EcoDry™ Cloning Kit (TakaraBio). A positive control (pUC19 and the control insert) and negative control (only the pETDuet-1 vector) were included in the experiment. The reaction setup can be seen in Table 17.

Table 17: Infusion reaction setup.

Reaction component	Cloning reaction	Negative control	Positive control
Purified PCR fragment	10-100 ng DNA	-	2 µL of 2 kb control insert
Linearised vector	50-100 ng DNA	1 µL	1 µL of pUC19 control vector
Deionised water	Fill to 10 µL	Fill to 10 µL	Fill to 10 µL

The cloning reactions were started by adding the purified *ACSM2B* fragment to the linear pETDuet-1 vector DNA in a 2:1 ratio. Nuclease-free ddH_2O was added to a final volume of 10 µL.

After that, the control reactions were set up according to Table 17. Each 10 µL mixture was then added to a separate In-Fusion HD EcoDry pellet and mixed well by pipetting up and down. The samples were then incubated at 37°C for 15 minutes and then at 50°C for 15 minutes. Thereafter, the samples were placed on ice. If the reactions were not used immediately, they could be stored at -20°C.

After the samples were placed on ice, they were immediately transformed into Stellar™ competent cells, as discussed in section 2.2.2.7. The transformants were verified by making spread plates of the transformants on LB agar plates that contained 100 ng/µL carbenicillin and 34 ng/µL chloramphenicol. A mini preparation was performed to purify the pETDuet-1_ACSM2B construct from the Stellar™ cells, as discussed in section 2.2.2.1. The pETDuet-1_ACSM2B plasmid DNA was stored at -20°C.

PCR was performed with the extracted pETDuet-1_ACSM2B as template DNA to amplify *ACSM2B* and confirm that *ACSM2B* is present in the pETDuet-1_ACSM2B construct. Agarose gel electrophoresis was performed to confirm the pETDuet-1_ACSM2B construct size. Sequence determination was performed to confirm that *ACSM2B* was present in the pETDuet-1 vector after gel extraction, as discussed in sections 2.2.2.5 and 2.2.2.9.

The pETDuet-1_ACSM2B construct was then used in a restriction enzyme digestion reaction (Table 18) to linearize the construct to amplify the pETDuet-1 vector to create the correct 5' and 3' ends for downstream InFusion® cloning with *GLYAT*.

Table 18: The reaction setup and conditions for linearizing the pETDuet-1_ACSM2B vector.

Target plasmid	Restriction enzyme	Restriction site	Buffer	Incubation temperature (°C)	Inactivation temperature (°C)	Incubation time (Hours)
pETDuet-1_ACSM2B	XmaJI(AvrII)	C/CTAGG	10X Buffer Tango	37	80	24

Gel extraction was performed as discussed in section 2.2.2.5 to retrieve the linear pETDuet-1_ACSM2B construct.

2.2.2.8.6 PCR amplification of the linearized pETDuet-1_ACSM2B construct

PCR was set up according to the CloneAmp™ HiFi PCR Premix protocol. The linearized pETDuet-1_ACSM2B was amplified to create the correct 5' and 3' vector ends for downstream InFusion® cloning with *GLYAT*.

The reagents and concentrations used can be seen in Table 19. The thermocycler conditions used for amplifying the pETDuet-1_ACSM2B construct can be seen in Table 20.

Table 19: PCR setup for the amplification of the pETDuet-1_ACSM2B vector.

Reagent	Concentration	Volume (µL)
Hifi CloneAMP premix	1X	12.5
Vector.For.*hGLYAT*	5-7.5 pmol (0.2-0.3 µM)	1
Vector.Rev.*hGLYAT*	5-7.5 pmol (0.2-0.3 µM)	1
ddH_2O (nuclease-free)	-	9.5
DNA template	< 100 ng/µl	1
Total	-	25

Table 20: Thermocycler conditions for the amplification of the pETDuet-1_ACSM2B vector.

Step	Cycles	Time (sec)	Temperature (°C)
Initial denaturation	1	60	95
Denaturation	30	30	98
Annealing		5	60
Extension		5	72
Final extension	1	300	72
Storage	-	∞	4

Agarose gel electrophoresis was used to confirm the amplification of the linearised pETDuet-1_ACSM2B construct, as discussed in section 2.2.2.3. Gel extraction was then performed to retrieve the linear amplified pETDuet-1_ACSM2B construct, as discussed in section 2.2.2.5. The plasmid DNA was stored at -20°C.

2.2.2.8.7 InFusion® cloning of *GLYAT* into the pETDuet-1_ACSM2B vector

Infusion® of the purified *GLYAT* fragment into linear pETDuet-1_ACSM2B was performed exactly as discussed in section 2.2.2.8.5, with the inserted gene being *GLYAT*.

After the samples were placed on ice, they were immediately transformed into OverExpress™ C41(DE3)pLysS competent cells, as discussed in section 2.2.2.7. The transformants were verified by making spread plates of the transformants on LB agar plates that contained 100 ng/µL and 34 ng/µL chloramphenicol. A mini preparation was performed to purify the pETDuet-1_ACSM2B/GLYAT construct from the C41(DE3)pLysS cells, as discussed in section 2.2.2.1. Glycerol stocks were made as discussed in section 2.2.1 with C41(DE3)pLysS cells that contained the pETDuet-1_ACSM2B/GLYAT construct.

Gel extraction was performed as discussed in section 2.2.2.5 to retrieve only the pETDuet-1_ACSM2B/GLYAT plasmid DNA for sequencing determination of *ACSM2B* and *GLYAT* to confirm that both genes are present in the vector.

Two PCRs were also performed with the extracted pETDuet-1_ACSM2B/GLYAT as template DNA to amplify *ACSM2B* and *GLYAT*, as discussed in section 2.2.2.8.4. Agarose gel electrophoresis was performed to confirm the gene sizes and confirm that both genes were in the vector, as discussed in section 2.2.2.3.

pETDuet-1_ACSM2B/GLYAT was stored at -20°C until needed.

2.2.2.9 DNA sequencing and data analysis

Verified plasmids were sent to the Central Analytical Facility (CAF) of the University of Stellenbosch for DNA sequence determination using the Sanger method. This was done to ensure that *ACSM2B* was correctly inserted into the pETDuet-1 vector in the first InFusion® cloning and that *GLYAT* was correctly inserted into the pETDuet-1_ACSM2B vector in the second InFusion® cloning. Sequence determination was also used to determine whether the genes were inserted in the correct orientation without any mutations being introduced.

The primers used for sequence determination can be seen in Table 21.

Table 21: The sequence determination primers for *ACSM2B* and *GLYAT*.

Gene	Primers used
ACSM2B	Vector.For.pETDuet_upstream
	Fragment.For.*ACSM2B*
	Fragment.Rev.*ACSM2B*
	Vector.Rev.*hGLYAT*
GLYAT	Vector.For.*ACSM2B*
	Fragment.For.*hGLYAT*
	Fragment.Rev.*hGLYAT*
	*T7 terminator

*The primer sequences can be seen in Table 9. The T7 terminator primer provided by the sequencing facility of CAF: Stellenbosch was used to sequence the end of *GLYAT* in the pETDuet-1_ACSM2B/GLYAT construct. The sequence of the T7 terminator primer is 5'-GCTAGTTATTGCTCAGCGG-3'.

Before the plasmid DNA was sent to CAF, the concentration and purity of the samples were determined, as discussed in section 2.2.2.2. CAF requires DNA samples to be > 100 ng/µL, and when the samples were < 100 ng/µL, the samples were concentrated using the SpeedVac concentrator (Thermo Scientific, Waltham, MA, USA).

The sequence chromatograms provided by CAF were analysed using FinchTv (v.1.4.0.) and aligned to the reference sequences with MEGAX (v.10.2.2.). These reference sequences were NM_001105069.2 for *ACSM2B* and NM_201648.3 for *GLYAT*. This gene variant for *GLYAT* was used in the Matsuo et al. (2012) study, whereas the gene variant for *ACSM2B* was used in the van der Sluis (2018) study. These gene variants were chosen due to having the highest haplotype frequencies observed. This variant for *GLYAT* also expressed GLYAT with the highest enzyme activity.

2.2.2.10 Coexpression of ACSM2B and GLYAT

Cells were taken from the glycerol stocks made in section 2.2.2.8.7 of the OverExpress® C41(DE3)pLysS that contained the pETDuet-1_ACSM2B/GLYAT construct and were inoculated into 25 mL of LB-broth.

The LB broth contained 34 ng/µl chloramphenicol, 100 ng/µl carbenicillin, 0.2% glucose (w/v), 0.5% glycine (w/v) and 5 mM pantothenic acid. The cells were grown overnight at 37°C with shaking at 200 rpm. The next day, 100 mL of LB broth containing 34 ng/µL chloramphenicol and 100 ng/µL carbenicillin was inoculated with 1 mL of the overnight culture. The cells were grown at 37°C until they reached an OD_{600} value of 0.8-1.0.

Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the cells at a final concentration of 1 mM. The cells were then induced overnight at 25°C with shaking at 200 rpm. The next day, the cells were placed on ice for 10 minutes, and then 50 mL of cells was centrifuged at 5000 x g for 10 minutes at 4°C to confirm the expression of the proteins. The other 50 mL of induced cells was also centrifuged at 5000 x g for 10 minutes at 4° for GC–MS analysis. The supernatant was removed and used in cell lysis, as discussed in section 2.2.2.11.

Two negative controls (the empty cells containing only the pLysS plasmid and the uninduced sample) were added and used for downstream analysis. No IPTG was added to the negative controls.

Figure 19 depicts a simplified workflow for sections 2.2.2.10-2.2.2.13.

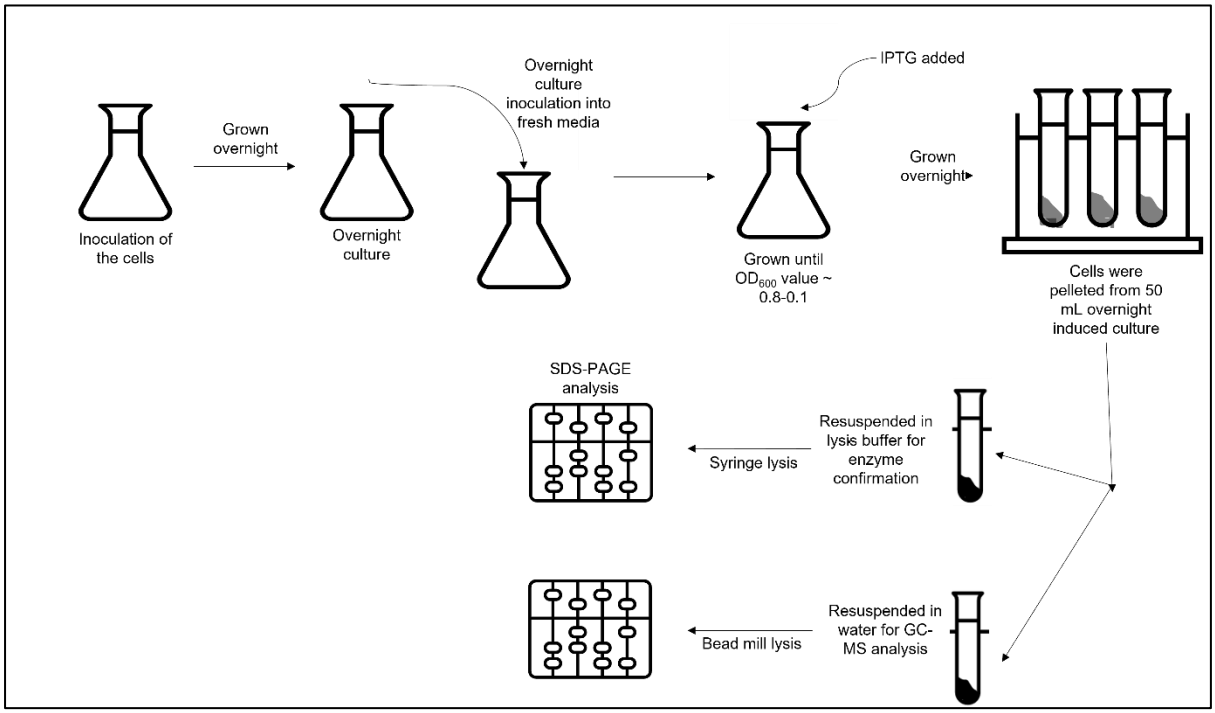


Figure 19: Simplified workflow for the experimental procedure in sections 2.2.2.10-2.2.2.13.

2.2.2.11 Cell lysis and lysate preparation for confirmation of the expression of soluble ACSM2B and GLYAT

A lysis buffer was prepared according to van der Sluis *et al.* (2013) by mixing 50 mM NaH_2PO_4 and 300 mM NaCl (pH 8.0). The solution was filter-sterilized using a sterile 0.8/0.2 µm Acrodisc PF Syringe Filter. After filtering the solution, 10% glycerol (v/v) and 1% Triton X-100 (v/v) were added. The lysis buffer was then stored at 4°C.

The supernatant was discarded from the tube, and the pellet was resuspended in 5 mL of lysis buffer. One microliter of Pierce® Universal Nuclease For Cell Lysis (5 kU) was added to the cell resuspension. The resuspended cells were placed on ice for 10 minutes before lysis with

a 22-G needle and 50 mL syringe. The cells were passed through the needle 7-9 times for cell disruption. The 22-G needle and syringe were washed with absolute ethanol and rinsed with ddH_2O to minimise cross-contamination between the samples.

The lysed cells were used as the total fraction, and 5 µL was then removed and placed into 500 µL Eppendorf® tubes that were placed in the 4°C refrigerator until needed. The rest of the lysed cells were centrifuged at 15000 x g for 20 minutes at 4°C. The supernatant fraction without cellular debris served as the soluble fraction. After centrifugation, the supernatant was removed from each sample and placed into a new 1.5 mL microcentrifuge tube, with care not to disturb the pellet. The soluble fraction was stored at 4°C if not used immediately.

The second lysis technique that was done was the vibration mill technique. This technique was employed as alternative should the buffers used in the needle and syringe technique interfere with later protein analysis on the GC-MS.

The prepared cell pellet was resuspended in 5 mL double distilled water. Glass beads were then added in a 2:1 (lysate: glass beads) ratio into 2 mL microcentrifuge tubes. For example, 1 mL resuspended pellet was added to ~500 µL glass beads. 1 µL of Pierce® Universal Nuclease For Cell Lysis (5 kU) was added to the resuspended cells.

The 2 mL microcentrifuge tubes were then placed into a Retsch (MM400) vibration mill for 40-50 minutes at a frequency of 30.0 (1/s). After lysis, the cells were then left for 20 minutes for further cell disruption. Thereafter, 5 µL of the total fraction was removed from the lysate and stored at 4°C until needed. The soluble fraction was obtained by centrifuging the lysate at 15000 x g for 20 minutes at 4°C. The clear soluble fractions were placed into new 1.5 mL microcentrifuge tubes and stored at 4°C if not used immediately.

2.2.2.12 Sodium dodecyl sulfate – Polyacrylamide gel electrophoresis (SDS–PAGE)

SDS–PAGE was used to confirm the successful expression, purification, and concentration of the proteins. The protocol used was adapted from the method described by Sambrook and Russell (2001) and Laemmli (1970).

Unless otherwise stated, 10% SDS–PAGE gels were made with the method of Sambrook and Russell (2001). To make a 1 x SDS–PAGE resolving gel, the reagents in Table 22 were mixed and pipetted into the vertical plates of the casting setup, up to 3 cm from the top of the glass plates.

Table 22: Reagent setup for casting 1 SDS–PAGE resolving gel.

Isopropanol was layered on top of the gel mixture, and the gels were allowed to solidify for 30-40 minutes. Filter paper was then used to remove the isopropanol layer, and the top part was rinsed with distilled water. The 5% stacking gel was made by mixing the reagents in Table 23. This amount is enough for 1 x SDS–PAGE stacking gel.

Reagent	Concentration	Volume (mL)
H_2O	-	0.68
Acrylamide mix	30%	0.17
Tris (pH 8.8)	1.0 M	0.13
SDS	10%	0.01
Ammonium persulfate	10%	0.01
TEMED	-	0.001
Total	-	1

Table 23: Reagent setup for casting 1 SDS–PAGE stacking gel.

Reagent	Concentration	Volume (mL)
H_2O	-	1.9
30% Acrylamide mix	-	1.7
Tris (pH 8.8)	1.5 M	1.3
SDS	10%	0.05
Ammonium persulfate	10%	0.05
TEMED	-	0.002
Total	-	5

The mixture was then pipetted on top of the resolving gel. A comb was placed inside the mixture to form the wells and was allowed to solidify for 30-40 minutes.

The gels in the glass plates were placed into the Bio-Rad SDS–PAGE running chamber. The chamber was then filled with 1x TGS (25 mM Tris-HCl, 192 mM glycine, and 0.1% SDS, pH 8.0) and added as a running buffer. The Precision Plus Protein™ Dual Colour Standards (10 kDa – 250 kDa) were used as a molecular weight marker, and 4 µL was loaded into every 1st well of the SDS–PAGE gels.

A fresh Laemmli buffer stock solution was prepared by adding the required amount of β-mercaptoethanol to the Laemmli buffer according to the manufacturer's suggestions on the container. Five microliters of Laemmli buffer stock solution was added to 10 µL of total protein sample (total fraction), and 15 µL of Laemmli buffer stock solution was added to 30 µL of soluble protein sample (soluble fraction). The samples were then boiled at 98°C for 5 minutes

and cooled. The samples are thus ready to be loaded into the wells of the SDS–PAGE gel. The gels were run at a constant 130 V for 75-80 minutes.

The gel was carefully removed from the glass plate and placed in a separate container. Approximately 40 mL of a Coomassie Brilliant Blue solution (1 g Coomassie Brilliant Blue R-250, 100 mL glacial acetic acid, 100 mL isopropanol, and 800 mL distilled water) was poured over the gel until completely covered and left overnight at room temperature, shaking at 50 rpm.

The next day, the Coomassie Brilliant Blue staining solution was decanted. The gel was destained with a destaining solution (100 mL glacial acetic acid, 100 mL isopropanol, and 800 mL distilled water). Approximately 30 mL of destaining solution was poured over the stained gel and placed at 37°C with shaking at 50 rpm. This step was repeated 4-5 times until the stained gel was completely destained. After destaining, the gel was covered with water to rehydrate it.

The ChemicDoc™ MP Imaging system (Bio-Rad) was used to image the destained SDS–PAGE gels. The white background was inserted into the system before imaging the destained gel. The software used to image the gels was ImageLab (v.5.2.1.). The Coomassie setting was selected along with the faint band setting.

2.2.2.13 SDS–PAGE western blot procedure

The gels for western blotting were made exactly as discussed in section 2.2.2.12. A western blot stack was prepared according to the Abcam general western blot protocol (https://www.abcam.com/protocols/general-western-blot-protocol). The western blot stack was prepared as follows: one western blot sponge, one blotting paper (MN 218 B, Macherey-Nagel), one supported nitrocellulose membrane (Cat. #162-0094, Bio Rad), the unstained SDS–PAGE gel was then placed on the nitrocellulose membrane, one blotting paper (MN 218 B, Macherey-Nagel), and one western blot sponge. The sponges and blotting paper were soaked in Towbin transferring buffer (3.03 g Tris base, 14.4 g glycine, 200 mL methanol, and 500 mL double distilled water). The buffer was brought up to 1 L, and the pH was set to 8.1-8.5.

The stack was placed into the western blotting cassette, which was then placed into the Trans-Blot® Turbo™ Transfer System (Bio-Rad). The proteins were transferred for 25 minutes at a constant 25 V.

After transferring the proteins to the nitrocellulose membrane (Bio-Rad), the stack was removed from the cassette and placed into a blocking buffer (1% Caesine in 1 x TBS buffer, Bio-Rad). The container with the gel was placed at 4°C overnight while shaking at 60 rpm.

The following day, the blocking buffer was decanted, and the nitrocellulose membrane containing the proteins was rinsed thrice with TNT buffer (0.05% Tween-20, 0.2 M NaCl, 0.05 M Tris-HCl, adjusted to 1 L). Primary antibodies were added to the TNT buffer at a 1:5000 ratio.

After rinsing the blocked nitrocellulose membrane, HRP-labelled polyclonal goat anti-S-tag antibody and HRP-labelled monoclonal mouse anti-His-tag antibody were added to the blocked nitrocellulose membrane. The antibodies were left to bind to the His-tagged proteins for at least 4-6 hours, shaking at 40-50 rpm.

The antibody solution was removed, and the nitrocellulose membrane was rinsed again with TNT buffer. After rinsing the nitrocellulose membrane, AEC solution (3-amino-9-ethylcarbazole) (Rodig, 2019) was poured over the nitrocellulose membrane. The AEC solution consisted of 0.4% AEC added to a 0.1 M sodium acetate solution. Hydrogen peroxide (0.15 mL, 12%) was added to the AEC solution. The solution was then filtered with filter paper before use.

The nitrocellulose membrane was then allowed to develop for 40-60 minutes at room temperature with shaking at 40-50 rpm. Distilled water was added to stop the reaction and development.

The ChemiDoc™ MP Imaging system (Bio-Rad) was used to image the nitrocellulose membrane. The white background was placed into the system before imaging. ImagingLab (v.5.2.1.) was used with the colorimetric setting along with the faint band setting.

2.3 Results and discussion of the pETDuet-1 expression system

2.3.1 Confirmation of the sizes of the vectors used in this study

All the vectors used in the pETDuet-1 system were run on one agarose gel, as discussed in section 2.2.2.3, to confirm each vector's size before use in downstream experiments. These included the unlinearized vectors from the DNA purification steps and the restriction enzyme digested vectors (linear vectors) before gel extraction. The vectors are pETDuet-1, pHis17_ACSM2B, and pHis17_GLYAT, as discussed in section 2.2.1. The pETDuet-1 vector was digested with the restriction enzyme BamHI-HF, whereas pHis_ACSM2B and pHis_GLYAT were digested with NdeI, as discussed in section 2.2.2.8.2. Each of the linear fragments was then compared to the molecular weight ladder.

For ease of comparison, the unlinearised vector was loaded first, followed by the corresponding linearised vector. The restriction enzyme digested vectors were used for size comparison since the unlinearised vectors do not represent the actual size of the vector. Multiple plasmid conformations exist, e.g., nicked, coiled and supercoiled plasmid conformations (Higgins *et al.*, 2015). The linear pETDuet-1 vector is marked with a red rectangle in Figure 20. The band in lane 3 can be observed between the 5000-6000 regions. This confirms the expected size of the pETDuet-1 vector of 5240 bp.

The linearized pHis17_ACSM2B vector band can be seen between 4000-5000 bp, which confirms the size of the expected pHis17_ACSM2B vector (4339 bp) in lane 5. The linearised pHis17_GLYAT vector band can be seen at the 3500 bp region, confirming the expected size of pHis17_GLYAT (3496 bp).

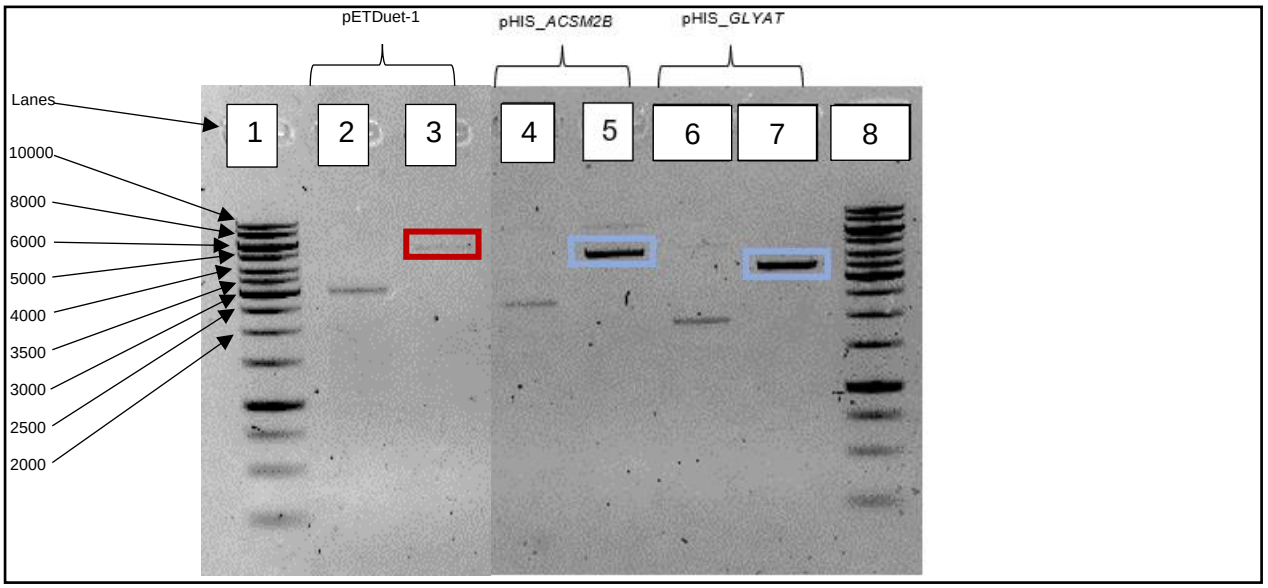


Figure 20: Image of the agarose gel with all plasmid vectors used in the pETDuet-1 system.
Lanes 1 and 8 contain the O'Gene Ruler™ 1 kb molecular weight ladder, lane 2 contains the unlinearised pETDuet-1 vector, lane 3 contains the linearised pETDuet-1 vector, lane 4 contains the unlinearised pHis17_ACSM2B vector, lane 5 contains the linearised pHis17_ACSM2B vector, lane 6 contains the unlinearised pHis17_GLYAT vector, and lane 7 contains the linearised pHis17_GLYAT vector. The red rectangle represents the linearised pETDuet-1 vector, whereas the blue rectangles represent the linearised pHis17_ACSM2B and pHis17_GLYAT vectors.

Because all the vectors were confirmed to have the correct sizes, the next step was to perform sequence determination of both MCSs of the pETDuet-1 vector.

2.3.2 Sequencing of MCS-1 and MCS-2 of the pETDuet-1 vector

Before InFusion® could be performed, sequence determination of both MCSs in the pETDuet-1 vector was performed, as discussed in section 2.2.2.9, to determine if the pETDuet-1 vector contains both the MCS-1 and MCS-2 sites with the correct sequences. Sequence determination was also performed to confirm that there were no mutations present in either

MCS. The primer design was dependent on the correct sequences of the MCSs to have correct binding in PCR and InFusion® reactions.

The sequences received from CAF: Stellenbosch (South Africa) were aligned to the reference sequence of pETDuet-1 (Novagen cat.no. 71146-3).

The alignments can be seen in Figure 21 and Figure 22, which indicate that both the MCS-1 and MCS-2 sites are present within the pETDuet-1 vector and have the correct sequences. It can also be seen that no mutations are present in either MCS.

Only the T7 terminator primer could be used for sequence determination since the pETDuet-1 vector contains two binding sites for the T7 promoter primer (one for each MCS). The T7 terminator sequence determination proved sufficient in determining the sequences for both MCSs.

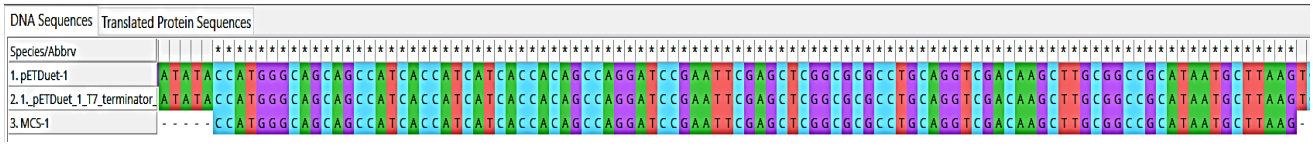


Figure 21: MCS-1 alignment of the T7 terminator sequence with the pETDuet-1 vector reference.

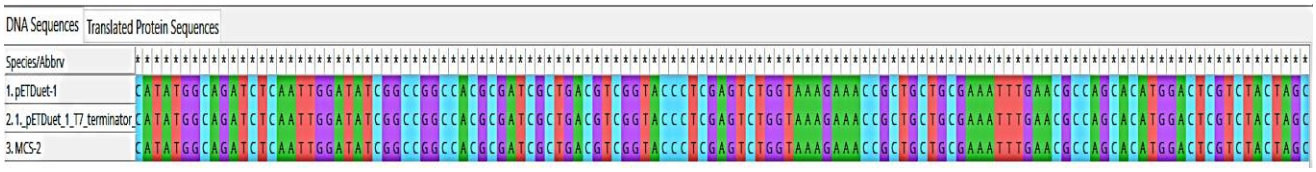


Figure 22: MCS-2 alignment of the T7 terminator sequence with the pETDuet-1 vector reference.

Because the sequence alignments were successful, the next step was to amplify the pETDuet-1 vector used for the first InFusion® reaction.

2.3.3 pETDuet-1 vector PCR amplification

When using InFusion® reactions, the linear fragment is inserted into the linear vector. Thus, the fragment and vector 5' and 3' ends need to be complementary to each other for InFusion® to be done correctly. The pETDuet-1 vector was restriction enzyme digested with BamHI-HF as described in section 2.2.2.8.2 to linearize it. Then, the pETDuet-1 vector was amplified to have the vector 5' and 3' ends match the fragment 15 bp overhangs (as discussed in section 2.2.2.8.1) to perform InFusion® cloning. The linearized pETDuet-1 vector was amplified with Vector.For. *ACSM2B* and Vector.Rev. *ACSM2B* primers as described in section 2.2.2.8.3.

As discussed in section 2.2.2.5, gel extraction of the linearized pETDuet-1 vector was performed to retrieve the linear pETDuet-1 vector for downstream InFusion® reactions. No

temperature optimisation was performed to amplify the pETDuet-1 vector because the primer melting temperatures were the same.

The linearized pETDuet-1 vector from section 2.2.2.8.2 was used as the positive control. Lane 1 contains the linearized positive control for the pETDuet-1 vector. 2-6 contain the amplified linear pETDuet-1 vector, shown in Figure 23. All of the bands can be seen between the 5000-6000 bp region. This confirms the correct size of the pETDuet-1 vector of 5420 bp. Single bands can be observed in every lane, which means that no off-target primer binding was observed and that the amplification of the pETDuet-1 vector was successful.

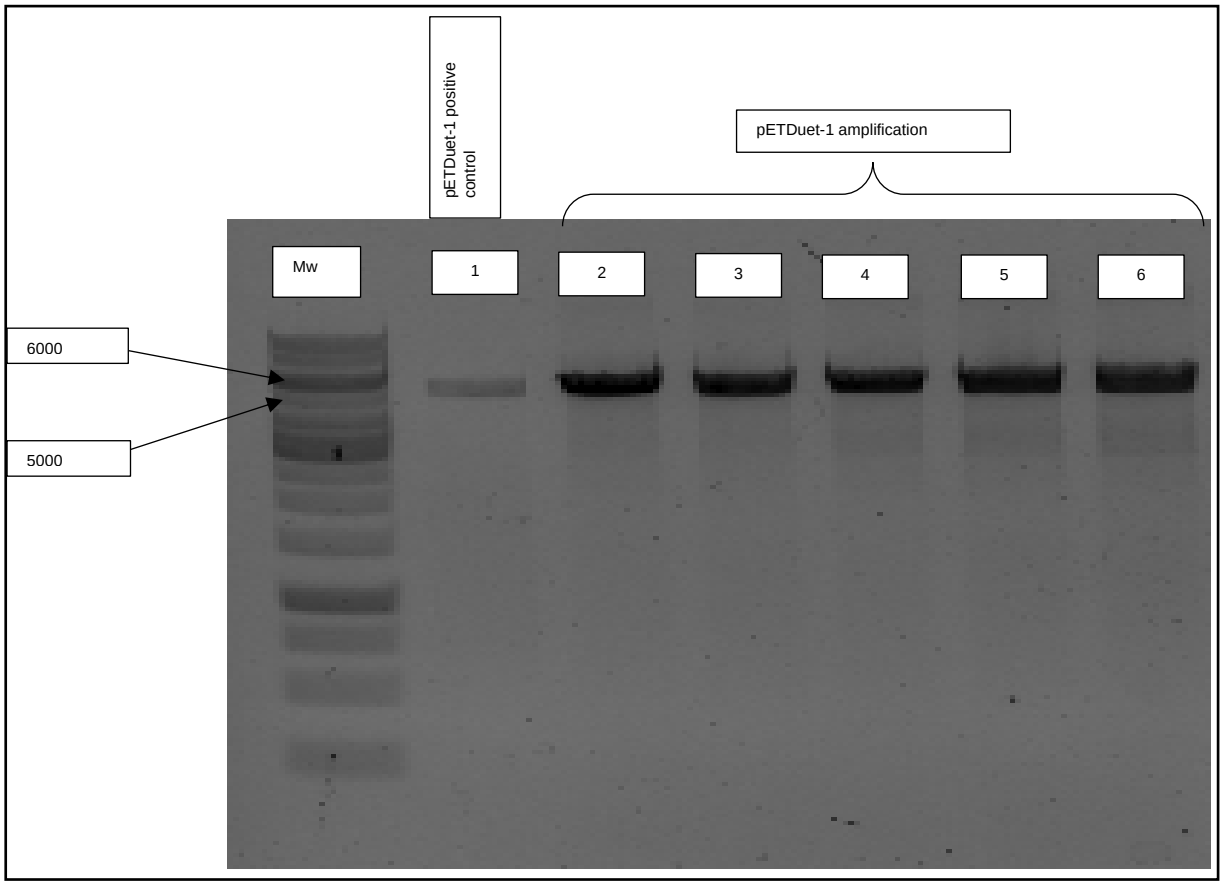


Figure 23: Amplification of the linearised pETDuet-1 vector.
The Mw-lane contains the O'GeneRuler ™ 1 kb molecular weight ladder, lane 1 contains the positive control for the linear pETDuet-1 vector, and lanes 2-6 contain the amplified linear pETDuet-1 vector.

The linear pETDuet-1 vector was successfully amplified. Therefore, the next step was to amplify *ACSM2B* and *GLYAT* using PCR. A temperature gradient was used to determine the optimal amplification conditions for the PCR amplification.

2.3.4 PCR temperature optimisation for the amplification of *ACSM2B* and *GLYAT*

Temperature optimisation for the amplification of *ACSM2B* and *GLYAT* was performed to determine the optimal temperature for downstream PCR amplification before InFusion® was performed. Both genes were amplified across a temperature gradient. The temperature

gradient for the amplification of *ACSM2B* ranged from 52-57°C, whereas the temperature gradient for *GLYAT* ranged from 52-65°C.

The bands in Figure 24 represent the temperature gradients performed for *ACSM2B* and *GLYAT*. Lanes 1-5 contained the amplified *ACSM2B* fragments, whereas lanes 6-10 contained the *GLYAT* fragments. Primer dimers can be seen at the bottom of the gel. These occur due to the elongation of the primers. The 5' ends of the primers hybridise due to complementary base pairs. The DNA polymerase then amplifies the rest of the primer on the 3' ends. These amplification products are the primer dimers reviewed in Simantini *et al.* (1999). However, the primer binding was highly specific due to no contaminating bands present in each lane. The *ACSM2B* bands can be seen between the 1500-2000 regions. This confirms the expected size of 1734 bp of *ACSM2B*. The *GLYAT* bands can be seen between 750-1000 bp, confirming the expected size of 891 bp of *GLYAT*.

For *ACSM2B* amplification, 55.1°C (in lane 4) was selected as the optimal temperature, whereas for *GLYAT* amplification, 52.0°C (in lane 10) was selected as the optimal temperature. Both of these temperatures were selected to be as far away from the extension temperature as possible while being ~5°C below the melting temperature of the primers

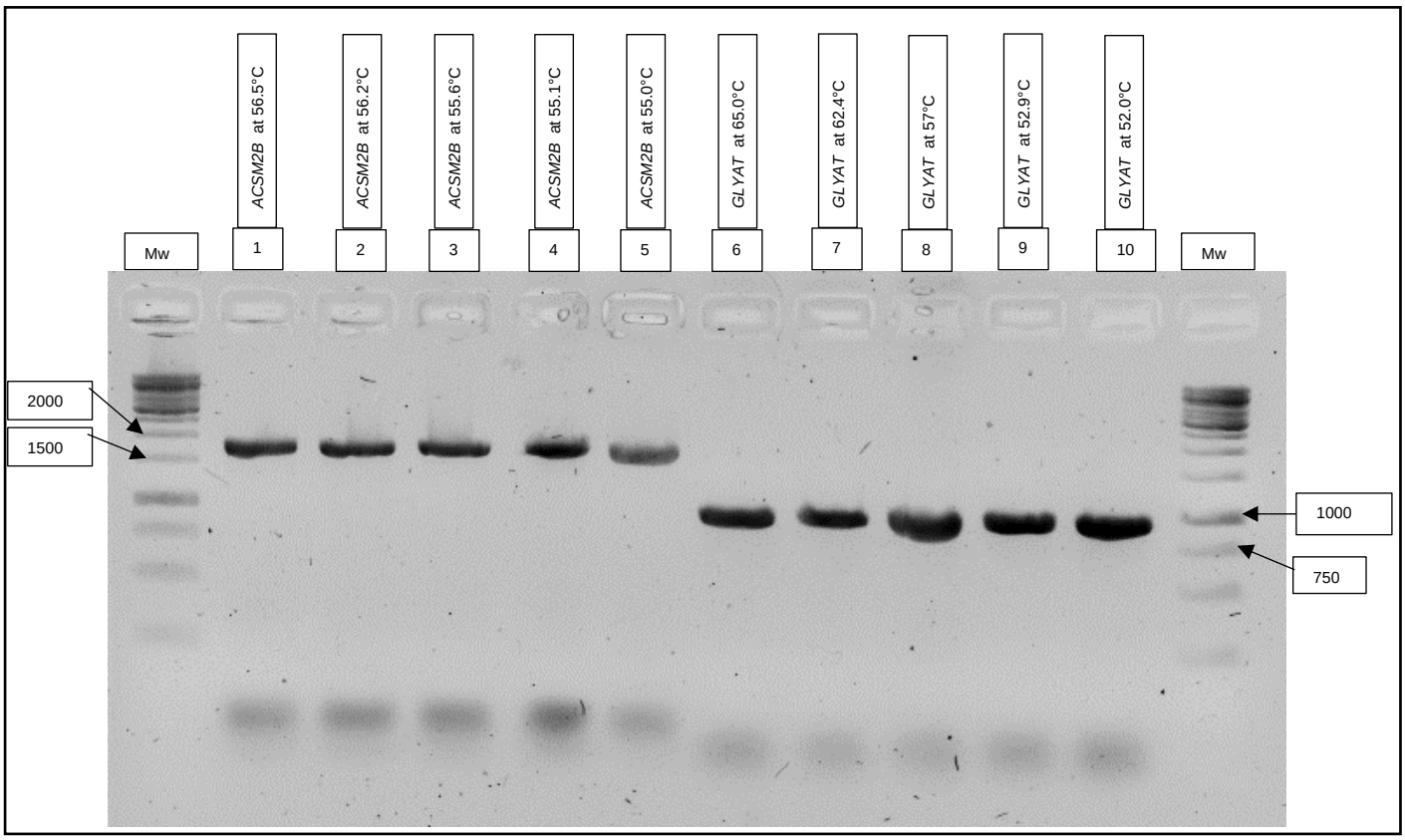


Figure 24: Temperature optimisation for ACSM2B and GLYAT.
Two O'Gene Ruler™ 1 kb molecular weight ladders were loaded in the first and last lanes. Lanes 1-5 contain the PCR products of *ACSM2B*, whereas lanes 6-10 contained the PCR product of *GLYAT*. The amplification temperatures are indicated in the lane names.

(Lorenz, 2012). These temperatures were chosen for the maximal amplification yield of *ACSM2B* and *GLYAT*.

Both *ACSM2B* and *GLYAT* were successfully amplified. The assembly of the pETDuet-1_ACSM2B/GLYAT construct could now start.

2.3.5 Creating the pETDuet-1_ACSM2B/GLYAT vector using InFusion®

The gel-purified linear fragments for *ACSM2B* and the linear pETDuet-1 vector were used to set up the first InFusion® reaction.

2.3.5.1 Confirmation of transformants of the first InFusion® reaction

After the InFusion® reaction of *ACSM2B* into the pETDuet-1 vector, the plasmid DNA was extracted from the selected colonies, as discussed in section 2.2.2.8.5.

PCR amplification of *ACSM2B* was then performed with the optimal PCR conditions from the temperature gradient discussed in section 2.2.2.8.4, using the purified plasmid DNA of each colony as a template. The amplified *ACSM2B* fragments were then run on an agarose gel as discussed in section 2.2.2.3 to determine whether *ACSM2B* was successfully inserted into the pETDuet-1 vector.

Primer dimers can be seen at the bottom of Figure 25 below the 250 bp size marker. Some pETDuet-1 vector DNA can be seen at the top of the bands in lanes 1-5 at the 10 000 bp region. This, however, does not represent the actual size of the pETDuet-1_ACSM2B vector since it is in the supercoiled conformation. The vector can be slightly seen because the DNA concentration used for the PCR was high enough to be visible on the agarose gel. Therefore, less pETDuet-1_ACSM2B vector template could be used for amplification.

It can further be seen from the amplification results that all the selected colonies contained the pETDuet-1_ACSM2B vector because all of the amplified *ACSM2B* bands were located between the 1500-2000 region and were at the correct location compared to the positive control (*ACSM2B* previously amplified in section 2.2.2.8.4). Therefore, all the selected colonies could be used for downstream applications. Glycerol stocks were made from every colony and stored.

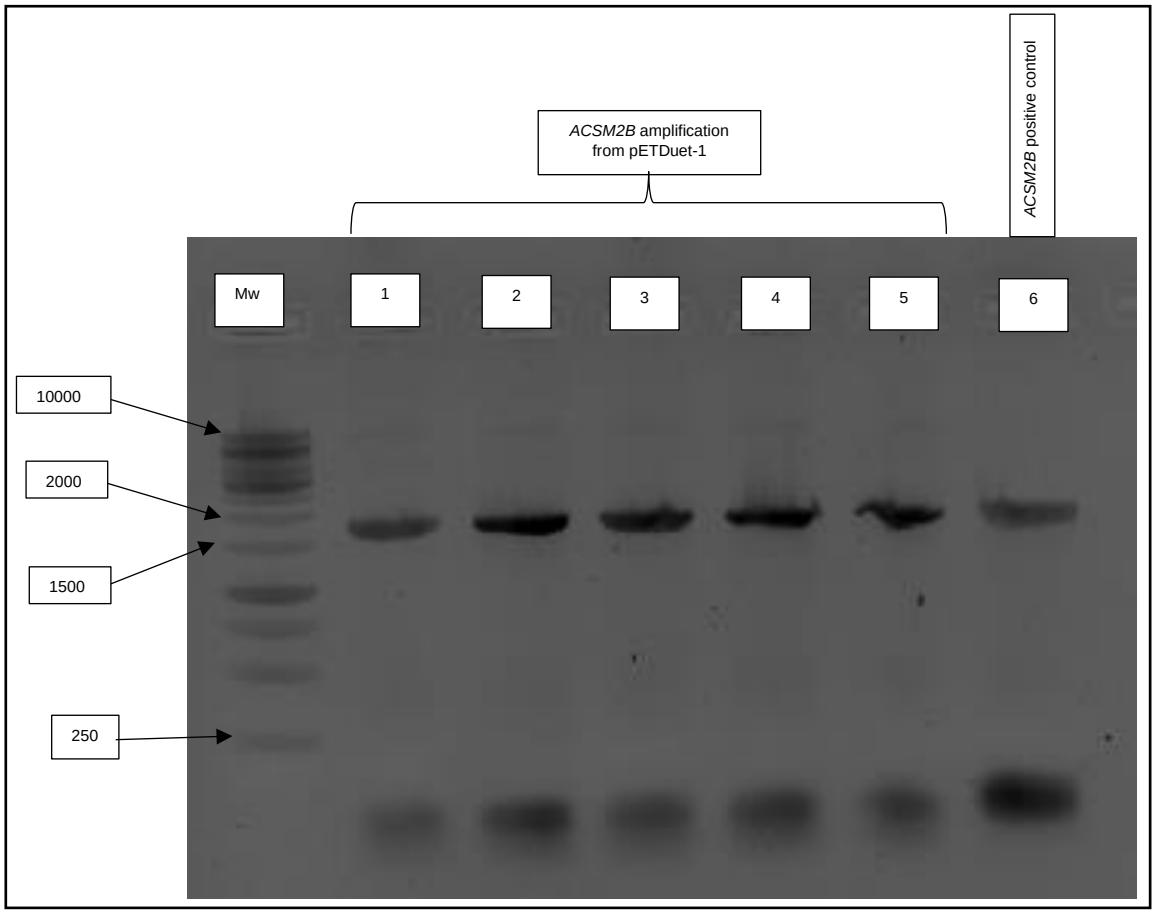


Figure 25: Amplification of *ACSM2B* from the pETDuet-1_ACSM2B vector.
The Mw-lane contains the O'GeneRuler™ 1 kb molecular weight ladder. Lanes 1-5 contain the amplification of *ACSM2B* from the extracted pETDuet-1_ACSM2B DNA from the selected colonies. Lane 6 contains the *ACSM2B* positive control.

To confirm that *ACSM2B* has the correct sequence and is in the correct orientation with no mutations introduced into *ACSM2B*, the sequence of *ACSM2B* was determined as discussed in section 2.2.2.9. Sequence determination of *ACSM2B* was performed with the Vector.For.pETDuet_upstream, Fragment.For. *ACSM2B,* Fragment.Rev. *ACSM2B,* and the Vector.Rev.*hGLYAT* primers.

The sequences received from CAF: Stellenbosch (South Africa) were aligned to the reference sequence for *ACSM2B* (NM_001105069.2). The alignment confirmed that *ACSM2B* was

correctly inserted into the pETDuet-1 vector in the correct orientation with no mutations inserted into *ACSM2B*. The alignment can be seen in Figure 26.

Figure 26: Sequence alignments of *ACSM2B* from the pETDuet-1_ACSM2B vector.

The second InFusion® step could now be performed since the *ACSM2B* fragment was correctly inserted into the pETDuet-1 vector.

2.3.5.2 Confirmation of transformants containing the pETDuet-1_ACSM2B/GLYAT construct

After the InFusion® of *GLYAT* was inserted into the pETDuet-1_ACSM2B vector, as discussed in section 2.2.2.8.7, the pETDuet-1_ACSM2B/GLYAT construct was transformed into C41(DE3)pLysS chemically competent cells. The plasmid DNA was then extracted from

selected colonies, and PCR amplification was performed for *ACSM2B* and *GLYAT,* as discussed in section 2.2.2.8.7. Agarose gel electrophoresis of the amplified *ACSM2B* and *GLYAT* was performed, as discussed in section 2.2.2.3. The gel containing the amplified fragments can be seen in Figure 27. The amplification results served as an additional qualitative confirmation that both genes were present in the pETDuet-1_ACSM2B/GLYAT construct.

Primer dimers can be seen in every lane. The purified pETDuet-1_ACSM2B/GLYAT construct template DNA used for these PCRs can be seen as a faint band in lanes 2-7 and lanes 10-15 at the 10 000 bp region and were not classified as contaminants. These were the supercoiled conformations of the pETDuet-1_ACSM2B/GLYAT construct and did not reflect the actual size of pETDuet-1_ACSM2B/GLYAT.

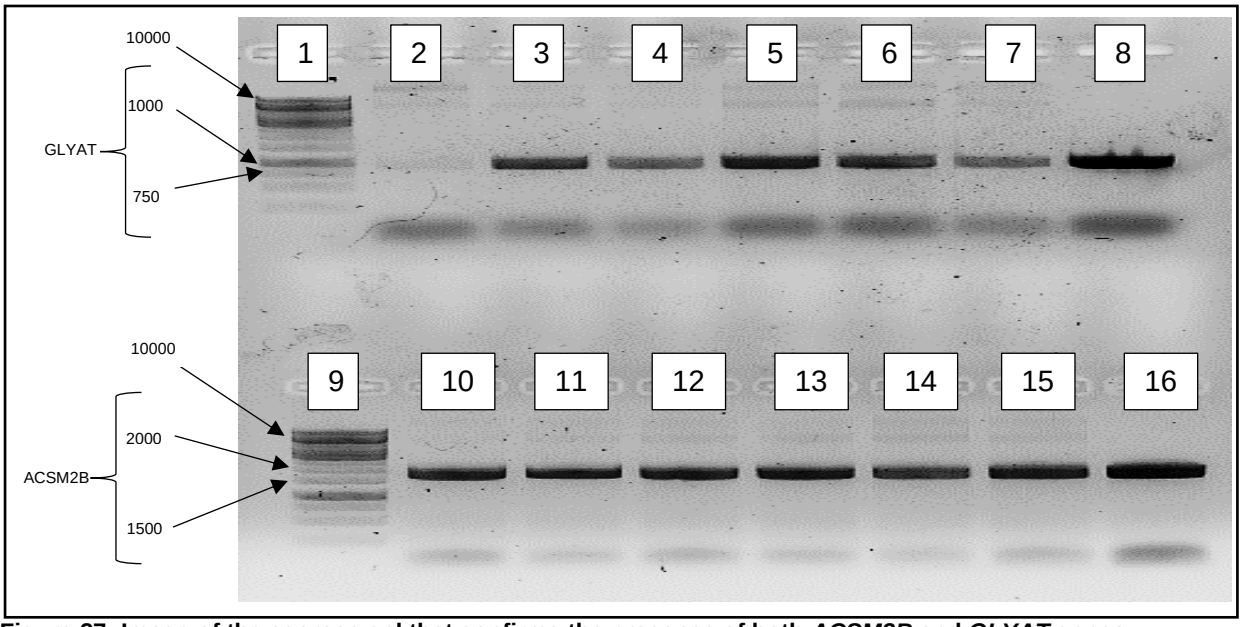


Figure 27: Image of the agarose gel that confirms the presence of both *ACSM2B* and *GLYAT* genes.
Lanes 1 and 9 contain the O'Gene Ruler 1 kb molecular weight ladder. Lanes 2-7 contain the PCR products of *GLYAT*, whereas lanes 10-15 contain the PCR products for *ACSM2B*. Lane 8 contains the positive control for *GLYAT* and lane 16 contains the positive control for *ACSM2B*.

This gel also confirms that the sizes of the inserts are correct because the bands are located between the correct size markers. The amplified *GLYAT* fragment can be seen in the top half of the gel in Figure 27 between the 750-1000 bp markers, whereas the bottom half of this gel contains the amplified *ACSM2B* fragment, which is located between the 1500-2000 bp markers.

Both the top and bottom halves of the gel contained positive controls from the previously amplified *ACSM2B* and *GLYAT* in the temperature gradient experiment of section 2.2.2.8.4. The positive controls can be seen in lane 8 for *GLYAT* and in lane 16 for *ACSM2B*, which also

confirms the presence of both genes in the *ACSM2B_pETDuet-1_GLYAT* construct. Colony 1 was discarded due to the faint band seen in lane 2.

The sequence alignments of *GLYAT* from CAF: Stellenbosch (South Africa) can be seen in Figure 28, whereas the alignments of *ACSM2B* can be seen in Figure 29. The primers used for the sequence determination of both *ACSM2B* and *GLYAT* were Vector.For.pETDuet_upstream, Fragment.For. *ACSM2B*, Fragment.Rev. *ACSM2B*, and Vector.Rev.h*GLYAT* for *ACSM2B* and Vector.For. *ACSM2B*, Fragment.For.h*GLYAT*, Fragment.Rev.h*GLYAT* and the T7 terminator primers for *GLYAT*. The sequences aligned to the reference sequence for *GLYAT* (NM_201648.3) and *ACSM2B* (NM_001105069.2).

Figure 28: Sequence alignments of *GLYAT* from the four sequencing primers used.

Figure 29: Sequence alignments of *ACSM2B* from the four sequencing primers used.

All the sequences from the four primers aligned perfectly with the reference gene for *GLYAT* and *ACSM2B,* and no mutations could be seen in these genes. Therefore, coexpression of ACSM2B and GLYAT in C41(DE3)pLysS cells was next attempted.

2.3.6 Analysis of the soluble expression of ACSM2B and GLYAT

A glycerol stock made from a confirmed colony containing pETDuet-1_ACSM2B/GLYAT in C41(DE3)pLysS cells was used in subsequent expression reactions, as discussed in section 2.2.2.10. An attempt was made to coexpress ACSM2B and GLYAT, and the cells were lysed, as discussed in section 2.2.2.11. The protein lysates were analysed using SDS–PAGE and western blot analysis to confirm the expression of ACSM2B and GLYAT, as discussed in sections 2.2.2.12 and 2.2.2.13.

The expected size of ACSM2B with the N-terminal 6x His-tag was 65.1 kDa, whereas the size of GLYAT with the N-terminal S-tag was 35.6 kDa. The two Coomassie-stained gels in Figure 30 indicated that the lysis technique discussed in sections 2.2.2.11 was adequate for completely lysing the C41(DE3)pLysS cells. These were the lysis buffer and syringe method for Part A and the vibration mill method for Part B.

The expected region of both ACSM2B and GLYAT can be seen in Figure 30, with ACSM2B marked with a blue rectangle and GLYAT marked with a green rectangle. A positive control that was expressed and purified by another MSc study in our laboratory was used. This His-tagged GLYAT positive control was expressed from pHis17_GLYAT. The positive controls can be seen in Figure 30, lanes 7 and 10. The bands were faint but can be seen marked with a red rectangle. The GLYAT positive control was added to observe whether GLYAT is present on the Coomassie-stained gels and confirm the size of the expressed GLYAT.

However, there were too many overlapping bands in the approximate regions of ACSM2B and especially GLYAT on the Coomassie-stained gels, which made ACSM2B and GLYAT indistinguishable from the background proteins. Western blots were, therefore, used to confirm the presence of both proteins within the soluble fraction of the lysate.

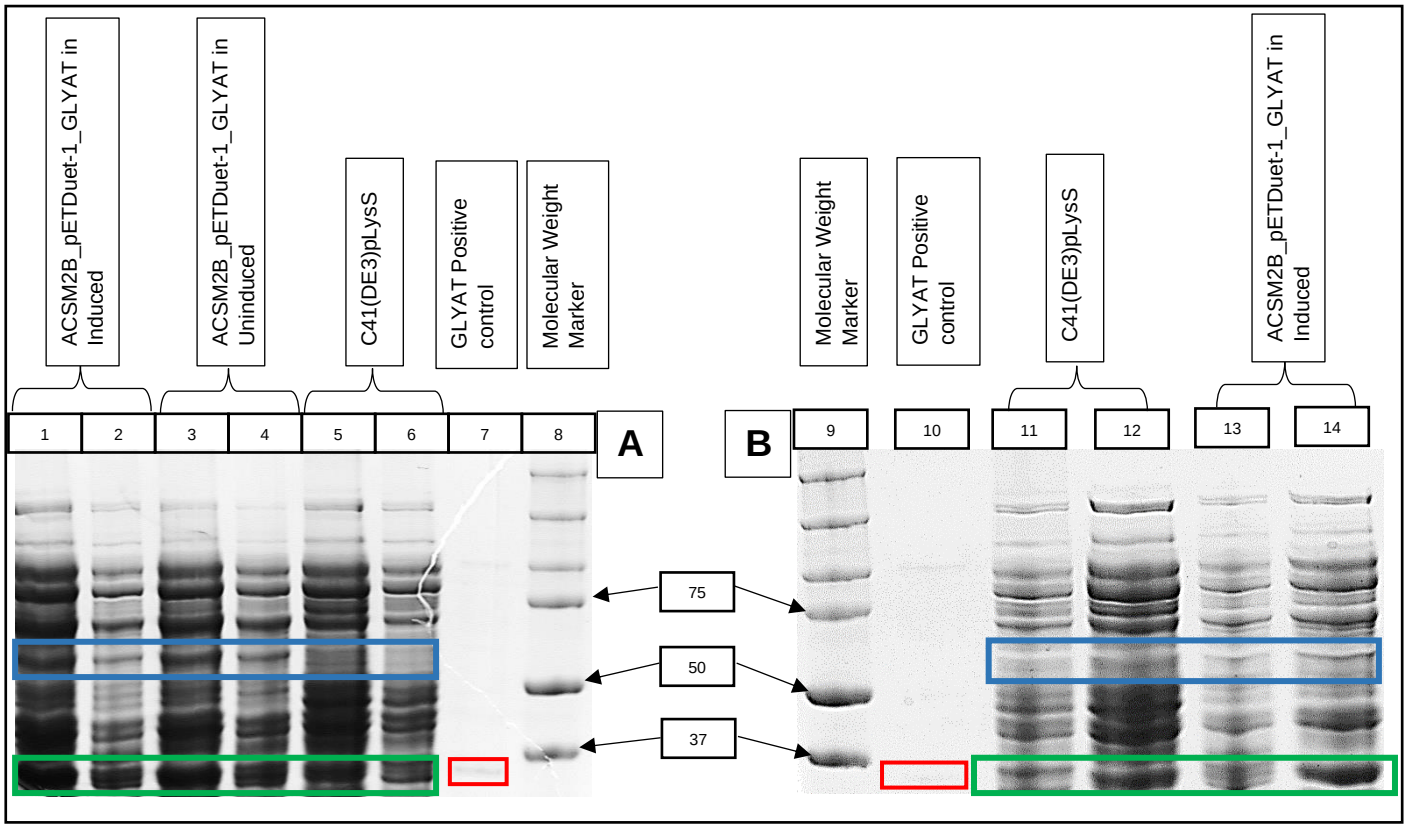


Figure 30: Image of the Coomassie-stained gels to determine the presence of ACSM2B and GLYAT.
The image of gel A represents the expression confirmation gel, whereas gel B represents the lysed samples using the vibration mill. For part A: Lane 8 contains the Precision Plus Protein™ Dual Color Standards. Lane 1 contains the soluble fraction of the induced C41(DE3)pLysS cells, lane 2 contains the total fraction of the induced C41(DE3)pLysS cells, lane 3 contains the soluble fraction of the uninduced C41(DE3)pLysS cells, lane 4 contains the total fraction of the uninduced C41(DE3)pLysS cells, lane 5 contains the soluble fraction of the C41(DE3)pLysS cells without the pETDuet-1_ACSM2B/GLYAT construct, lane 6 contains the total fraction of the C41(DE3)pLysS cells without the pETDuet-1_ACSM2B/GLYAT construct, and lane 7 contains the GLYAT positive control. For part B: Lane 9 contains the Precision Plus Protein™ Dual Color Standards. Lane 10 contains the GLYAT positive control, lane 11 contains the total fraction of the C41(DE3)pLysS cells without the pETDuet-1_ACSM2B/GLYAT construct, lane 12 contains the soluble fraction of the C41(DE3)pLysS cells without the pETDuet-1_ACSM2B/GLYAT construct, lane 13 contains the total fraction of the induced C41(DE3)pLysS cells, and lane 14 contains the soluble fraction of the induced C41(DE3)pLysS cells. The positive controls are marked with a red rectangle, whereas ACSM2B was marked with a blue rectangle, and GLYAT was marked with a green rectangle.

The nitrocellulose membrane was cut into two pieces after transferring the proteins to the nitrocellulose membrane from the SDS–PAGE gels. The two pieces were then incubated separately with antibodies to avoid mixing the HRP-labelled polyclonal goat anti-S-tag antibody (which binds to the S-tag on GLYAT) and the HRP-labelled monoclonal mouse anti-His-tag antibody (which binds to the His-tag on ACSM2B) with each other.

Western blot analysis confirmed that GLYAT was expressed in lanes 1-4 of Part A and lanes 13-14 of Part B marked with green rectangles in Figure 31. The positive control (previously expressed His-tagged GLYAT) can also be seen in lanes 7 and 10, marked with red rectangles. His-tagged ACSM2B was not expressed because no bands can be seen in the expected region marked with the dark blue rectangle. His-tagged GLYAT (positive control) did,

however, develop. Therefore, the problem could not have been the antibody since the development of the positive control was observed.

Leaky expression was observed in lanes 3 and 4 (the uninduced C41(DE3)pLysS cells), even though glucose was added to decrease it. Both western blots confirmed the correct size of GLYAT at approximately 35.6 kDa.

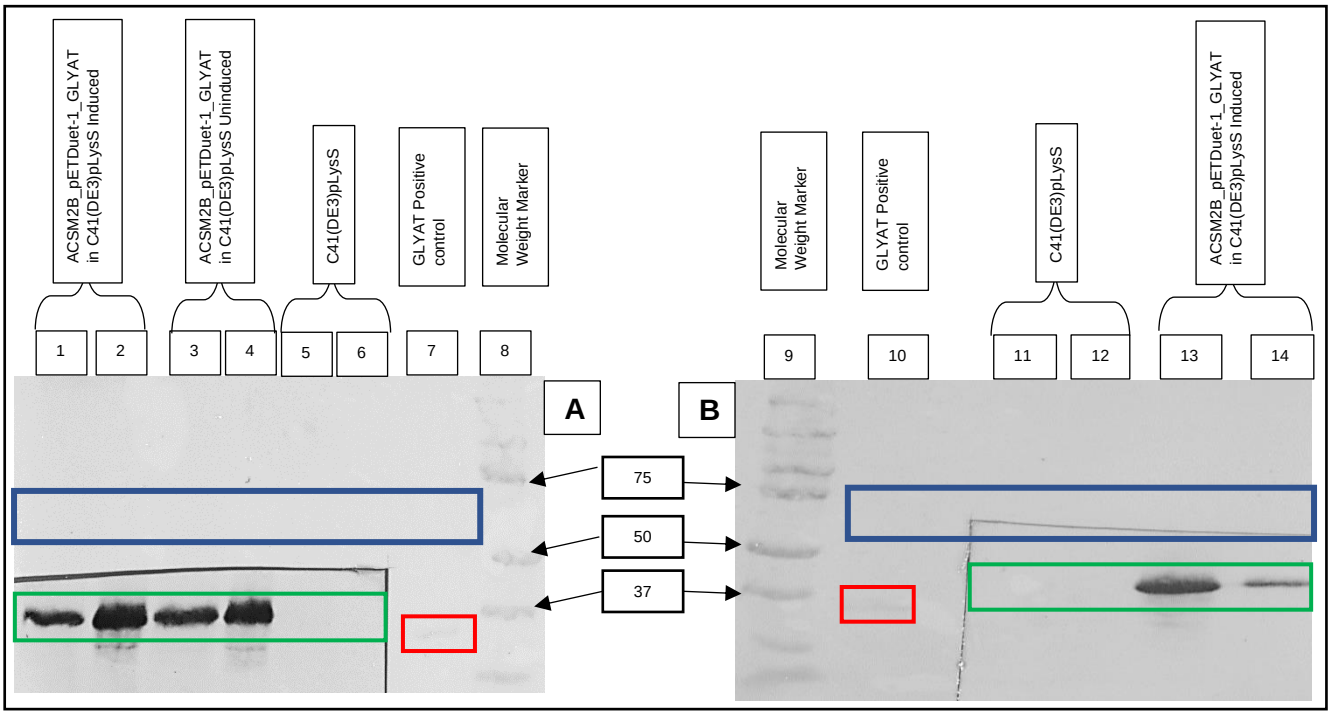


Figure 31: Images of the western blot confirmation of the presence of ACSM2B and GLYAT.
Nitrocellulose membrane A serves a qualitative purpose to observe the presence of GLYAT. In contrast, nitrocellulose membrane B confirms GLYAT in the soluble lysate of the cells lysed with the vibration mill. Nitrocellulose A: Lane 1 contains the soluble fraction of the induced C41(DE3)pLysS cells, lane 2 contains the soluble fraction of the induced C41(DE3)pLysS, lane 3 contains the total fraction of the uninduced C41(DE3)pLysS cells, lane 5-6 contains the total and soluble lysates of the C41(DE3)pLysS cells without the pETDuet-1_ACSM2B/GLYAT construct, and lane 7 contains the GLYAT positive control. Lanes 8-9 conta the Precision Plus Protein™ Dual Color standards. Nitrocellulose membrane B: Lane 10 contains the GLYAT positive control. Lanes 11-12 contain the total and soluble lysates of the C41(DE3)pLysS cells without the pETDuet-1_ACSM2B/GLYAT plasmid, lane 13 contains the total fraction of the induced C41(DE3)pLysS cells, and lane 14 contains the soluble fraction of the induced C41(DE3)pLysS cells.

Because no ACSM2B was expressed, the first suspected problem was the expression media (in this case, LB broth). The expression of two recombinant proteins may have caused a high metabolic strain on the C41(DE3)pLysS cells. This increased metabolic strain may have then influenced the expression of the more toxic enzyme, ACSM2B. The toxicity of ACSM2B to the cells may be explained by the formation of CoA thioesters (Brass, 1994). Testing the expression again with a new expression medium was the quickest and most cost-effective method. However, several other changes can be made if time constraints and cost-effectiveness are not a problem. These additional techniques are thoroughly explained in Saida *et al.* (2006). Another possibility could be adding a highly soluble fusion tag such as the

Trx tag (thioredoxin tag) to ACSM2B to aid soluble expression (Malhotra, 2009; Sachdev & Chirgwin, 1998). The expression of soluble ACSM2B was not further investigated in this study since it was part of another MSc study in our department.

The media was changed to the expression media used in (van der Sluis *et al.*, 2013) because it contained more complex nutrients, resulting in the soluble expression of GLYAT in that study. The new expression media contained 2% bacto-tryptone, 1.25% yeast extract, 0.625% NaCl, 0.5% $Na_2HPO_4.12H_2O$, 0.1% KH_2PO_4, and 0.2% glucose. The expression steps were repeated as discussed in section 2.2.2.10.

Figure 32 shows that the C41(DE3)pLysS cells did not express ACSM2B when the media was changed. Therefore, this aspect of the study needs more investigation.

In Figure 32, the positive control GLYAT (previously expressed His-tagged GLYAT) is marked with red rectangles, and the predicted region for ACSM2B is marked with green rectangles for both parts A and B. No ACSM2B was expressed in the soluble or total lysate fractions.

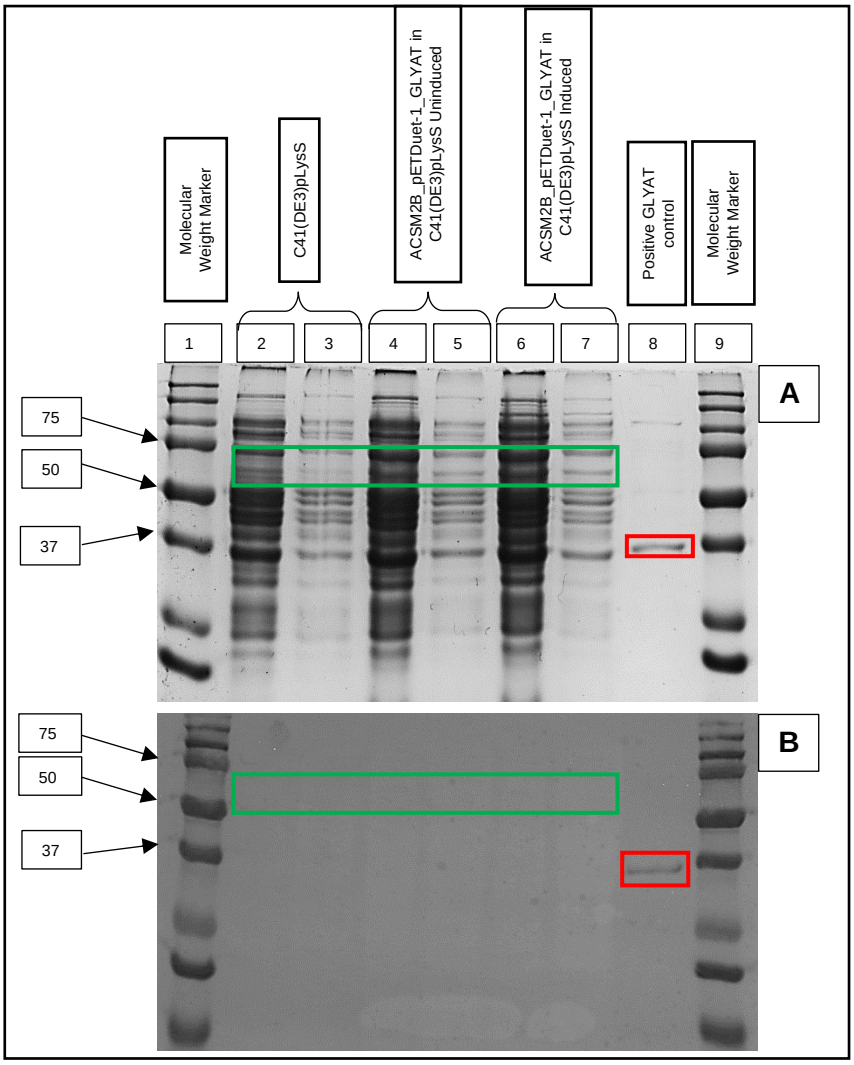


Figure 32: SDS-PAGE and western blot analysis of no *ACSM2B* expression.
Part A represents the Coomassie-stained gel of the additional attempt to express ACSM2B, whereas part B contains the nitrocellulose membrane for the additional attempt to express ACSM2B. Lane 2 contains the Precision Plus Protein™ Dual Color standards (Bio-Rad). Lane 3 contains the total fraction of the C41(DE3)pLysS cells without the pETDuet-1_ACSM2B/GLYAT construct, lane 4 contains the soluble fraction of the C41(DE3)pLysS cells without the pETDuet-1_ACSM2B/GLYAT construct, lane 5 contains the total fraction of the uninduced C41(DE3)pLysS cells, lane 5 contains the soluble fraction of the uninduced C41(DE3)pLysS cells, lane 6 contains the total fraction of the induced C41(DE3)pLysS cells, and lane 7 contains the soluble fraction of the induced C41(DE3)pLysS cells. Lane 8 contains the positive GLYAT control, and lanes 1 and 9 contain the Precision Plus Protein™ Dual Color Standards (Bio-Rad).

2.4 Summary

The main goal of this chapter of the study was to generate the pETDuet-1_ACSM2B/GLYAT construct and to attempt to coexpress soluble ACSM2B and GLYAT in a single host.

The construct was created using two InFusion® reactions. The first was used to insert *ACSM2B* into the pETDuet-1 vector, and the second was used to insert *GLYAT* into the pETDuet-1_ACSM2B vector. This part was successfully performed and verified with PCR amplification and sequence determination of the inserted *ACSM2B* and *GLYAT*. The construct was transformed into the C41(DE3)pLysS host. The bands were present at the correct regions after PCR verification, and the sequence alignments indicated that the genes were successfully inserted in the correct orientation with no mutations observed in either *ACSM2B* or *GLYAT*. After sequence determination was performed, the next goal of this chapter was to attempt the coexpression of ACSM2B and GLYAT in C41(DE3)pLysS cells. Two attempts were made to coexpress ACSM2B and GLYAT in C41(DE3)pLysS cells.

The two coexpression attempts did not successfully express ACSM2B. This suggested that the media may not have been the problem. Some other measures could still be tested in the future, as described in Saida *et al.* (2006).

GLYAT was overexpressed very well, and therefore, the first objective to coexpress ACSM2B and GLYAT was only partially achieved. The long-term goal remains to coexpress ACSM2B and GLYAT in an established model to characterise the different components of the glycine conjugation pathway to establish an enzymatic model. Furthermore, this established coexpression model can then be used to study the effect of genetic variants of both enzymes on the glycine conjugation pathway. However, attempts to express different variants of *GLYAT* can be attempted using this system since GLYAT was overexpressed.

Isovaleric acidemia causes adverse health effects for patients with the isovaleryl-CoA dehydrogenase defect. Because this organic acid is endogenously formed in the human body, the question remains whether GLYAT or one of its paralogues, such as GLYATL1, can aid in eliminating isovaleryl-CoA by conjugation to either glycine or glutamine. The partial success of the pETDuet-1 vector may, therefore, still contribute to answering this question by characterising and observing the decrease in isovaleryl-CoA and increase in isovalerylglycine/isovalerylglutamine related to one of the amino acid conjugation reactions. Moreover, these conjugation reactions with GLYAT can identify possible substrate competition occurring in the glycine conjugation pathway.

Furthermore, the pETDuet-1 system can still be used to determine the influence of genetic variants on glycine conjugation in Phase II since the pETDuet-1 system successfully expressed soluble GLYAT.

Chapter 3: Establishment of an expression system to express soluble GLYAT and GLYATL1

3.1 Introduction

After two attempts, ACSM2B could not be coexpressed with GLYAT using the pETDuet-1 system in the C41(DE3)pLysS cell line. Instead, only the overexpression of GLYAT was observed. Therefore, it is suggested that this host is well suited for expressing *GLYAT* and perhaps some of its paralogues. Other expression systems that express ACSM2B as a soluble protein were not investigated further since it was part of another MSc (Master of Science) study conducted in our laboratory.

Isovaleric acidemia (IVA) is caused by a defective isovaleryl-CoA dehydrogenase (IVD) enzyme in the leucine catabolic pathway. This leads to the accumulation of isovaleric acid and isovaleryl-CoA (Krieger & Tanaka, 1976; Tanaka *et al.*, 1966). Previous studies identified isovalerylglycine and isovalerylglutamine in the urine of patients with IVA (Ensenauer *et al.*, 2004; Wysocki *et al.*, 1983), which may suggest the involvement of the glycine conjugation pathway in alleviating the bioaccumulation of isovaleric acid and isovaleryl-CoA. GLYAT forms part of the second phase of the glycine conjugation pathway discussed in section 1.3, which detoxifies endogenous and exogenous xenobiotics. There are, however, few studies that support the involvement of GLYAT in forming isovaleryl-CoA conjugates.

Two paralogues of *GLYAT* exist, namely, *GLYATL1* and *GLYAL2*. These express GLYATL1 and GLYATL2, respectively. GLYATL1 is expressed in the kidneys (Zhang *et al.*, 2007). *GLYATL1* is discussed in detail in section 1.3.3. Furthermore, GLYATL1 was also shown to prefer glutamine for conjugation reactions (Matsuo *et al.*, 2012), which might suggest that it is responsible for conjugating isovaleryl-CoA to glutamine in the kidneys. It is therefore essential to verify the possible involvement of GLYAT and GLYATL1 in isovaleryl-CoA conjugation.

To verify the involvement of GLYAT and/or GLYATL1 in isovaleryl-CoA conjugation reactions, His-tagged GLYAT was expressed using the pHis17_GLYAT vector available in our laboratory, while His-tagged GLYATL1 was expressed from the pET23a(+)_GLYATL1 vector ordered from GenScript.

A simplified workflow for this chapter can be seen in Figure 33. The figure only depicts the main steps of the experimental procedure, followed by a more detailed description of the methods used in this chapter.

Simplified workflow of Chapter 3

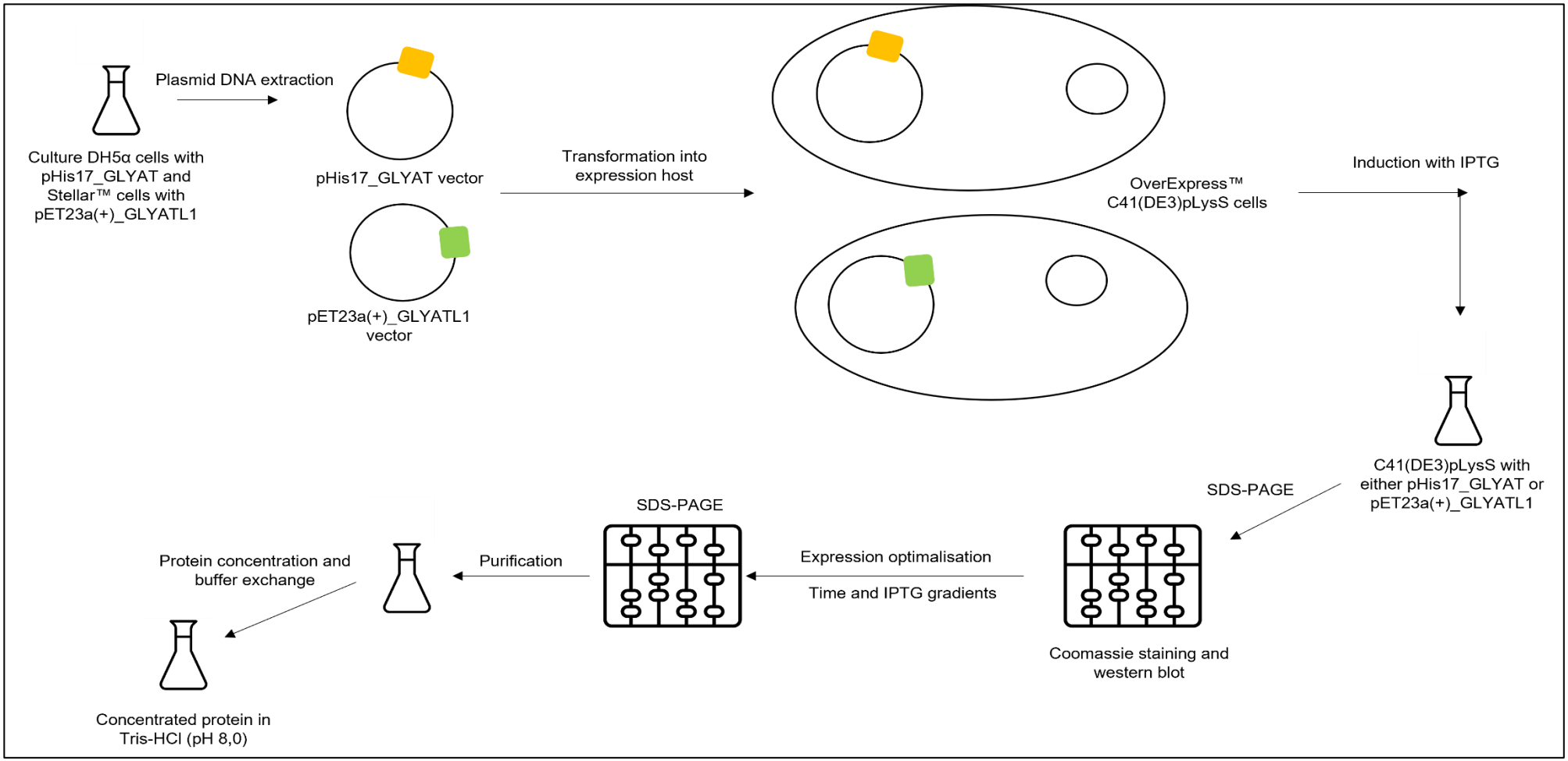


Figure 33: Simplified workflow of methods in Chapter 3.
The orange block on the plasmid represents GLYAT, whereas the green block on the other plasmid represents GLYATL1. This image is only a representation of the main workflow concepts used in Chapter 3. The smaller plasmid represents the pLysS plasmid.

3.2 Materials and methods

3.2.1 Materials

A list containing the reagents used and manufacturers of the reagents is included in Appendix A.

The reference sequences were chosen according to the haplotype with the highest haplotype frequency (for *GLYAT* and *GLYATL1*) in a worldwide population using the data from the GenomAD Browser (v.2.1.1.). For *GLYAT*, the highest haplotype reference sequence was NM_201648.3, and for *GLYATL1,* it was NM_001220496.4. The pHis17_GLYAT vector was already available in our laboratory. pET23a(+) with the subcloned GLYATL1 was ordered from GenScript Biotech (Netherlands) B.V. The pHis17_GLYAT vector can be seen in Figure 17, whereas the purchased pET23a(+)_GLYATL1 vector can be seen in Figure 34.

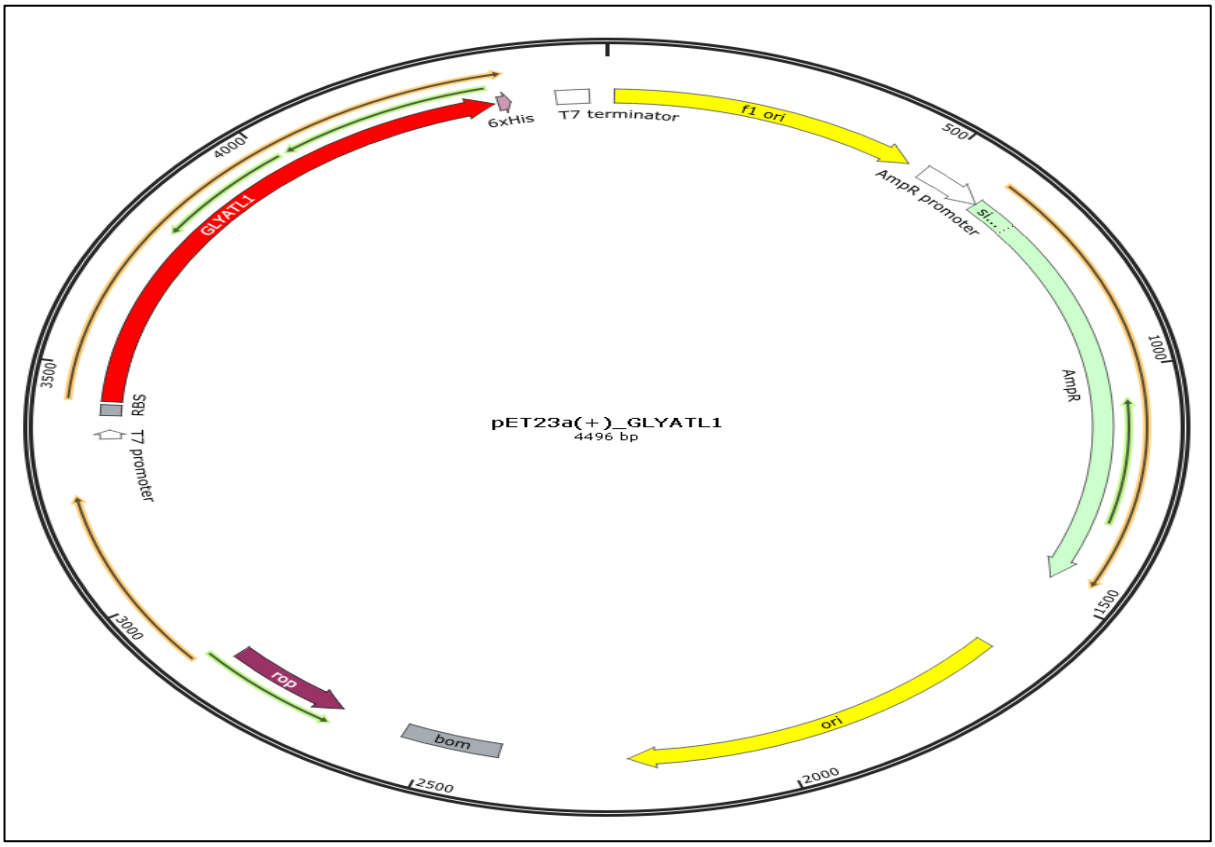


Figure 34: The pET23a(+)_GLYATL1 vector with *GLYATL1*.
The relevant features of this vector were the ori site, T7 promoter and terminator, the AmpR gene, the *GLYATL1*, and the 6xHis tag. This figure was compiled in Snapgene v.3.2.1.

Glycerol stocks of the C41(DE3)pLysS cells with the pHis17_GLYAT vector and the C41(DE3)pLysS cells and JM109 cells with the pET23a(+)_GLYATL1 vector were prepared as described in section 2.2.1.

The JM109 chemically competent cells for cloning and the OverExpress™ C41(DE3)pLysS chemically competent cells (Sigma–Aldrich) were available in our laboratory. The JM109 chemically competent cells were used for cloning, and the OverExpress™ C41(DE3)pLysS chemically competent cells (Sigma–Aldrich) were used for protein expression. All competent cells were stored at -80°C.

Before every induction step, a fresh 1 M IPTG stock solution was prepared in a 1.5 mL microcentrifuge tube with ddH2O.

1 x LEW and 1 x Elution buffer working solution were made according to the Protino® Ni-TED 2000 Packed Columns (Macherey-Nagel) with ddH_2O before purification.

3.2.2 Methods

3.2.2.1 Transformation into JM109 and C41(DE3)pLysS

Transformation was performed as discussed in section 2.2.2.7. pHis17_GLYAT was transformed into OverExpress™ C41(DE3)pLysS competent cells, whereas pET23a(+)_GLYATL1 was transformed into JM109 competent and OverExpress™ C41(DE3)pLysS competent cells. After that, glycerol stocks of the selected transformants were made for long-term storage, as discussed in section 2.2.1.

3.2.2.2 Sequence determination of pHis17_GLYAT and pET23a(+)_GLYATL1

The pHis17_GLYAT vector was extracted from C41(DE3)pLysS, and pET23a(+)_GLYATL1 was extracted from JM109 and C41(DE3)pLysS cells using the mini preparation extraction method discussed in section 2.2.2.1. The pHis17_GLYAT and pET23a(+)_GLYATL1 vector DNA were sent for sequence determination as discussed in section 2.2.2.9 to confirm that the sequences for *GLYAT* and *GLYATL1* were correct, that the fragment was in the correct orientation and that no mutations were introduced. The sequence determination primers are included in Table 24.

Table 24: The primers used for the sequencing determination of *GLYAT* and *GLYATL1* from pHis17_GLYAT and pET23a(+)_GLYATL1, respectively.

Primer name	Primer sequence (5'-3')
T7 Promoter	TAATACGACTCACTATAGGG
T7 Terminator	GCTAGTTATTGCTCAGCGG

3.2.2.3 Determining the optimal IPTG concentration to express soluble GLYAT and GLYATL1

An IPTG gradient was evaluated to determine at which concentration of IPTG the soluble expression yield was maximal.

The C41(DE3)pLysS cells containing pHis17_GLYAT and those containing pET23a(+)_GLYATL1 were taken from the glycerol stocks and inoculated into 25 mL of LB broth that contained 100 µg/mL carbenicillin and 34 µg/mL chloramphenicol. C41(DE3)pLysS cells that did not contain the recombinant vector were also inoculated into a separate flask of 25 mL of LB broth that contained only 34 µg/mL chloramphenicol. The inoculants were grown overnight at 37°C with shaking at 200 rpm. The recombinant vectors used in this section were the pHis17_GLYAT and pET23a(+)_GLYATL1 vectors.

The next day, 9 mL of the cells containing the recombinant vectors were inoculated into 400 mL of LB broth with 100 ng/µL carbenicillin and 34 ng/µL chloramphenicol. Three milliliters of C41(DE3)pLysS cells that did not contain the recombinant vector were inoculated into 100 mL of LB broth with 34 ng/µL chloramphenicol. Glycine (5 mM) was added to the cells that contained the pHis17_GLYAT vector, whereas 5 mM glutamine was added to the cells containing the pET23a(+)_GLYATL1 vector. Glucose (5 mM) was added to all flasks. The cells were grown at 37°C and shaken at 200 rpm until they reached an OD_{600} value of ~ 0.5.

The 400 mL cells were then aliquoted into 4 x 100 mL cultures in new flasks. IPTG was added to final concentrations of 30 mM, 60 mM, and 90 mM in separate 100 mL cultures. No IPTG was added to the cells that did not contain the recombinant vector and the uninduced cells (cells that contained the recombinant vector but were not induced). The cells were induced for 24 hours at 25°C.

The next day, the 100 mL cell cultures were centrifuged at 4000 x g for 15 minutes at 4°C. After that, the cells were lysed, and SDS–PAGE analysis (both Coomassie staining and western blotting) was performed as discussed in sections 2.2.2.12 and 2.2.2.13.

Important note

Whenever GLYATL1 was expressed, the media described in van der Sluis *et al.* (2013) was used. This expression media contained 2% bacto-tryptone, 1.25% yeast extract, 0.625% NaCl, 0.5% $Na_2HPO_4.12H_2O$, 0.1% KH_2PO_4, and 0.2% glucose.

3.2.2.4 Determining the optimal induction time to express soluble GLYAT and GLYATL1

After the IPTG gradient was evaluated and the optimal IPTG concentration was determined, the time gradient was performed exactly as discussed in section 3.2.2.3 to determine the optimal induction time for maximal soluble protein recovery. The induction times chosen were 8, 16 and 24 hours.

3.2.2.5 Purification of GLYAT and GLYATL1 from the C41(DE3)pLysS cell lysate

After determining the optimal IPTG concentration and expression time, these parameters were used to express the soluble recombinant enzymes. The cells were lysed, and the cleared lysate was immediately used for purification.

In this study, purification was performed according to the manufacturer's protocol with Protino® Ni-TED 2000 Packed columns. Protino® Ni-TED is a dry silica-based resin precharged with Ni^{2+} ions. The His residues interact with the Ni^{2+} ions. The gravity flow procedure was used to purify the recombinant proteins.

The Protino® Ni-TED Packed Columns were equilibrated by adding 4 mL of the prepared 1 x LEW buffer into the column. The column was then allowed to drain by gravity. Next, the cleared lysate was added to the column and allowed to drain by gravity. This step was repeated thrice. The column was then washed with 4 mL 1 x LEW buffer and allowed to drain by gravity. This step was repeated twice.

A new collection tube was placed underneath the column, and 3 mL of the 1 x prepared elution buffer was added to the column and allowed to drain into the new collection tube. This step was repeated twice. Thus, there were three elution fractions. All of the flow-through fractions were kept and stored at 4°C. SDS–PAGE analysis (Coomassie staining and western blotting) was performed as discussed in sections 2.2.2.12 and 2.2.2.13 to confirm the success of the purification process.

3.2.2.6 Concentrations of GLYAT and GLYATL1 and buffer exchange

After purification, concentration and buffer exchange were performed using the Pierce™ Protein Concentrator PES, 30K MWCO, 5-20 mL manual to have concentrated GLYAT and GLYATL1 available to perform downstream reactions on the GC–MS for comparison purposes. The buffer in which GLYAT and GLYATL1 were eluted was changed because of

the elution buffer's high salt content, which interferes with the reaction run during GC–MS. The chosen exchange buffer was Tris-HCl (pH 8.0).

All of elution fraction 1 was loaded into the concentrator sample chamber. An appropriate balance was also made. The concentrator sample chambers were then centrifuged at 6000 x g for 10 minutes (or until the sample volume was reduced by 90-95%). The flow-through was discarded.

The same volume (compared to elution fraction volume 1) of the new dilution buffer (Tris-HCl, pH 8.0) was added to dilute the concentrate in the concentrator sample chamber. The concentrator tubes were then centrifuged at 6000 x g for 10 minutes. This step was repeated 4 times.

The concentrated sample in Tris-HCl (pH 8.0) was stored at 4°C. SDS–PAGE (both Coomassie staining and western blotting) was performed as discussed in sections 2.2.2.12 and 2.2.2.13 to confirm that the protein concentration was successful and that the proteins were present in the exchange buffer.

3.2.2.7 Protein quantification using the Qubit 2.0 Fluorometer

The Qubit™ Protein and Protein Broad Range (BR) Assay kit (Thermo Fisher) was purchased. Qubit 2.0 already had the latest software installed before use. The proteins were quantified in the concentrated sample to determine how much protein to add to each enzyme reaction prepared with the purified, concentrated GLYAT and GLYATL1.

The Qubit™ Protein and Protein Broad Range (BR) Assay kit (Thermo Fisher) can measure initial protein sample concentrations between 12.5 µg/mL and 5 mg/mL and exhibits low protein-to-protein variation. This method measures the fluorescence (dependent on the protein concentration in the sample) and compares it to prediluted Qubit™ BSA (Bovine Serum Albumin) standards.

A Qubit™ working solution was prepared at a final ratio of 1:200 with Qubit™ Protein reagent in Qubit™ Protein Buffer. This was done by adding 5 µL of Qubit™ Protein reagent to 995 µL Qubit™ Protein Buffer. This was enough for the 3 prediluted BSA standards and 2 protein samples. Then, 5 x 0.5 mL thin-wall, clear PCR tubes were labelled, and 190 µL Qubit™ working solution was added to each tube.

Ten microliters of the three prediluted Qubit™ BSA standards were added to 3 different tubes and briefly vortexed, with care not to create bubbles. Ten microliters of the purified protein sample was then added to 190 µL of Qubit™ working solution and vortexed briefly. The tubes were incubated at room temperature for 15 minutes.

The **protein** setting was selected on the device, and the command prompts were followed. The three Qubit™ BSA protein standards were measured first, then the two samples.

3.3 Results and discussion

3.3.1 Sequence determination after pHis17_GLYAT and pET23a(+)_GLYATL1 were transformed into C41(DE3)pLysS cells

The received sequences were processed as discussed in 2.2.2.9. *GLYATL1* was aligned to the *GLYATL1* reference sequence NM_001220496.4 and is shown in Figure 35. *GLYAT* alignment to the *GLYAT* reference sequence NM_201648.3 can be seen in Figure 36. This was done to determine if the sequences were correct and in the correct orientation with no mutations introduced.

The sequence alignments for *GLYAT* and *GLYATL1* to the respective reference sequences confirmed *GLYAT* in pHis17_GLYAT and *GLYATL1* in pET23a(+)_GLYATL1. Furthermore, both *GLYAT* and *GLYATL1* were in the correct orientation, and no mutations were present in the genes after transformation into C41(DE3)pLysS cells.

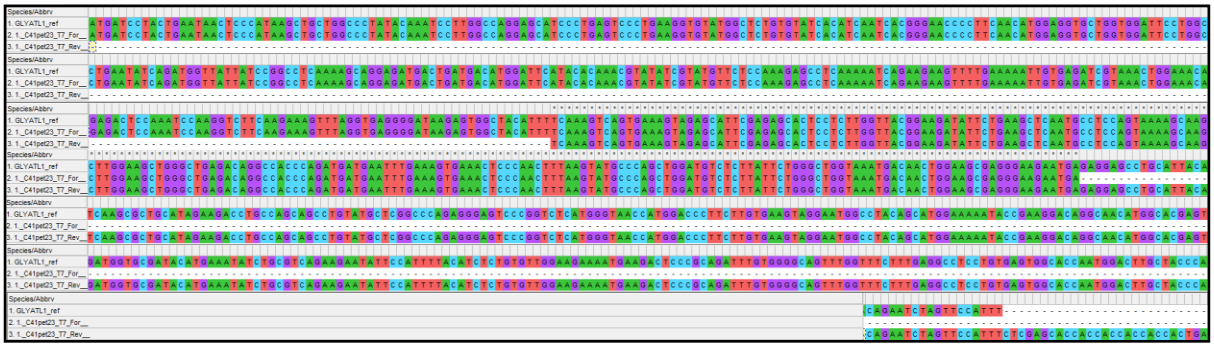


Figure 35: *GLYATL1* alignment to reference sequence.

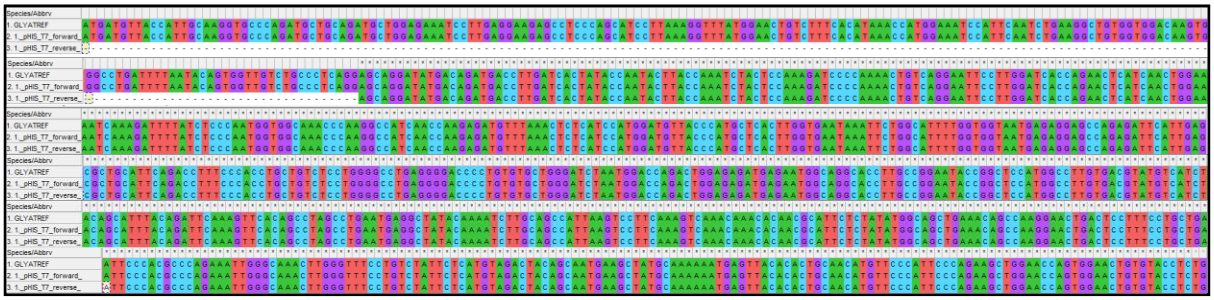


Figure 36: *GLYAT* alignment to the reference sequence.

After both alignments were performed and were deemed correct, the next step was to attempt to express soluble GLYAT and GLYATL1 in the C41(DE3)pLysS host.

3.3.2 Determining the optimal IPTG concentration to express GLYAT and GLYATL1

GLYAT and GLYATL1 were expressed in C41(DE3)pLysS cells using an IPTG concentration gradient to observe at which IPTG concentration maximal soluble GLYAT was yielded. SDS–PAGE analysis was performed and can be seen in Figure 37 and Figure 38.

The Coomassie-stained gels for GLYAT and GLYATL1 indicated good expression of the cellular proteins and adequate lysis. However, background proteins make it difficult to distinguish GLYAT and GLYATL1. Therefore, western blotting was performed to determine whether soluble GLYAT and GLYATL1 were present.

The nitrocellulose membranes were incubated with HRP-labelled monoclonal mouse anti-His-tag antibodies. The nitrocellulose membranes were developed, and it can be seen that His-tagged GLYAT and His-tagged GLYATL1 were present on the western blot in Figure 37 and Figure 38. The bands in Figure 37 marked with a yellow rectangle represent GLYAT. The bands in Figure 38, marked with a brown rectangle, represent GLYATL1. The cells expressing GLYAT had the lowest yield for the uninduced cells and the highest yield when induced at 0.6 mM IPTG overnight. The uninduced cells did express protein even though glucose was added to prevent it. The presence of glucose reduces the intracellular cAMP due to it being easier metabolizable carbon source other than lactose. Therefore, in the presence of glucose the target gene is expressed at a lower concentration (Novy & Morris, s.a.). For GLYATL1 expression, the highest yield was observed at an IPTG concentration of 0.6 mM.

The western blot in Figure 37 indicated that His-tagged GLYAT was expressed at the correct size of 34.74 kDa and that GLYATL1 was correctly expressed at 36.2 kDa.

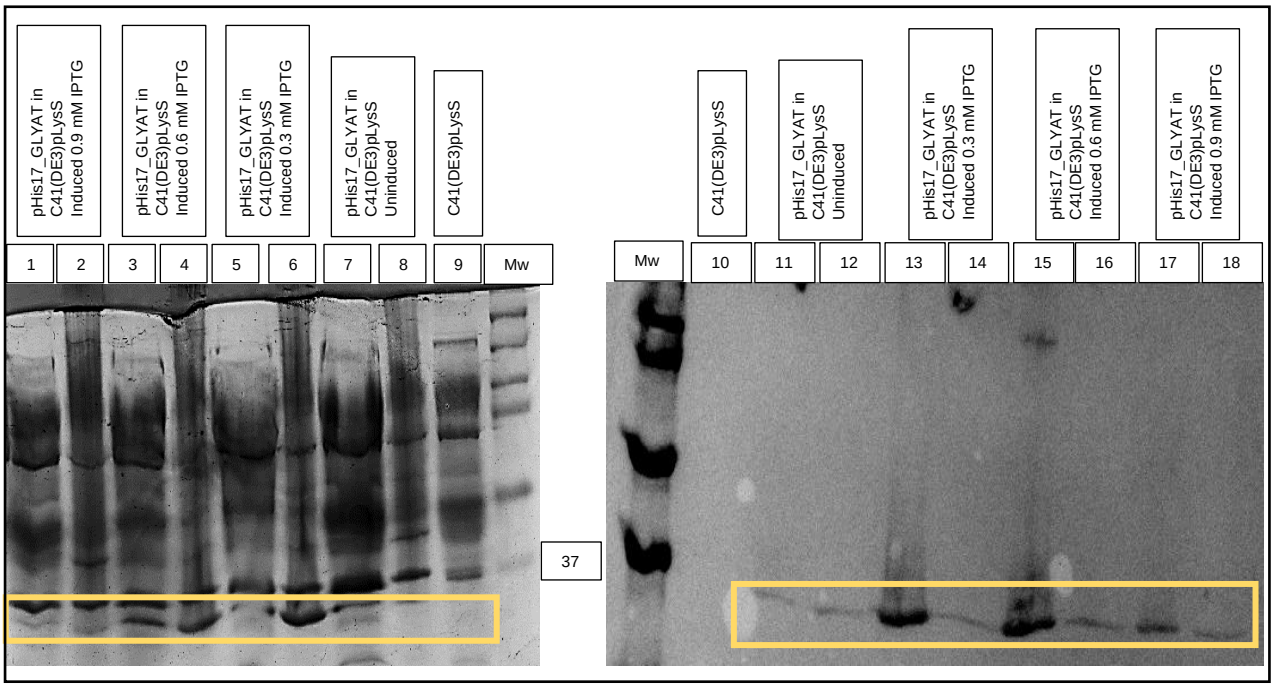


Figure 37: IPTG gradient for GLYAT expression.
Lanes 1, 3, and 5 contain the total fraction of the induced C41(DE3)pLysS cells with the IPTG gradient concentrations. Lanes 2, 4, and 6 contain the soluble fraction of the induced C41(DE3)pLysS cells with the IPTG gradient concentrations. Lane 7 contains the total fractions of the uninduced C41(DE3)pLysS cells, and lane 8 contains the soluble fraction of the uninduced C41(DE3)pLysS cells. Lane 9 contains the soluble fraction of the C41(DE3)pLysS cells without the pHis17_GLYAT. The Mw-lanes contain the Precision Plus Protein™ Dual Color Standards (Bio-Rad). The nitrocellulose membrane is represented by lanes 10-18. Lane 10 contained the soluble fraction of the C41(DE3)pLysS cells without the pHis17_GLYAT vector. Lane 11 contains the total fraction of the uninduced C41(DE3)pLysS cells, and lane 12 contains the soluble fraction of the uninduced C41(DE3)pLysS cells. Lanes 13-18 contain the induced C41(DE3)pLysS cells with the respective IPTG concentrations.

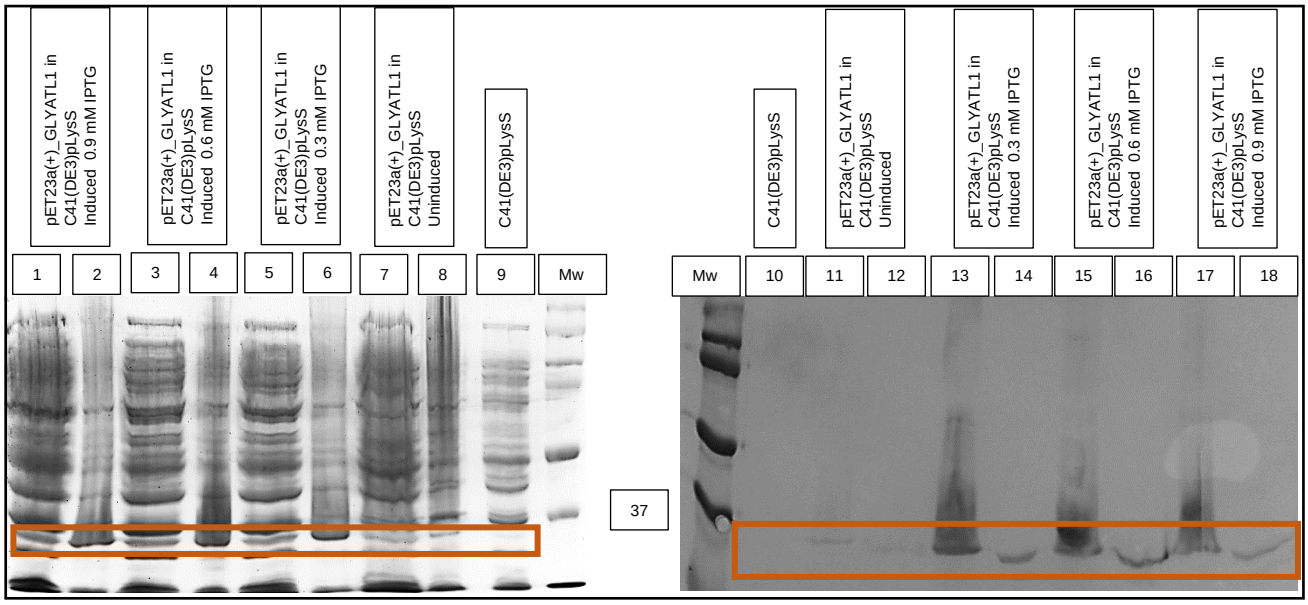


Figure 38: IPTG gradient GLYATL1.
Lanes 1, 3, and 5 contain the total fraction of the induced C41(DE3)pLysS cells with the IPTG gradient concentrations. Lanes 2, 4, and 6 contain the soluble fraction of the induced C41(DE3)pLysS cells with the IPTG gradient concentrations. Lane 7 contains the total fractions of the uninduced C41(DE3)pLysS cells, and lane 8 contains the soluble fraction of the uninduced C41(DE3)pLysS cells. Lane 9 contains the soluble fraction of the C41(DE3)pLysS cells without the pET23a(+)_GLYATL1. The Mw-lanes contain the Precision Plus Protein™ Dual Color Standards (Bio-Rad). The nitrocellulose membrane is represented by lanes 10-18. Lane 10 contained the soluble fraction of the C41(DE3)pLysS cells without the pET23a(+)_GLYATL1 vector. Lane 11 contains the total fraction of the uninduced C41(DE3)pLysS cells, and lane 12 contains the soluble fraction of the uninduced C41(DE3)pLysS cells. Lanes 13-18 contain the induced C41(DE3)pLysS cells with the respective IPTG concentrations.

Because the optimal IPTG induction concentration could be determined for both GLYAT and GLYATL1, the next step would be to determine the optimal induction temperature to obtain maximal protein yields.

3.3.3 Determining the optimal induction time to express soluble GLYAT and GLYATL1

After the optimal IPTG concentration was determined for both GLYAT and GLYATL1 (0.6 mM), a time gradient was performed to determine the optimal induction time at which soluble GLYAT and GLYATL1 were maximally expressed.

GLYAT and GLYATL1 were indistinguishable from the background proteins on the Coomassie-stained gels, and therefore, western blots were performed to confirm the presence of His-tagged GLYAT and His-tagged GLYATL1.

The highest yield was observed after allowing expression for 24 hours for His-tagged GLYAT and at 16 hours for His-tagged GLYATL1. GLYAT was expressed at the correct size of 34.74 kDa, and GLYATL1 was correctly expressed at 36.2 kDa. The expression of GLYAT is shown in Figure 39, whereas the expression of GLYATL1 is shown in Figure 40.

The image in Figure 39 was accidentally taken skew on the imager and was only realised when the nitrocellulose membrane was already overdeveloped. Therefore, it looks skewed in

the image as well. If it was to be straightened, the bands would appear above the 37 kDa size marker.

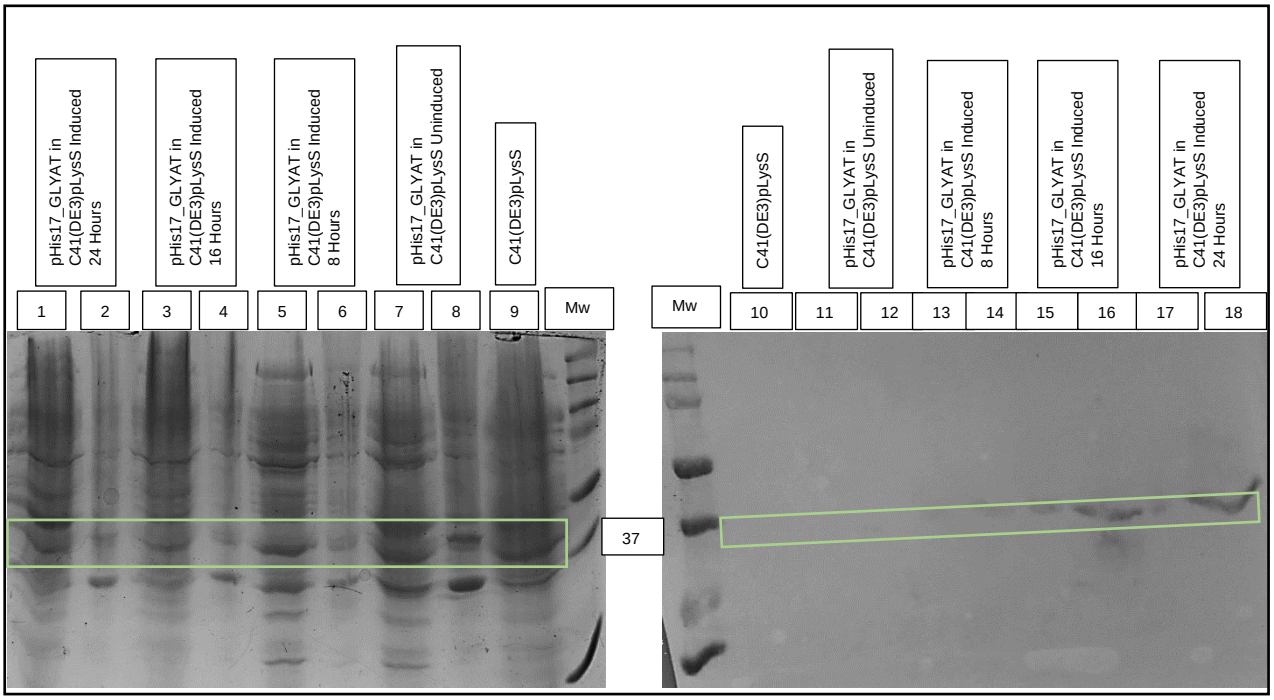


Figure 39: Time gradient of GLYAT expression.
Lanes 1, 3, and 5 contain the total fraction of the induced C41(DE3)pLysS cells at specific time intervals. Lanes 2, 4, and 6 contain the soluble fraction of the induced C41(DE3)pLysS cells at specific time intervals. Lane 7 contains the total fractions of the uninduced C41(DE3)pLysS cells, lane 8 contains the soluble fraction of the uninduced C41(DE3)pLysS cells. Lane 9 contains the soluble fraction of the C41(DE3)pLysS cells without the pHis17_GLYAT. The Mw-lanes contain the Precision Plus Protein™ Dual Color Standards (Bio-Rad). The nitrocellulose membrane is represented by lanes 10-18. Lane 10 contained the soluble fraction of the C41(DE3)pLysS cells without the pHis17_GLYAT vector. Lane 11 contains the total fraction of the uninduced C41(DE3)pLysS cells and lane 12 contains the soluble fraction of the uninduced C41(DE3)pLysS cells. Lanes 13-18 contain the induced C41(DE3)pLysS cells at the respective time intervals.

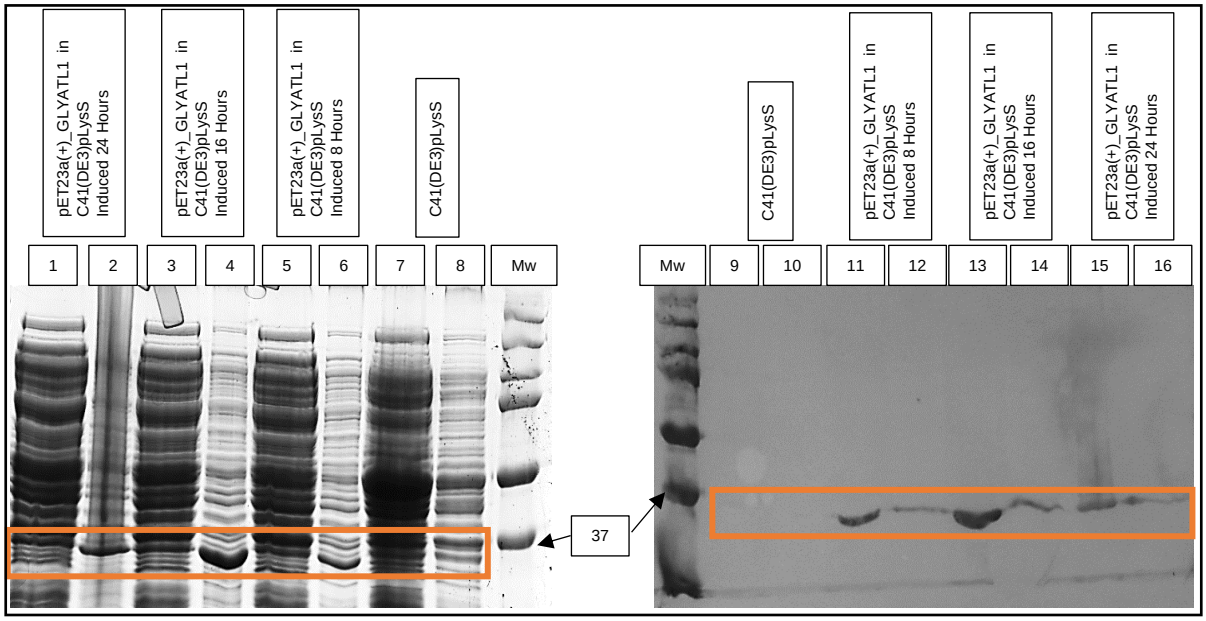


Figure 40: Time gradient of GLYATL1.
Lanes 1, 3, and 5 contained the total fraction of the induced C41(DE3)pLysS cells at specific time intervals. Lanes 2, 4, and 6 contained the soluble fraction of the induced C41(DE3)pLysS cells at specific time intervals. Lane 7 contained the total fraction of the C41(DE3)pLysS cells without the pET23a(+)_GLYATL1. Lane 8 contained the soluble fraction of the C41(DE3)pLysS cells without the pHis17_GLYAT. The Mw-lanes contain the Precision Plus Protein™ Dual Color Standards (Bio-Rad). The nitrocellulose membrane is represented by lanes 9-16. Lane 9 contains the total fraction of the C41(DE3)pLysS cells without the pET23a(+)_GLYATL1. Lane 10 contained the soluble fraction of the C41(DE3)pLysS cells without the pHis17_GLYAT vector. Lanes 11-16 contain the induced C41(DE3)pLysS cells at specific intervals.

Since the expression system for GLYAT and GLYATL1 was optimised, purification could now be performed to retrieve the soluble His-tagged GLYAT and His-tagged GLYATL1.

3.3.4 Purification of GLYAT and GLYATL1 from the cleared lysate of C41(DE3)pLysS cells

Soluble GLYAT was purified after expression using the optimal induction time and IPTG concentration. The purification was successful and can be seen in Figure 41 for His-tagged GLYAT, marked with a yellow rectangle, and Figure 42 for His-tagged GLYATL1, marked with an orange rectangle. Both GLYAT and GLYATL1 had to be purified and concentrated to be compared to previous literature. Furthermore, GLYAT and GLYATL1 cannot be quantified in the soluble lysate and, therefore, need to be purified for quantification.

GLYAT and GLYATL1 were, therefore, purified and concentrated to make this comparison between the results obtained and previous literature. The purification was performed as discussed in section 3.2.2.5.

SDS–PAGE analysis was performed using the stored fractions of purified His-tagged GLYAT and His-tagged GLYATL1 in two separate purification columns, as discussed in sections 2.2.2.12 and 2.2.2.13. Only the western blot for the purified His-tagged GLYAT is displayed

here because no distinction between protein bands could be made on the Coomassie-stained gel of the His-tagged GLYAT. Both the Coomassie-stained gel and the western blot are shown for His-tagged GLYATL1.

Most of the soluble His-tagged GLYAT and soluble His-tagged GLYATL1 were eluted in the first elution fraction. His-tagged GLYAT is at the correct size of 34.74 kDa. His-tagged GLYATL1 was also 36.2 kDa in size. The positive control (His-tagged GLYAT previously purified in another MSc study) in lanes 8 (Figure 41) and 16 (Figure 42) confirms that the purified GLYAT and GLYATL1 are at the correct size.

Some His-tagged GLYAT and His-tagged GLYATL1 remained in the post binding, wash 1 and wash 2 steps. The purified His-tagged GLYAT and His-tagged GLYATL1 were stored at 4°C until needed.

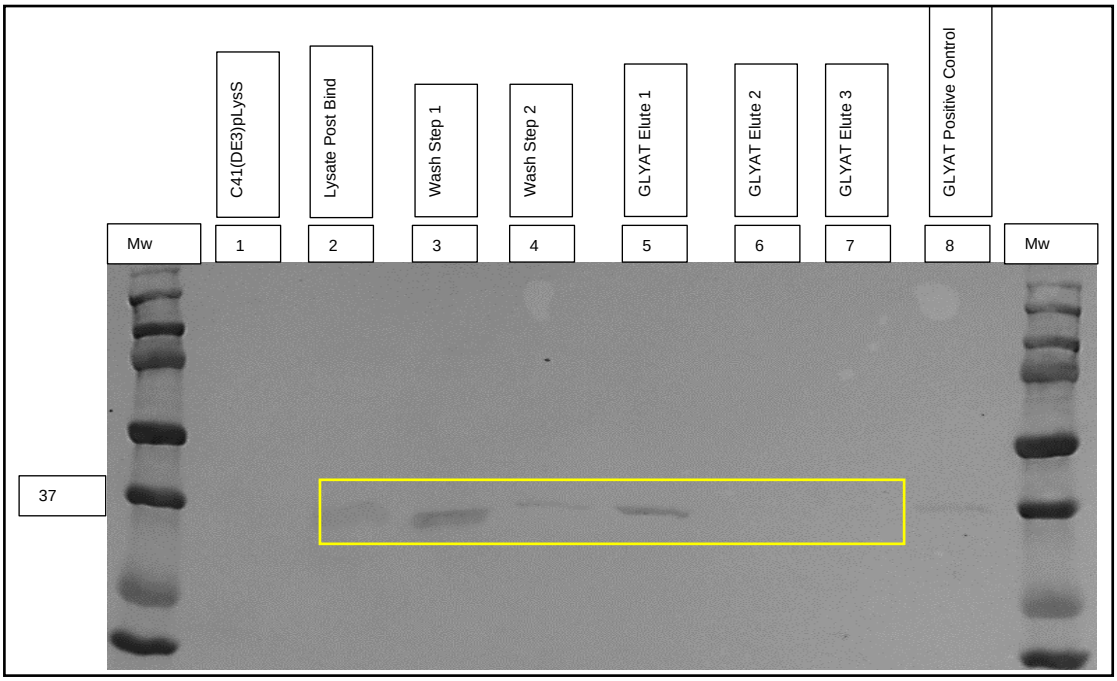


Figure 41: The nitrocellulose membrane of the purification of GLYAT from the C41(DE3)pLysS after induction at the optimal expression conditions.
The Mw-lanes contain the Precision Plus Protein™ Dual Color Standards (Bio-Rad). Lane 1 contained the soluble fraction of the C41(DE3)pLysS cells without the pHis17_GLYAT. Lane 2 contained the fraction of the post-bind step, lane 3 contained the fraction of the first wash step, lane 4 contained the fraction of the second wash step, lane 5 contained the fraction of the first, lane 6 contained the fraction of the second elution, lane 7 contained the fraction of the third elution. Lane 8 contained the GLYAT positive control.

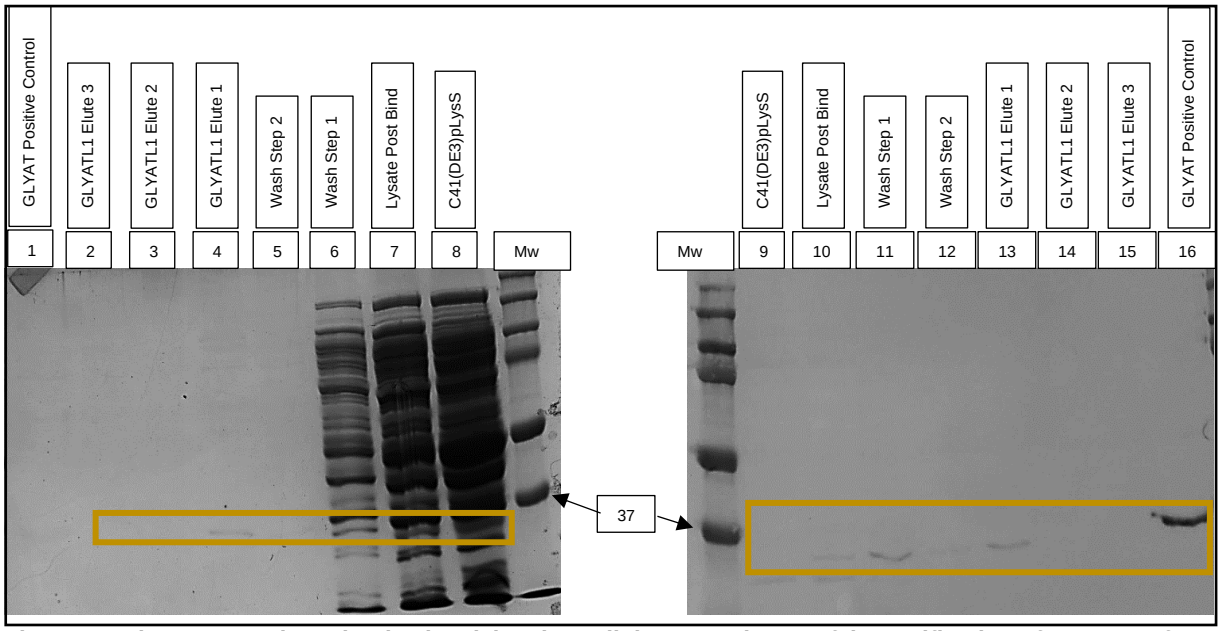


Figure 42: The Coomassie-stained gel and the nitrocellulose membrane of the purification of GLYATL1 from the C41(DE3)pLysS after induction at the optimal expression conditions.
The image on the left is the Coomassie-stained gel and the right is the nitrocellulose membrane. Lane 1 contained the positive GLYAT control, lane 2 contained the fraction of the third elution, lane 3 contained the fraction of the second elution, lane 4 contained the fraction of the first elution, lane 5 contained the fraction of the second wash step, lane 6 contained the fraction of the first wash step, lane 7 contains the post-bind fraction, and lane 8 contained the soluble lysate of the C41(DE3)pLysS cells without the pHis17-GLYAT. The Mw-lanes contain the Precision Plus Protein™ Dual Color Standard (Bio-Rad). Lane 9 contained the soluble lysate of the C41(DE3)pLysS cells without the pHis17_GLYAT. Lane 10 contained the post-bind fraction, lane 11 contained the fraction of the first wash step, lane 12 contained the fraction of the second wash step, lane 13 contained the fraction of the first elution, lane 14 contained the fraction of the second elution, lane 15 contained the fraction of the third elution, and lane 16 contained the positive GLYAT control.

Since both His-tagged GLYAT and His-tagged GLYATL1 were successfully purified, the next step was to concentrate the proteins and exchange the storage and reaction buffer because the purification buffer causes problems during the GC–MS sample preparation.

3.3.5 Concentration and buffer exchange of purified GLYAT and GLYATL1

The purified His-tagged GLYAT and His-tagged GLYATL1 were concentrated for use in downstream GC-MC analysis. The buffer was changed because the elution buffer containing a mixture of salts posed a problem in downstream GC–MS sample preparation steps. These salts are concentrated during the sample preparation step, as discussed in section 3.2.2.6. This influences the EIC during GC–MS analysis (Silvestro *et al.*, 2013). The chosen exchange buffer was then Tris-HCl (pH 8.0).

The concentrated His-tagged GLYAT is shown in Figure 43, and the concentrated His-tagged GLYATL1 is shown in Figure 44. Concentrated His-tagged GLYAT and His-tagged GLYATL1 were present at the correct size and seen in the Coomassie-stained gel and the western blot. The presence of the positive control (His-tagged GLYAT purified from another MSc study) confirms the correct size of the expressed GLYAT. Some contaminating bands can be seen

on the Coomassie-stained gels. The contaminating proteins were larger than the molecular weight cut-off value of the protein concentrator sample tubes used in section 3.2.2.6. For example, when the purified protein sample was concentrated, contaminating proteins larger than 30 kDa were not removed. These proteins may be residual proteins carried over from the purification process, which indicates that these proteins also have an affinity toward the Ni^{2+}-charged purification columns. These contaminating proteins may be removed by adding an additional step during protein purification, as discussed in section 3.2.2.5. The additional step would be to further concentrate the concentrated GLYAT and GLYATL1 samples with a higher MWCO protein concentrator unit (such as 50 MWCO) using the same buffer continuously. GLYAT and GLYATL1 would be small enough to pass through the larger MWCO protein concentrator unit, but the contaminating proteins would remain in the leftover buffer on top of the protein concentrator unit. Thus, the flow-through with only GLYAT and GLYATL1 would not contain the contaminating proteins and, therefore, be purer.

The wrong sample of GLYATL1 was loaded into lanes 2 and 5 of Figure 43. Therefore, SDS–PAGE analysis was performed again, as discussed in sections 2.2.2.12 and 2.2.2.13, with the concentrated His-tagged GLYATL1 sample shown in Figure 44.

During the concentration of the purified His-tagged GLYAT and His-tagged GLYATL1, the buffer was changed to Tris-HCl (pH 8.0), which was selected as a suspension buffer because the protein activity is at a more basic pH.

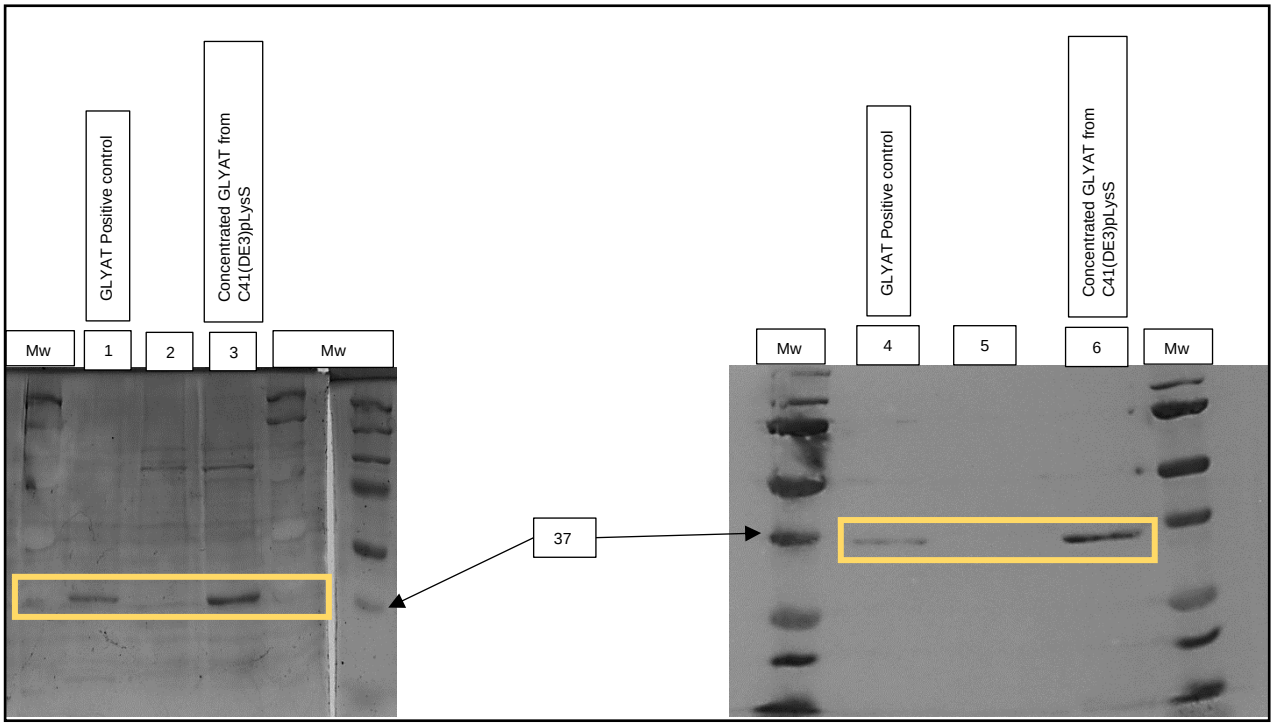


Figure 43: Concentration of GLYAT from the purified first elution fraction.
The image on the left contained the Coomassie-stained gel, whereas the right contained the nitrocellulose membrane. The Mw-lanes contained the Precision Plus Protein™ Dual Color Standards. Lane 1 contained the positive GLYAT control. Lane 3 contained the concentrated GLYAT from the purified fraction. Lane 4 contained the positive GLYAT control, and lane 6 contained the concentrated GLYAT from the purified fraction.

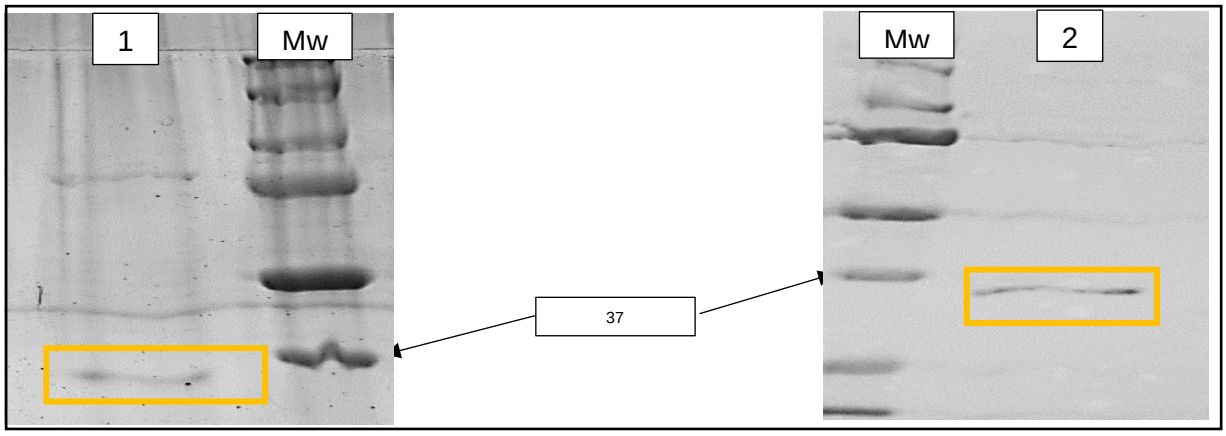


Figure 44: Concentration of GLYATL1 from the first elution fraction of the purification.
The image on the left is the Coomassie-stained gel and the right was the nitrocellulose membrane. Lane 1 and 2 contained the concentrated GLYATL1 from the first elution fraction. The Mw-lanes contained the Precision Plus Protein™ Dual Color Standards.

Now that both proteins were concentrated successfully and the buffer was changed, quantification can thus be performed to determine the concentration of GLYAT and GLYATL1 in the concentrated samples.

3.3.6 Protein quantification results from the Qubit™ 2.0 fluorometer

The protein samples were quantified as discussed in section 3.2.2.7 before GC–MS analysis to compare the same concentration of enzyme in each enzyme reaction that was set up with the concentrated protein. When GC–MS was performed with the concentrated His-tagged GLYAT and His-tagged GLYATL1 reactions, the results could be related to a quantified enzyme concentration within the reactions. It was not possible to determine the enzyme concentrations within the cell lysate due to the cell lysate containing many other cellular expressed proteins.

Therefore, the concentrated His-tagged GLYAT and His-tagged GLYATL1 were quantified in Tris-HCl (pH 8.0) buffer. The concentrated protein concentration (Table 25) was determined with a specific formula seen below after the value was obtained from the Qubit™ 2.0 fluorometer:

→ Qf – the sample value read from Qubit™ 2.0.

→X - µL of sample added to the reaction

$$[sample] = Qf\ value \times \frac{200}{x}$$

Table 25: Quantification results from the Qubit™ 2.0 Fluorometer.

Sample	Qf (µg/mL)	X (µL)	Sample concentration (µg/mL)
GLYAT	1.82	10	36.4
GLYATL1	2.47	10	49.4

3.4 Section summary

The goal of this chapter of the study was to establish and optimise an expression system for soluble His-tagged GLYAT and His-tagged GLYATL1. This was attempted using the already available pHis17_GLYAT vector and the purchased pET23a(+)_GLYATL1 vector.

The vectors were transformed into C41(DE3)pLysS cells, and the sequences were determined for both genes *(GLYAT* and *GLYATL1*). The sequence alignments indicated that the

sequences for both *GLYAT* and *GLYATL1* were correct and in the correct orientation with no mutations in the sequences. Therefore, the optimisation of the expression conditions could be done next.

Both His-tagged GLYAT and His-tagged GLYATL1 were expressed very well. Therefore, the third objective was to identify and optimise an expression system that would result in the expression of soluble His-tagged GLYAT, and His-tagged GLYATL1 was achieved successfully.

This chapter's findings will still contribute to the qualitative characterisation of *GLYAT* and one of its paralogues, *GLYATL1*, after being expressed as soluble His-tagged GLYAT or His-tagged GLYATL1. This system is optimised for the expression of soluble GLYAT and GLYATL1 and can, therefore, be used to express different variants of both *GLYAT* and *GLYATL1* to try and characterise the effect of different variants on the enzyme activity of GLYAT and GLYATL1. It might also be used to confirm the involvement of isovaleryl-CoA in the glycine conjugation pathway. Moreover, it may also be determined if *GLYAT* and its paralogues work in synergy to eliminate different types of xenobiotics-CoA and OA-CoA molecules. This can be determined when a mixture of the substrates is present in reactions with either GLYAT or GLYATL1. The substrate and product analysis of these reactions may then elucidate which enzyme detoxifies which type of xenobiotics by determining the consequent decrease in substrate and increase in products when one of the enzymes is present in the reactions.

Chapter 4: Analysis of the fluctuation of selected metabolites detoxified by the glycine conjugation pathway using an established GC–MS method

4.1 Introduction

GC–MS was chosen to analyse the biotransformation of selected metabolites of the glycine conjugation pathway in a bacterial C41(DE3)pLysS expression host. The original aim of the study was to determine the biotransformation of specific metabolites in the presence of soluble biologically active ACSM2B and GLYAT. However, this aim could not be completed since only soluble GLYAT could be successfully expressed in the pETDuet-1 expression system. Therefore, the focus shifted to the analyses of the specific metabolites (benzoyl-CoA, phenylacetyl-CoA, and isovaleryl-CoA) biotransformed by GLYAT or GLYATL1. In this chapter, the xenobiotic-CoA and OA (organic acid)-CoA decrease and hippuric acid increase were measured. The GC–MS method was used to analyse the enzymes expressed by the pETDuet-1_ACSM2B/GLYAT expression system and the pHis17_GLYAT and pET23a(+)_GLYATL1 systems.

GLYAT was expressed with a His-tag rather than an S-tag for purification purposes. Therefore, two systems expressed soluble GLYAT, and one expressed soluble GLYATL1. The GLYATL1 expressed from the second system also contained a His-tag for purification purposes. GLYATL1 activity was measured here due to the suspected involvement in the glycine conjugation pathway and the elimination of isovaleryl-CoA from patients who suffer from isovaleric acidemia.

Reactions with xenobiotic and OA substrates were, therefore, set up containing the respective preferred amino acid substrate. For example, GLYAT prefers to conjugate xenobiotics to glycine (Matsuo *et al.*, 2012; Rohwer *et al.*, 2021; Schutte *et al.*, 2019; van der Sluis *et al.*, 2013; van der Sluis *et al.*, 2017; van der Westhuizen *et al.*, 2000), whereas GLYATL1 prefers to conjugate xenobiotics with glutamine (Matsuo *et al.*, 2012). The xenobiotic substrates analysed were benzoyl-CoA and phenylacetyl-CoA, and the OA substrate was isovaleryl-CoA.

The sample reactions were prepared for GC–MS analysis and derivatized. Derivatization was necessary for the metabolites to be detected by GC–MS. Silylation was used to derivatise the metabolites after sample preparation for this study. This step was performed for all the reactions to be run on the GC–MS system. Silylation increased the volatility and decreased the polarity of the compounds. Moreover, it increased the compounds' stability and improved the GC behaviour (Moldoveanu & David, 2018). A general GC–MS system configuration was followed thereafter to analyse the derivatised samples. A simplified reaction setup diagram

can be seen in Figure 45. The reactions were set up with the soluble lysate containing GLYAT or GLYATL1 and the purified concentrated GLYAT and GLYATL1 in Tris-HCl (pH 8.0).

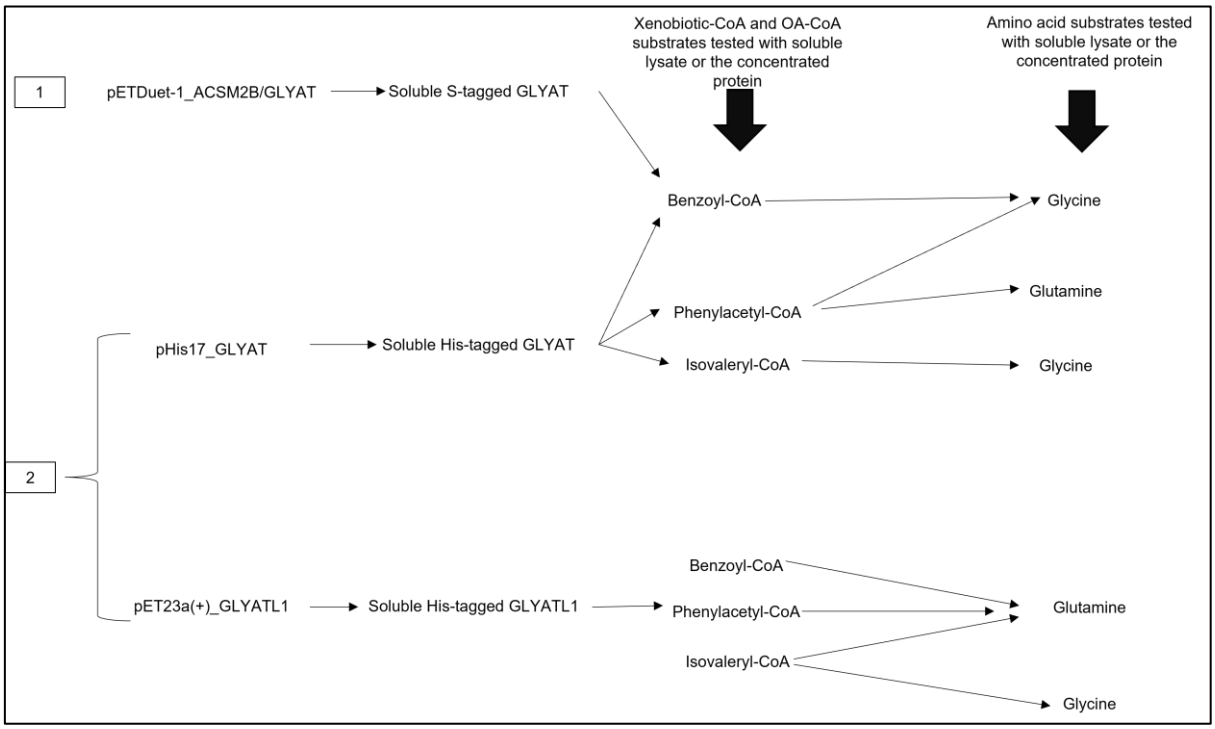


Figure 45: Reaction set-up of the soluble lysate or concentrated proteins with the xenobiotic-CoA, OA-CoA, and amino acid substrates.
Number 1 indicates the pETDuet-1 expression system and number 2 indicates the pHis17_GLYAT and pET23a(+)_GLYATL1 expression system. The next column indicates the xenobiotic-CoA or OA-CoA substrate, and the last column indicates the amino acid substrate.

The soluble lysate and the purified concentrated GLYAT and GLYATL1 were then compared to observe whether biologically active GLYAT and GLYATL1 were expressed and yielded similar results in the purified concentrated enzymes as in the soluble lysate reactions. Furthermore, the results were also compared to previous literature that determined the enzyme kinetic parameters of GLYAT (Matsuo *et al.*, 2012; Rohwer *et al.*, 2021; Schutte *et al.*, 2019; van der Sluis *et al.*, 2013; van der Sluis *et al.*, 2017; van der Westhuizen *et al.*, 2000) and GLYATL1 (Matsuo *et al.*, 2012).

4.2 Materials and methods

4.2.1 Materials

A table containing the reagents needed and the manufacturers can be seen in Appendix A.

Stock solutions

Stock solutions (0.5 mM) were created of benzoyl-CoA, isovaleryl-CoA, phenylacetyl-CoA, hippuric acid, and isovalerylglycine using nuclease-free ddH_2O.

Then, 20 mg/mL MOX (methyloxyamine hydrochloride) (Sigma–Aldrich) in pyridine (Sigma–Aldrich) was prepared.

1% BSTFA (N,O-bis(trimethylsilyl)trifluoroacetamide with trimethylchlorosilane) for GC derivatisation (Sigma–Aldrich) was purchased.

Stock solutions of 0.5 M Tris-HCl (pH 8.0) and 0.5 M glycine were made with ddH_2O.

4.2.2 Methods for the GC–MS analysis of the conversion reactions in the soluble lysate containing either GLYAT or GLYATL1

4.2.2.1 Cell lysis and lysate preparation of samples to be used for GC–MS analysis

The prepared cell pellet for GC–MS analysis was resuspended in 5 mL double distilled water. Glass beads were then added in a 2:1 (lysate: glass beads) ratio into 2 mL microcentrifuge tubes. For example, 1 mL resuspended pellet was added to ~500 µL glass beads. One microliter of Pierce® Universal Nuclease For Cell Lysis (5 kU) was added to the resuspended cells.

The 2 mL microcentrifuge tubes were then placed into a Retsch (MM400) vibration mill for 40-50 minutes at a frequency of 30.0 (1/s). After lysis, the cells were left for 20 minutes for further cell disruption. After that, 5 µL of the total fraction was removed from the lysate and stored at 4°C until needed. The soluble fraction was obtained by centrifuging the lysate at 15000 x g for 20 minutes at 4°C. The clear soluble fractions were placed into new 1.5 mL microcentrifuge tubes and stored at 4°C if not used immediately.

The cell pellet used for GC–MS was not lysed with lysis buffer due to the ingredients of the prepared lysis buffer interfering with the lysate preparation step before GC–MS.

4.2.2.2 Retention time determination of substrates and expected products

A generalised untargeted method for organic acid detection was used, as discussed in section 4.2.2.5. Therefore, the method was not optimised and was only used for the qualitative observation of the conversion reactions, e.g., benzoyl-CoA to hippuric acid. Benzoic acid (the hydrolysed form of benzoyl-CoA) and hippuric acid were qualitatively observed in this case.

The retention time of the substrates and detectable products were evaluated before setting up reactions for the GC–MS. This was done to observe the retention time at which the substrates/products would elute. The substrate (benzoyl-CoA) and products, hippuric acid and isovalerylglycine, were used. These samples were prepared by adding 1 µL of the 0.5 mM stock solutions (isovaleryl-CoA, benzoyl-CoA, hippuric acid, and isovalerylglycine) to 199 µL nuclease-free ddH_2O. The reactions were prepared for GC–MS analysis, as discussed in section 4.2.2.4.

The substrate (phenylacetyl-CoA) and other products (phenylacetylglycine, phenylacetylglutamine, isovalerylglutamine and benzoylglutamine) were not available in our laboratory. It was, however, attempted to perform a NIST (National Institute of Standards and Technology) library search to observe where the unavailable products eluted. This was done using the NIST MS Search (v.2.) software.

After the retention times were determined, the reactions were set up with the soluble and purified concentrated GLYAT and GLYATL1 and were run on the GC–MS.

4.2.2.3 Enzyme reaction setup for GLYAT and GLYATL1 in the soluble lysate of the cell

After the retention time was determined for the substrates and products measured in this study, as discussed in section 4.2.2.2, reactions containing soluble GLYAT and GLYATL1 were set up with the soluble lysate. The soluble lysates used were obtained from section 4.2.2.1. Negative controls of each substrate were added to the study, first to determine whether there were matrix effects in the soluble lysates that did not contain the soluble GLYAT or GLYATL1, and second to be compared to the reactions that did contain soluble GLYAT or GLYATL1. These comparisons were used to observe the substrate decrease (benzoyl-CoA, isovaleryl-CoA, and phenylacetyl-CoA) and product increase (hippuric acid, benzoylglutamine, isovalerylglycine, isovalerylglutamine, phenylacetylglycine, and phenylacetylglutamine).

Enzyme reactions were set up according to van der Sluis *et al.* (2017) with some amendments. Table 26 indicates the reaction setup for GC–MS analysis. These reactions were set up using the soluble lysate of the C41(DE3)pLysS cells that contained either GLYAT or GLYATL1. To avoid batch effects on the results obtained, the C41(DE3)pLysS cells were doubled in volume during expression at the optimal IPTG concentration and time. The cells were then lysed, as discussed in section 4.2.2.1. The soluble fraction of the cells was then obtained. Half of the soluble fraction was used for GC–MS analysis reaction setup, and the other half was used for purification, concentration, and buffer exchange of the soluble GLYAT and GLYATL1. Therefore, all reactions were performed with the same batch of expressed GLYAT or GLYATL1.

The soluble lysates were used for GC–MS analysis first. The concentrated GLYAT and GLYATL1 were stored at 4°C until needed. When the concentrated GLYAT and GLYATL1 were used for GC–MS analysis, all the reactions were set up with the same conditions for that of soluble lysate.

Table 26: The enzyme reaction setup for GLYAT and GLYATL1 with the soluble lysate.

Reagent	Final concentration	Reaction volume (µL)	Volume of substrate added (µL)
Benzoyl-CoA or Isovaleryl-CoA or Phenylacetyl-CoA	2.5 µM	200	1
Tris-HCl (pH 8.0)	25 mM		10
Glycine or glutamine	2.5 mM		1
Soluble lysate with either GLYAT or GLYATL1	-		188

The soluble lysate containing GLYAT or GLYATL1 was used, and the concentrations of the reactions were as follows: 2.5 µM of the respective substrate (Benzoyl-CoA, Isovaleryl-CoA, or phenylacetyl-CoA), 2.5 mM glycine and 25 mM Tris-HCl (pH 8.0). The samples that only contained benzoyl-CoA were run for 6 minutes. The samples that only contained isovaleryl-CoA were run for 20 minutes, and the samples that had both benzoyl-CoA and isovaleryl-CoA were run for 20 minutes. The samples that contained phenylacetyl-CoA were run for 30 minutes. GLYAT was always mixed with glycine as a substrate unless otherwise stated, and GLYATL1 was always mixed with glutamine as a substrate unless otherwise stated.

The samples were set up as shown in Table 27.

Table 27: The reactions and reaction combinations setup when using the soluble lysate of the induced cells and the soluble lysate of the cells that did not contain GLYAT or GLYATL1.

Enzyme	Reason	Substrate	Run time (minutes)
No enzyme	Negative controls and to determine detection in the soluble lysate	Benzoyl-CoA	6
No enzyme		Isovaleryl-CoA	20
No enzyme		Phenylacetyl-CoA	30
GLYAT	Observe benzoic acid decrease and hippuric acid increase	Benzoyl-CoA	6
GLYAT	Same as above but also to observe the effect of isovaleryl-CoA on benzoyl-CoA elimination	Benzoyl-CoA and Isovaleryl-CoA	20
GLYAT	Observe isovaleric acid decrease and isovalerylglycine increase	Isovaleryl-CoA	20
GLYAT	Observe phenylacetic acid decrease and phenylacetylglycine increase	Phenylacetyl-CoA	30
GLYAT	Observe whether GLYAT can use glutamine to convert phenylacetyl-CoA but also measure the decrease of phenylacetic acid and increase of phenylacetylglutamine	Phenylacetyl-CoA and glutamine	30
GLYATL1	Observe whether GLYATL1 can convert benzoyl-CoA and	Benzoyl-CoA	6

	observe the decrease of benzoic acid and increase of hippuric acid		
GLYATL1	Observe the effect of benzoyl-CoA in phenylacetyl-CoA elimination and measure the decrease of phenylacetic acid and increase of phenylacetylglutamine	Benzoyl-CoA and Phenylacetyl-CoA	30
GLYATL1	Observe the decrease of phenylacetic acid and the increase of phenylacetylglutamine	Phenylacetyl-CoA	30
GLYATL1	Observe whether GLYATL1 can convert isovaleryl-CoA with glycine	Isovaleryl-CoA and glycine	20
GLYATL1	Observe whether GLYATL1 can covert isovaleryl-CoA with glutamine	Isovaleryl-CoA	20

4.2.2.4 Sample preparation for GC–MS

After lysis of the cells and supernatant collection as discussed in section 4.2.2.1, the reactions in the soluble cell lysate in Table 27 were completed, ice-cold 100% methanol was added in a 1 sample:3 methanol (v/v) ratio and centrifuged for 20 minutes at 15000 x g at room temperature. The samples were dried using nitrogen gas at 60°C for 45-60 minutes before derivatisation in clear Screw Cap Vials 2 mL (Agilent Technologies). The dried samples can be stored at 4°C if not used immediately.

Derivatization was performed by adding 50 µL of 1% BSTFA and 50 µL of MOX in pyridine to the dried samples and left at room temperature for 60 minutes. MOX is added to form oxime

derivatives of molecules that contains either aldehyde or ketone groups. This step is preformed prior to silylation. After that, the mixture was pipetted into vial inserts (250 µL pulled point glass) (Agilent Technologies) and placed back into the screw cap vials with lids screwed on afterwards. The samples were then ready for GC–MS analysis.

4.2.2.5 GC–MS method

A similar method to Reinecke *et al.* (2012) and Erasmus *et al.* (2019), which detected organic acids in urinary samples, was used to detect the xenobiotic and organic acid substrates and conjugated products in both the soluble lysate and the purified concentrated GLYAT and GLYATL1. The substrates that were targeted were benzoic acid (hydrolysed form of benzoyl-CoA), isovaleric acid (hydrolysed form of isovaleryl-CoA), and phenylacetic acid (hydrolysed form of phenylacetyl-CoA). The amino acid conjugates of these substrates were also targeted. These were benzoylglycine, benzoylglutamine, isovalerylglycine, isovalerylglutamine, phenylacetylglycine and phenylacetylglutamine.

Measuring the metabolites after derivatization is important to verify the biological activity of GLYAT and GLYATL1 and to observe any decreased substrates and increased product concentrations . The samples were first placed into the autosampler sample rack, and all external instrument checks were performed.

The metabolites were then measured with an Agilent Technologies 7890A front-end system with an Agilent Technologies 7693 autosampler coupled to a Pegasus HT TOFMS (High Throughput Time-of-Flight Mass Spectrometer) (LECO®). The Agilent Technologies 7693 autosampler injected 1 µL/run of the sample into the Restex Rxi-5Sil MS (30 m x 250 µm x 0.25 µm) column in the 10-split mode. The inlet temperature was kept at a constant 250°C during the run. Helium gas (He) was used as a mobile phase at a 1.5 mL/minute constant flow rate. The oven temperature started at 70°C for 2 minutes and was then increased to 120°C. The temperature increase was 7°C/minute.

After the initial increase, the oven temperature increased to 230°C. The temperature increase was 10°C/minute. The last oven temperature increase was to 300°C at 13°C/minute. The oven was then kept at 300°C for 2 minutes. Thereafter, the oven temperature returned to the initial temperature of 70°C. Each sample was run for a total of 22.3 minutes while the transfer line was kept at 222°C and the inlet was kept at 199°C.

Ionisation was performed with electron impact (EI^+) at an electron energy of -70 eV. This method had a solvent delay of 4 minutes and 10 seconds, followed by spectral data acquisition over a scan range of 50-950 amu.

Important note

Before each run, a Tune and Leak test was done to determine whether the GC–MS is functioning properly and that there aren't any leaks that may influence the results.

4.2.2.6 Data processing and visualisation of the EIC obtained from the GC–MS

LECO® ChromaTOF® software (v.4.50.8) was used to visualise the data and extract the ion chromatograms. The EICs were then overlaid with the negative controls to compare the substrate decrease and product increase.

4.2.3 Methods for the GC–MS analysis of the conversion reactions using concentrated GLYAT or GLYATL1

4.2.3.1 Determining if Tris-HCl (pH 8.0) is a good storage and reaction buffer for downstream experiments

Determining whether Tris-HCl (pH 8.0) would affect the samples prepared for GC–MS or interfere with the resulting output of the GC–MS is important in this study since GLYAT and GLYATL1 were stored in Tris-HCl (pH 8.0). If the buffer would affect the GC–MS system, a new reaction buffer must be tested to find the most adequate buffer to be used that will not interfere with the downstream GC–MS analysis, or the concentration should be drastically decreased to 25 mM Tris-HCl (pH 8.0), as previously reported (van der Sluis *et al.*, 2013; van der Sluis *et al.*, 2017).

Two samples were set up according to Table 28 to determine the efficacy of the Tris-HCl (pH 8.0) buffer.

Reagent	Initial concentration (M)	Final concentration (M)	Volume added (µL)	Total reaction volume (µL)
Reaction 1				
Tris-HCl (pH 8.0)	0.5	0.495	198	200
Glycine	0.5	0.005	2	
Reaction 2				
ddH_2O	-	-	198	200
Glycine	0.5	0.005	2	

Table 28: Storage and reaction setup of concentrated GLYAT and GLYATL1 in Tris-HCl (pH 8.0).

The samples were prepared for GC–MS analysis as discussed in section 4.2.2.4 and run on the GC–MS as discussed in section 4.2.2.5.4.3 Results and discussion of the GC–MS analysis of the enzyme reactions using the soluble lysate

4.3.1 GC–MS analysis of the pure substrate and some products to determine retention times

Before the metabolites for GLYAT could be determined, a pure substrate and some products were analysed (as discussed in section 4.2.2.2) by the chosen GC–MS method discussed in section 4.2.2.5. This was done to ensure that the retention times of the metabolites were within the parameters of the GC–MS method and that the chosen concentration to be added to each reaction as discussed in section 4.2.2.3 is detectable by the GC–MS technique. Therefore, these results were used to observe whether the GC–MS configuration needed to be adjusted.

When the EIC of the pure substrate and products were analysed, it could first be observed that the EIC indicated that the retention time for each metabolite was within the detection range of the chosen GC–MS method. Second, the chosen concentration (2.5 µM) was sufficient to be detected by GC–MS analysis. The EIC in Figure 46 represents the retention time for benzoic acid (hydrolysed form of benzoyl-CoA), Figure 47 represents the retention time of hippuric acid (benzoylglycine), and the EIC in Figure 48 represents the retention time for isovalerylglycine.

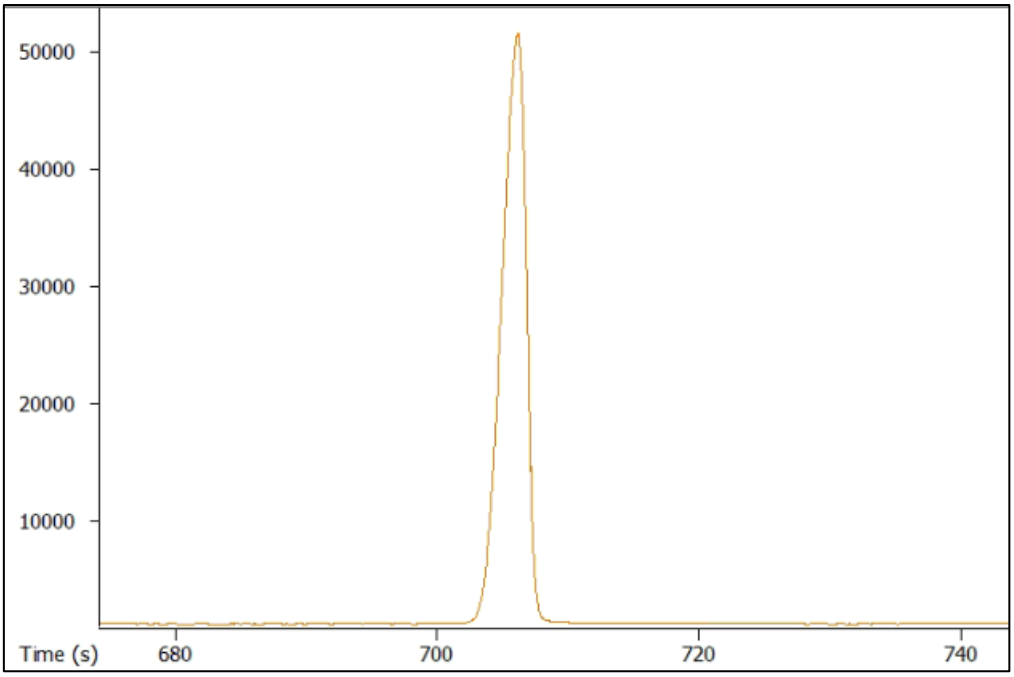


Figure 46: EIC of benzoic acid (hydrolysed form of benzoyl-CoA).

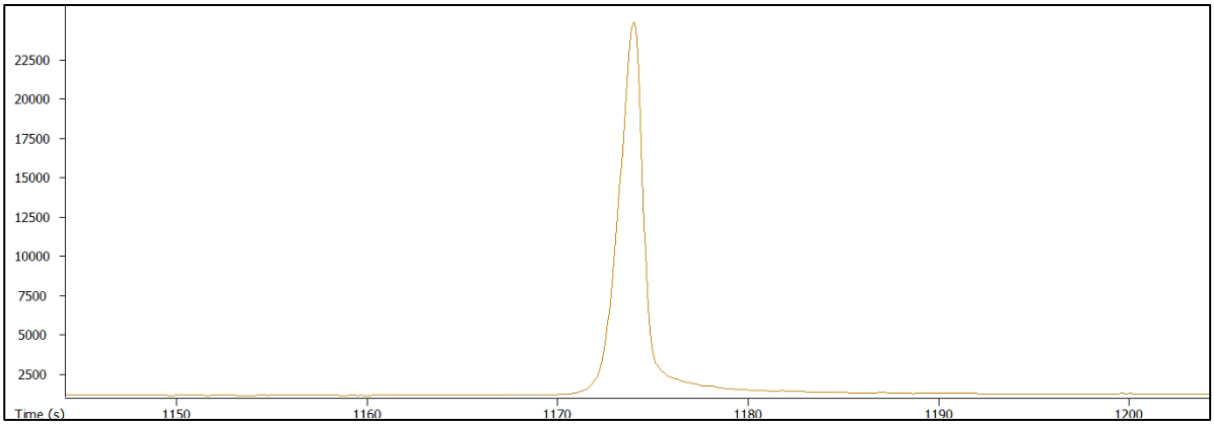


Figure 47: EIC of hippuric acid (benzoylglycine).

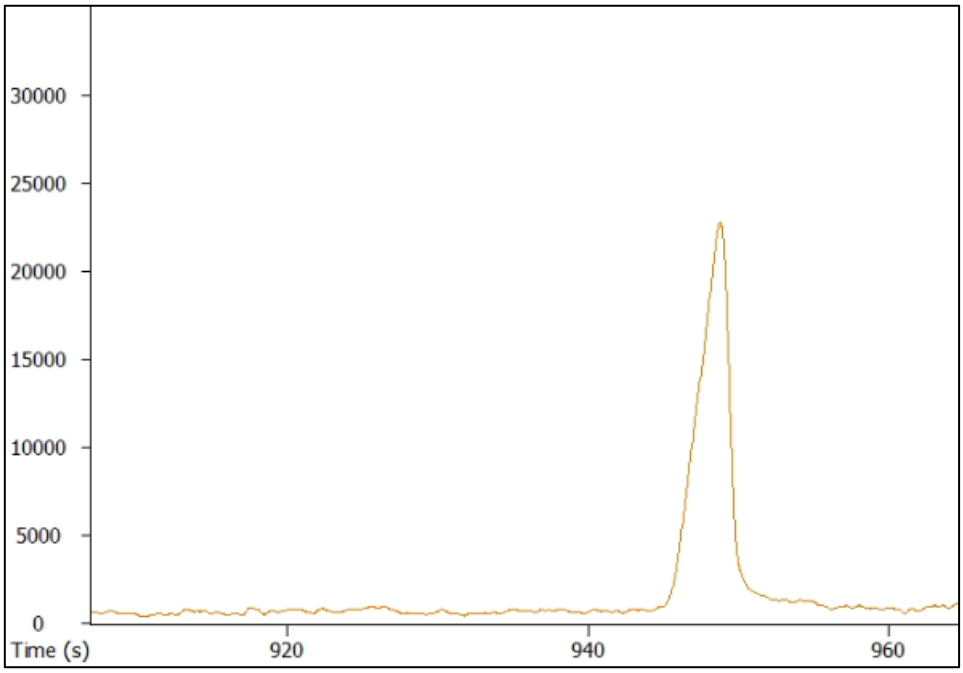


Figure 48: EIC of isovalerylglycine.

GC–MS with the electron impact (EI) ionisation technique is considered a hard ionisation technique. Therefore, during the hard ionisation step, the molecules in the gaseous mobile phase are split into their respective fragments (Kiseleva *et al.*, 2021). Moreover, GC–MS is unable to detect CoA esters mainly due to the size of CoA since GC–MS cannot detect large molecules due to its limited mass range (Vivekanandan-Giri *et al.*, 2011). For example, when benzoyl-CoA is ionised, the molecule is split into two representative fragments, namely, benzoic acid and CoA. Because GC–MS methods cannot detect CoA fragments, the only fragment related to benzoyl-CoA would be the hydrolysed form, benzoic acid. Therefore, the decrease in benzoic acid would be representative of a decrease in benzoyl-CoA. In the EIC of Figure 46 and Figure 47, it can be seen that benzoic acid (hydrolysed form of benzoyl-CoA) and hippurate elute at two different positions. Therefore, both benzoic acid (hydrolysed form

of benzoyl-CoA) that served as the substrate and hippurate that served as the product could be analysed separately.

After the retention times of some of the substrates and some products were determined, the next step was to set up reactions with the soluble lysate with and without soluble GLYAT.

4.3.2 Results of GC–MS analysis of GLYAT from the pETDuet-1 expression system cell lysate using the 105 m/z ion for metabolite change detection

The GC–MS analysis of the soluble lysate without GLYAT expression indicated that the matrix did not affect the retention times of the substrate and products. Therefore, this GC–MS method as discussed in section 4.2.2.5 was considered to be sufficient for GC–MS analysis of the reactions that were set up with the soluble lysates that contained GLYAT. The results in section 4.3.1 confirm that the GC–MS method would be able to measure the metabolites in the soluble lysates since it detected the pure substrate and products at the same retention times for benzoic acid (hydrolysed form of benzoyl-CoA), hippuric acid (benzoylglycine) and isovalerylglycine.

The metabolites were then measured by using a general GC–MS system to observe a decrease in the substrate benzoic acid (hydrolysed form of benzoyl-CoA) and an increase in the end-products (hippuric acid), which might qualitatively indicate biologically active GLYAT.

The GC–MS metabolite detection was untargeted, and therefore, additional peaks can be seen in the EICs. These additional peaks were caused by other metabolites produced by the host C41(DE3)pLysS cells during cellular growth and the expression of GLYAT. These metabolites were then present in the soluble lysate before the GC–MS sample preparation step and were, therefore, also detected by GC–MS. An example of the additional peaks is shown in Figure 49.

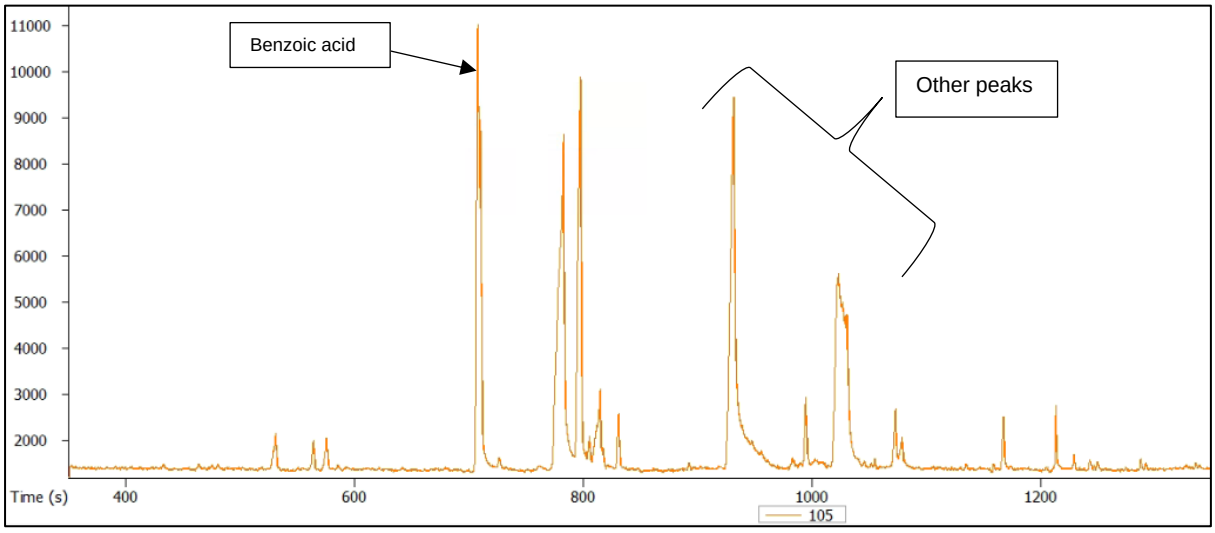


Figure 49: EIC indicating multiple peaks due to metabolites in the cleared lysate.
One of the examples indicating multiple peaks in the reaction with the soluble lysate.

The chromatograms below indicate the reduction in the concentration of benzoic acid (hydrolysed form of benzoyl-CoA) and the increased concentration of hippuric acid in the presence of soluble GLYAT. The soluble lysate of induced C41(DE3)pLysS cells (orange line, Figure 50) was compared to the soluble lysate of C41(DE3)pLysS cells (which did not contain GLYAT) (green line, Figure 50). Benzoic acid (hydrolysed form of benzoyl-CoA) and hippuric acid gave a characteristic 105 m/z ion (fragment), which was used to visualise the respective peaks more clearly.

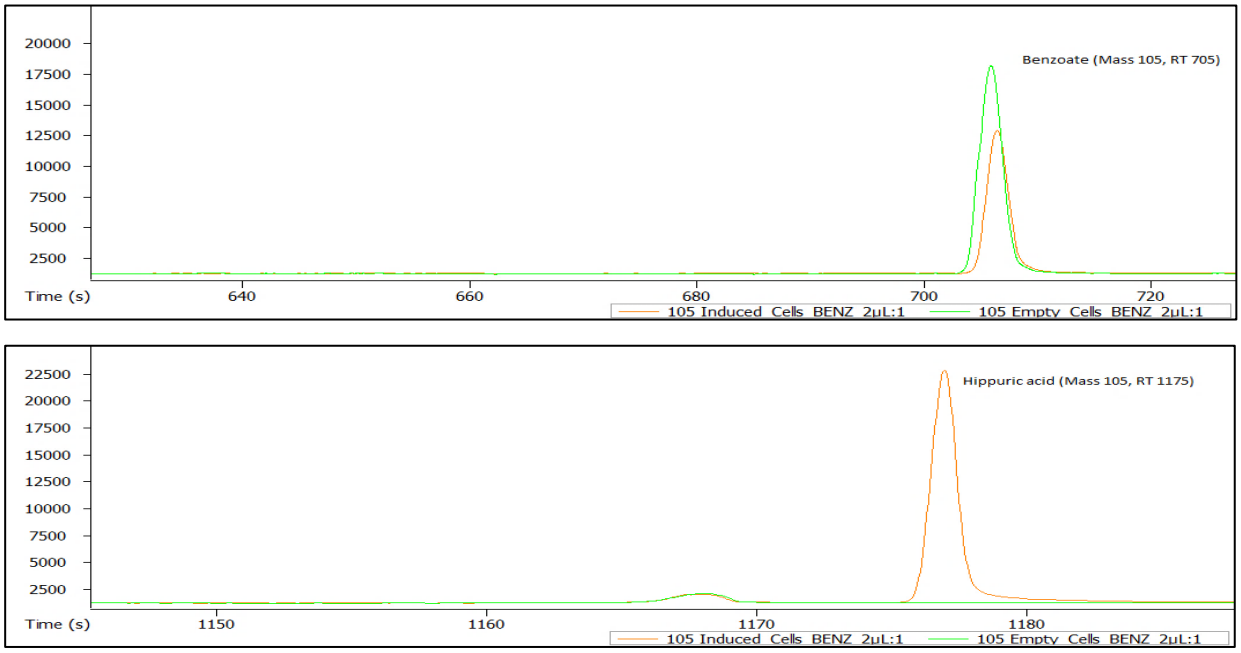


Figure 50: GC-MS results of benzoyl-CoA in the pETDuet-1 expression system.
The green EIC represents the C41(DE3)pLysS lysate without soluble GLYAT (Empty cells). The orange EIC represents the C41(DE3)pLysS lysate with soluble GLYAT. Benzoyl-CoA decreased in this reaction while the hippuric acid increased in this reaction.

A clear decrease in the benzoic acid (hydrolysed form of benzoyl-CoA) concentration can be seen in Figure 50, whereas a clear increase in hippuric acid (benzoylglycine) can be seen. The hippuric acid only formed in the cells that contained GLYAT, which confirms that the C41(DE3)pLysS cells cannot naturally metabolism benzoyl-CoA to hippuric acid. The possible problem with natural benzoate degradative pathways mentioned by Chen *et al.* (2014) and Davidson *et al.* (2005) is thus not applicable to C41(DE3)pLysS cells. The increase in hippurate concentration also indicates that the GLYAT enzyme expressed in the pETDuet-1 expression system is biologically active in the soluble lysate of the cells.

Therefore, C41(DE3)pLysS cells are a good host for expressing biologically active soluble GLYAT. This is indicated by the biotransformation of benzoyl-CoA to hippuric acid in the

C41(DE3)pLysS cells containing soluble GLYAT. This biotransformation further suggests that GLYAT is indeed responsible for eliminating xenobiotic-CoA metabolites (such as benzoyl-CoA) in the second phase of the glycine conjugation pathway and qualitatively confirms the findings of previous studies (Bartlett & Gompertz, 1974; Kelley & Vessey, 1994; Matsuo *et al.*, 2012; Mawal & Qureshi, 1994; Rohwer *et al.*, 2021; Schutte *et al.*, 2019; van der Sluis *et al.*, 2013; van der Sluis *et al.*, 2017; van der Westhuizen *et al.*, 2000). These studies, however, measured the enzyme kinetics with enzyme assays. Due to the high sensitivity of GC–MS to detect very low concentrations of metabolites compared to the inability of enzyme assays to detect very low concentrations of metabolites, the GC–MS method was preferred over enzyme assays for this study.

Because the biotransformation of benzoyl-CoA to hippuric acid can specifically be observed by this GC–MS configuration, the second objective was partially successful. It is only partially successful since ACSM2B and GLYAT could not be analysed simultaneously in the same soluble cell lysate as a biphasic pathway.

4.3.3 GC–MS analyses of expressed GLYAT using the pHis17 expression system

In this section, the decrease in some of the known substrates for GLYAT was measured along with the increase in the respective end-products. These substrates were benzoic acid (hydrolysed form of benzoyl-CoA), isovaleric acid (hydrolysed form of isovaleryl-CoA), and phenylacetic acid (hydrolysed form of phenylacetyl-CoA). The biotransformation of these substrates would then indicate whether GLYAT can conjugate these substrates to glycine. One of the reactions was set up with phenylacetyl-CoA and glutamine to test the conjugation ability of GLYAT when these substrates were present.

The chromatograms below indicate a decrease in benzoic acid (hydrolysed form of benzoyl-CoA) and an increase in hippuric acid in the presence of soluble GLYAT. The soluble lysate of induced C41(DE3)pLysS cells (green line, Figure 51) was compared to the soluble lysate of C41(DE3)pLysS cells (which did not contain GLYAT) (orange line, Figure 51).

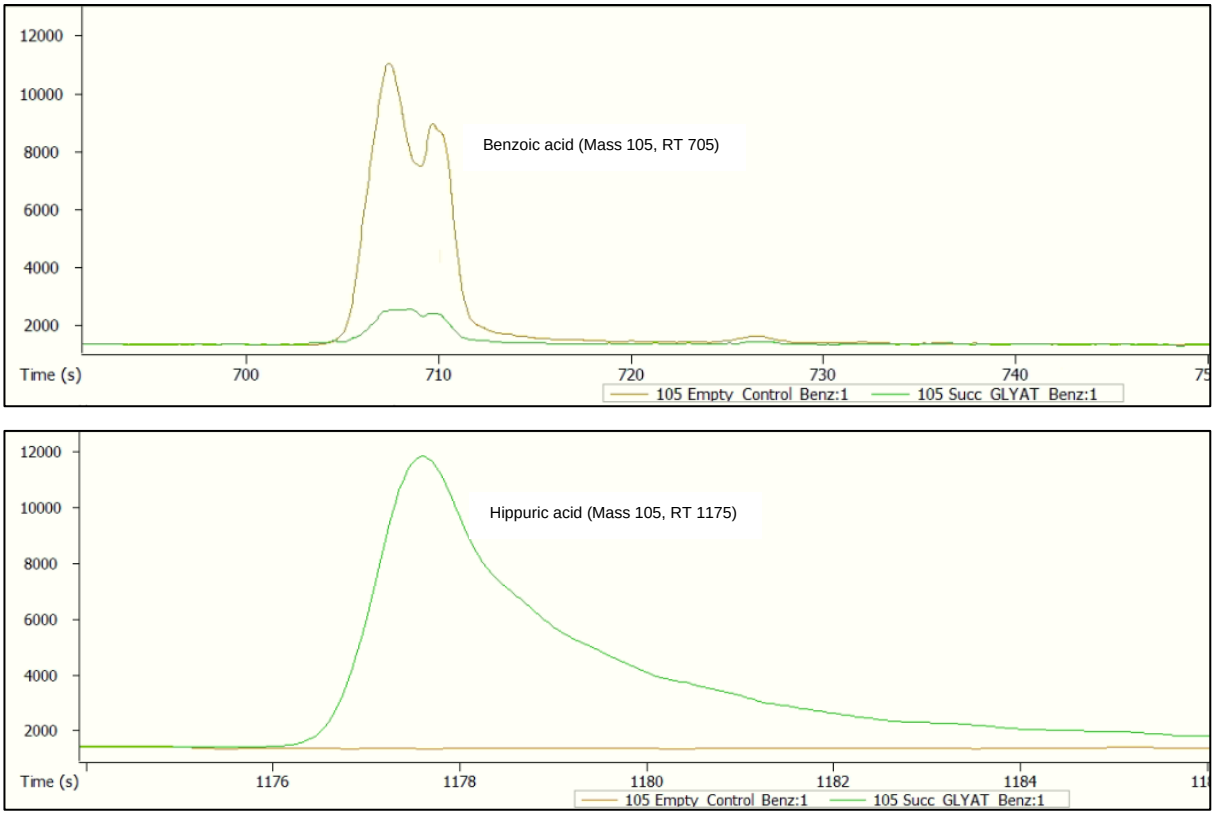


Figure 51: GLYAT and only Benzoyl-CoA added in the pHis17_GLYAT expression system.
The orange EIC represented the induced C41(DE3)pLysS lysate with soluble GLYAT. The green EIC represents the C41(DE3)pLysS lysate without soluble GLYAT (Empty cells). Benzoyl-CoA decreased in this reaction while the hippuric acid increased.

A clear decrease in the benzoic acid (from benzoyl-CoA) concentration can be seen in Figure 51, whereas a clear increase in hippuric acid (benzoylglycine) can be seen. Therefore, C41(DE3)pLysS cells are a good host for expressing biologically active soluble GLYAT using the pHis17 vector. The results are similar to those of the pETDuet-1_ACSM2B/GLYAT system and confirm the presence of biologically active GLYAT.

There are, however, some peak tailing visible that cause lower peak resolution. Peak tailing is caused by the bad placement of the column into the inlet, the cut of the column at the inlet is bad or the liner should be replaced/trimmed at the front of the column (CrawfordScientific, 2022). However, quantification was not performed, and enzyme activity was merely investigated qualitatively. Thus, peak tailing does not have that great of an influence on the results.

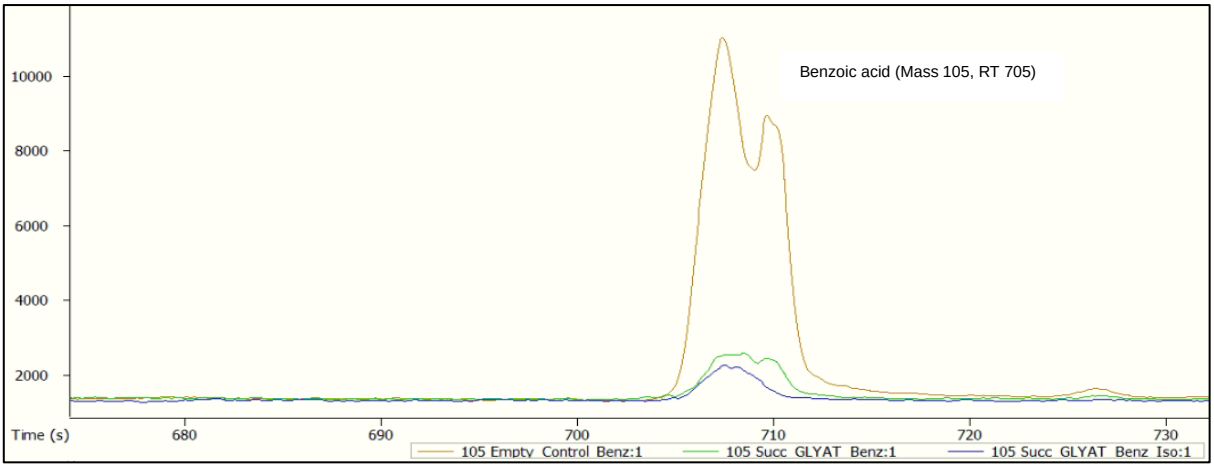
Isovaleric acid (hydrolysed form of isovaleryl-CoA) and isovalerylglycine were not detected by the GC–MS system in the cell lysate containing GLYAT. These metabolites could not even be detected in the cells that did not contain biologically active GLYAT, which suggests that there may be natural catabolism of isovaleryl-CoA in C41(DE3)pLysS cells. Although research on

isovaleryl-CoA dehydrogenases in bacteria is lacking, a study indicated that molecules that have β-methyl groups (such as isovaleric acid) are carboxylated, and then an acetate group is removed from this molecule (Forster-Fromme & Jendrossek, 2008). This explains the inability of GC–MS to detect the change in isovaleric acid (hydrolysed isovaleryl-CoA) since isovaleric acid is already catabolized during the reaction set up using soluble lysate.

Due to the inability to observe a decrease in isovaleric acid (hydrolysed form of isovaleryl-CoA) and an increase in isovalerylglycine, no chromatograms were shown.

Interestingly, it was shown that in the presence of isovaleric acid (in the form of isovaleryl-CoA), a decrease in the biotransformation rate of benzoic acid (in the form of benzoyl-CoA) to hippuric acid (benzoylglycine) was observed. This may indicate an inhibitory/substrate competition effect of isovaleryl-CoA on the glycine conjugation pathway. It was shown in section 1.3.2 (Table 3) that isovaleryl-CoA is not a preferred substrate (K_m – 180 µM (Bartlett & Gompertz, 1974) compared to benzoyl-CoA K_m – 97±3 µM (Rohwer *et al.*, 2021)) for GLYAT and may, therefore, decrease benzoyl-CoA elimination. An inhibitory effect of salicylic acid elimination was already previously shown (Levy & Amsel, 1966). The study of Levy and Amsel (1966), together with this current study, indicates that a type of inhibition is observed between substrates that are conjugated by the same enzyme.

The effect can be seen in Figure 52, where the orange EICs represent the cells that did not contain soluble biologically active GLYAT. The green EIC represents the cells that contained soluble biologically active GLYAT but only had benzoyl-CoA added. Blue EIC represents the cells that also contained soluble biologically active GLYAT, but benzoyl-CoA and isovaleryl-CoA were both added to the reaction.



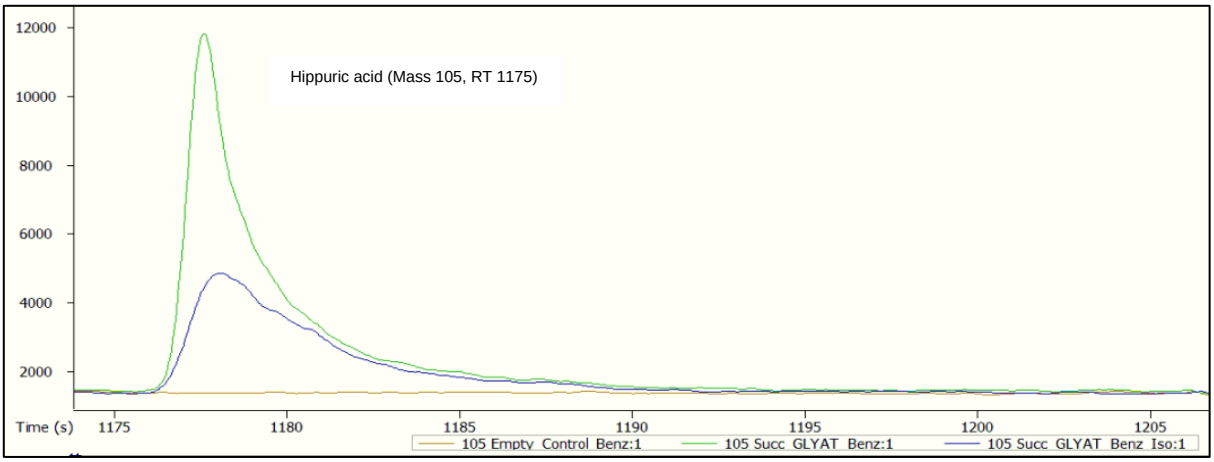


Figure 52: GLYAT with Benzoyl-CoA added and Isovaleryl-CoA added in the pHis17_GLYAT expression system.
The orange EIC represents the C41(DE3)pLysS lysate without soluble GLYAT (Empty cells). The green EIC represents the induced C41(DE3)pLysS lysate with soluble GLYAT, and the blue EIC represents the induced C41(DE3)pLysS lysate with soluble GLYAT, but benzoyl-CoA and isovaleryl-CoA was added to the reaction. This reaction indicated a decrease in benzoyl-CoA in the presence of isovaleryl-CoA, but not as effective as the reaction that only contained benzoyl-CoA.

4.3.4 GC–MS analysis of the biotransformation of phenylacetyl-CoA by GLYAT and GLYATL1 in the soluble lysate

4.3.4.1 Biotransformation of phenylacetyl-CoA in the presence of GLYAT and glycine

In the reactions of the EIC in Figure 53, the decrease in phenylacetic acid (hydrolysed form of phenylacetyl-CoA) as a substrate was determined in the presence of either excess glycine or glutamine. The concentration of phenylacetic acid (hydrolysed form of phenylacetyl-CoA) in the C41(DE3)pLysS clear lysate that did not contain GLYAT was then compared to the reactions that did contain soluble GLYAT. Furthermore, the amino acid added was varied to observe which is more likely to be conjugated to phenylacetyl-CoA.

The orange EIC represents the reaction that did not contain GLYAT, the green EIC reaction contained GLYAT but had glycine added as an amino acid substrate, and the blue EIC represents the second reaction that contained GLYAT but had glutamine as an amino acid substrate. The orange and blue EIC did not indicate a clear decrease, whereas the green EIC indicated a partial decrease in phenylacetic acid (hydrolysed form of phenylacetyl-CoA). The decrease in the green EIC was small.

Due to the small decrease in phenylacetic acid (hydrolysed form of phenylacetyl-CoA), three interpretations were made from this comparative EIC. The first is that glutamine may not be a preferred substrate for reactions with GLYAT and phenylacetyl-CoA. Second, when the preferred amino acid for GLYAT is present, only some phenylacetyl-CoA is conjugated to glycine. Both the first two observations agree with the findings of (Matsuo *et al.*, 2012) that GLYAT is not the main enzyme responsible for the elimination of phenylacetic acid. Third, the phenylacetic acid degradation pathway is not active in C41(DE3)pLysS cells since no decrease in phenylacetic acid (hydrolysed form of phenylacetyl-CoA) could be observed in the cells that did not contain GLYAT. The phenylacetic acid degradation pathway used by bacteria was described in the study of Teufel *et al.* (2010).

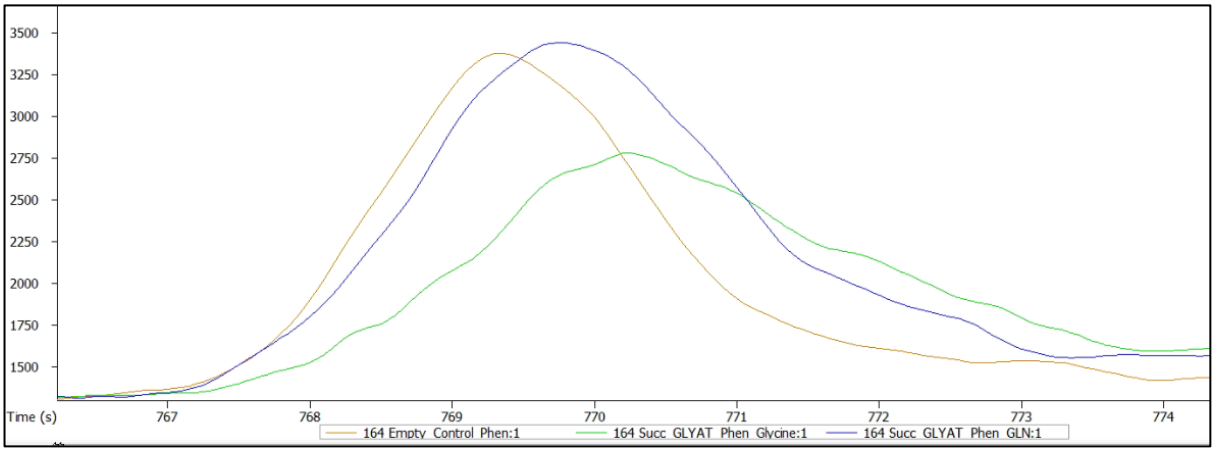


Figure 53: Phenylacetyl-Co biotransformation with GLYAT and glycine or glutamine.
The orange EIC represents the C41(DE3)pLysS lysate without soluble GLYAT (Empty cells). The green EIC represents the C41(DE3)pLysS lysate with soluble GLYAT in the presence of phenylacetyl-CoA and glycine as substrates. The blue EIC represents the C41(DE3)pLysS lysate with soluble GLYAT, and phenylacetyl-CoA and glutamine were added as substrates. The cells that had glycine as substrates indicated some conversion of the phenylacetyl-CoA, whereas those with glutamine as substrate did not indicate any conversion of phenylacetyl-CoA.

4.3.4.2 Biotransformation of phenylacetyl-CoA in the presence of GLYATL1 and glutamine

The reactions were set up in the presence of GLYATL1 with either excess amounts of glycine or glutamine. Phenylacetyl-CoA was also added as a substrate in these reactions. Another

reaction was also set up. This was the C41(DE3)pLysS cell lysate reaction that did not contain GLYATL1. The decrease, if any, was then measured with the GC–MS system.

The decrease in phenylacetic acid (hydrolysed form of phenylacetyl-CoA) was observed only in the reactions containing GLYATL1 and can be seen in Figure 54. The orange EIC represents the reaction that did not contain GLYATL1, whereas the green EIC represents the reaction with GLYATL1 and glutamine, and the blue EIC represents the reaction that contained GLYATL1 and glycine.

Three main observations were made by comparing these EICs in Figure 54. The first is that phenylacetyl-CoA in the presence of GLYATL1 is possibly conjugated to either glutamine or glycine. Second, there may be a difference in substrate preference for GLYAT and GLYATL1 when phenylacetyl-CoA is considered since GLYAT could not conjugate phenylacetyl-CoA as well as GLYATL1. This observation further supports the findings of Matsuo *et al.* (2012) that phenylacetyl-CoA is indeed a better substrate for GLYATL1. Finally, even in the presence of benzoyl-CoA, GLYATL1 was able to conjugate phenylacetyl-CoA to either amino acid. This phenomenon can be seen in the blue EIC, which contained phenylacetyl-CoA and benzoyl-CoA in the same reaction.

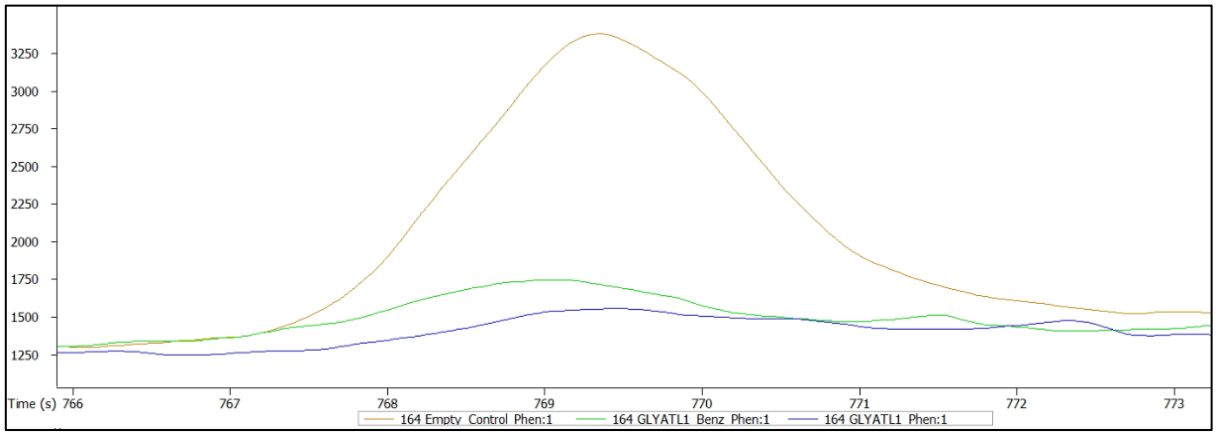


Figure 54: Phenylacetyl-CoA in GLYATL1 with glutamine and presence of benzoyl-CoA.
The orange EIC represents the C41(DE3)pLysS lysate without soluble GLYATL1 (Empty cells). The green EIC represents the C41(DE3)pLysS lysate with soluble GLYATL1, and Phenylacetyl-CoA and glutamine were added as substrates. The blue EIC represents the C41(DE3)pLysS with soluble GLYATL1, and phenylacetyl-CoA, benzoyl-CoA, and glutamine were added as substrates. These reactions indicated that phenylacetyl-CoA was converted in the induced C41(DE3)pLysS lysate but not in the cells that did not contain soluble GLYATL1.

Another focus was to observe whether benzoyl-CoA could be biotransformed to hippuric acid by GLYATL1. After GC–MS analysis was performed, the EIC of the reaction containing benzoyl-CoA and GLYATL1 was compared to reactions containing neither GLYAT nor GLYATL1 and the reaction that contained GLYAT and benzoyl-CoA and glycine, which can be seen in Figure 55.

A difference could be seen in the reaction containing phenylacetyl-CoA and benzoyl-CoA. This difference is more directed towards the biotransformation of benzoyl-CoA by GLYATL1. The orange EIC was the reaction that did not contain GLYAT or GLYATL1, the green EIC was the reaction that contained GLYAT in the presence of benzoyl-CoA and glycine, and the blue EIC was the reaction that contained GLYATL1, benzoyl-CoA and glutamine. The reaction did not biotransform benzoyl-CoA to hippuric acid since there was no hippuric acid formed in the reaction with benzoyl-CoA and GLYATL1. This phenomenon can be related to the very low activity of GLYATL1 to benzoyl-CoA in the presence of glutamine in the study by Matsuo *et al.* (2012).

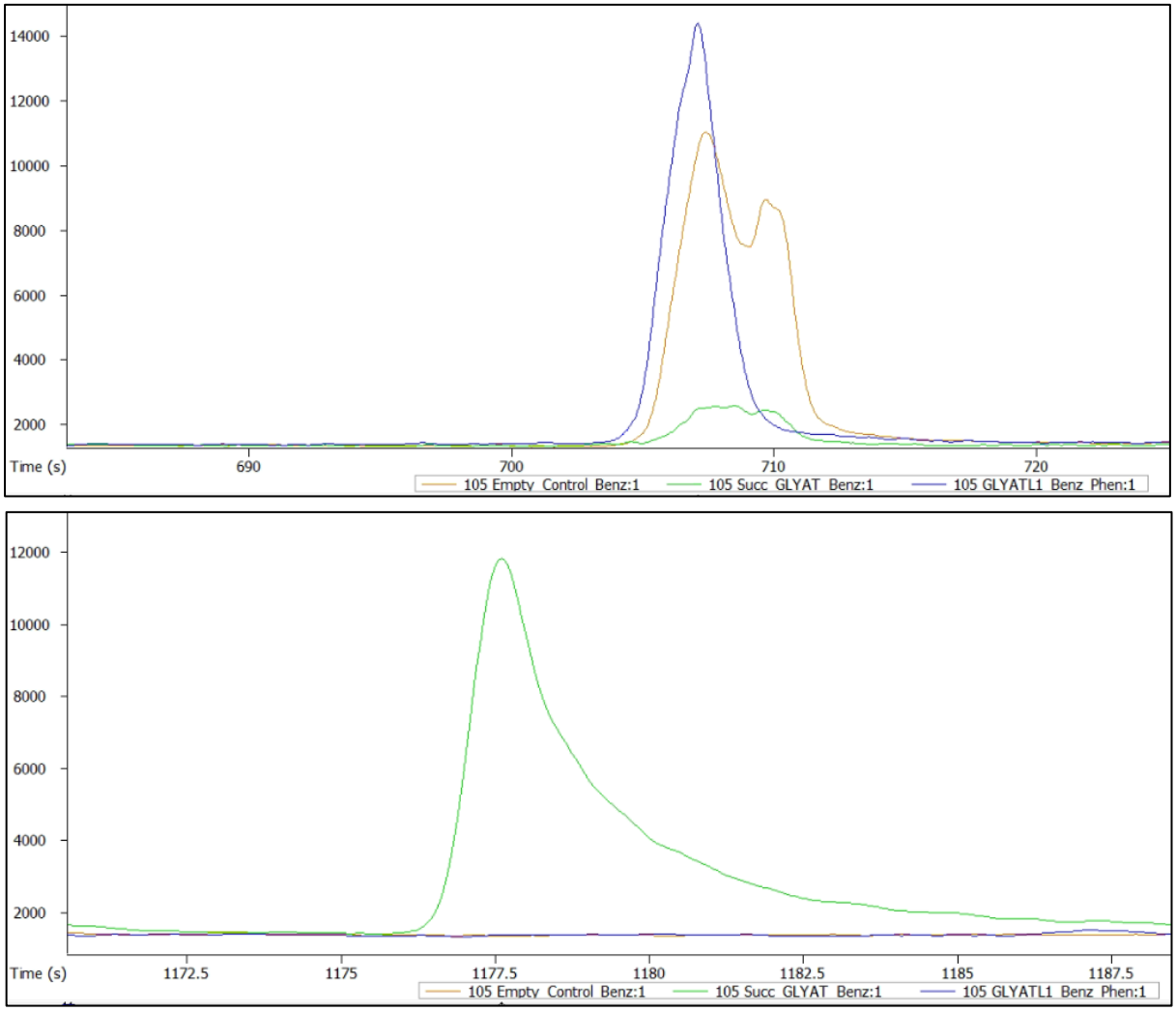


Figure 55:No hippuric acid formation in lysate with GLYATL1.
Then orange EIC represents the C41(DE3)pLysS lysate without GLYATL1 (Empty cells). The green EIC represents the induced C41(DE3)pLysS lysate with soluble GLYAT. The blue EIC represents the induced C41(DE3)pLysS lysate with soluble GLYATL1. Benzoyl-CoA decreased while hippuric acid increased in the lysate that contained GLYAT, but not in the lysate that contained GLYATL1.

The expected amino acid conjugates for phenylacetyl-CoA were phenylacetylglycine and phenylacetylglutamine by GLYAT and GLYATL1, respectively. Neither phenylacetylglycine nor phenylacetylglutamine was detected by GC–MS in this study. The only studies found related to phenylacetylglutamine/phenylacetylglycine currently were based on the localisation and effects of these metabolites on the human body. I could not, to my knowledge, find studies related to phenylacetylglutamine/phenylacetylglycine catabolism through bacterial enzymes.

To establish an enzyme model for the glycine conjugation pathway, the amino acid substrates (glycine and glutamine) were measured with the established GC–MS method. For free amino acids, such as glycine and glutamine to be analysed, an extraction step is usually included (Azevedo *et al.*, 2017; Bale *et al.*, 2020; Perez-Palacios *et al.*, 2015) before GC–MS analysis. However, in this study, the focus was to measure the metabolites directly in the enzyme reactions that were set up with the soluble lysate. Therefore, no amino acid extraction was performed.

It can be observed in Figure 56 that the glycine concentration could not be clearly distinguished between the reactions to be compared. The soluble lysate that contained GLYAT (in the presence of glycine and phenylacetyl-CoA) is represented by the green line in Figure 56, whereas the reactions that contained GLYAT (in the presence of glutamine and phenylacetyl-CoA) are represented by the orange line in Figure 56. The amino acid EICs were, therefore, not used as comparative results for the enzyme reactions set up in the soluble lysate of the cells that contained either GLYAT or GLYATL1.

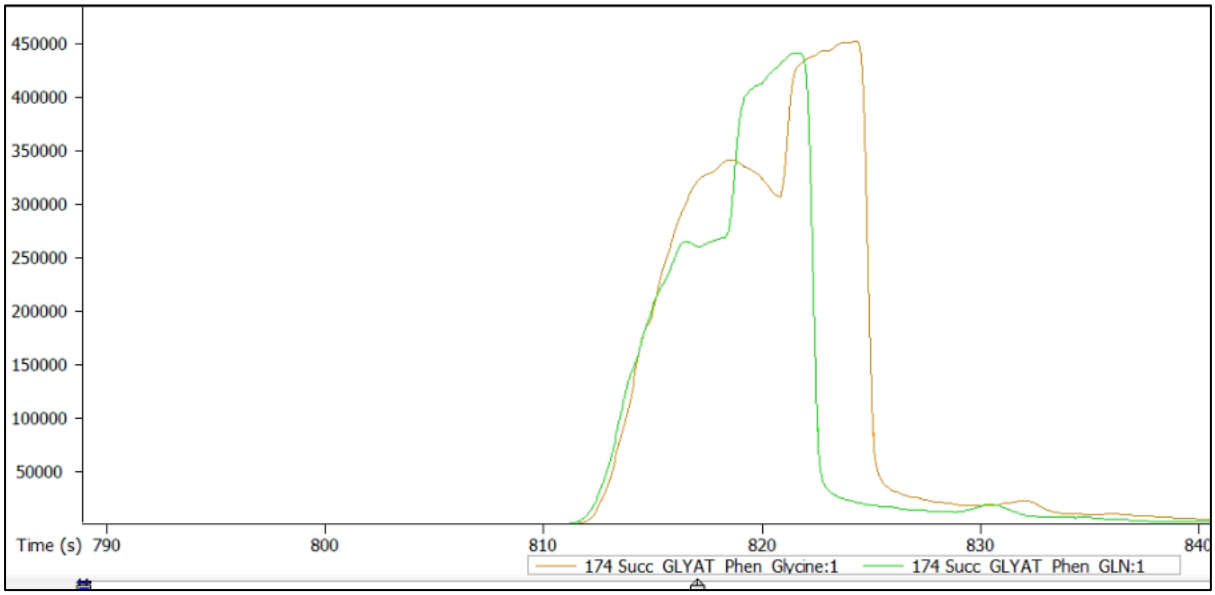


Figure 56: EIC for glycine concentrations in reactions with phenylacetyl-CoA as substrate.

Because these chromatograms indicated promising results, the next step would be to run these reactions with the concentrated and quantified GLYAT and GLYATL1 that were stored in the Tris-HCl (pH 8.0) buffer and observe if similar results are obtained.

4.4 The results and discussion of the GC–MS analysis of enzyme reactions containing purified, concentrated GLYAT or GLYATL1

4.4.1 GC–MS analysis of purified, concentrated GLYAT and GLYATL1 in Tris-HCl (pH 8.0)

Because the purified concentrated GLYAT and GLYATL1 were stored in a high concentration of Tris-HCl (pH 8.0), the matrix effect had to be evaluated. Furthermore, the effectiveness of Tris-HCl (pH 8.0) as a reaction buffer for subsequent reactions on the GC–MS was determined. This was done to observe if changes seen in the EIC of the concentrated GLYAT and GLYATL1 can be compared to the successful EIC of the cleared lysate reactions.

The analysis was performed by preparing a pseudoreaction that only contained Tris-HCl (pH 8.0) and glycine and another reaction with water and glycine. The samples were prepared as discussed in section 4.2.3.1. The samples were then analysed on the GC–MS with the same configuration as the reactions of the cleared lysates as discussed in 4.2.2.5.

In Figure 57, green EIC represents the first setup reaction (with Tris-HCl and glycine), and orange EIC represents the second reaction (with water and glycine) described in section 4.2.3.1. The analysis of Tris-HCl (pH 8.0) was important since downstream applications could

be performed to confirm the observations of the previously described chromatograms with quantified and concentrated GLYAT and GLYATL1.

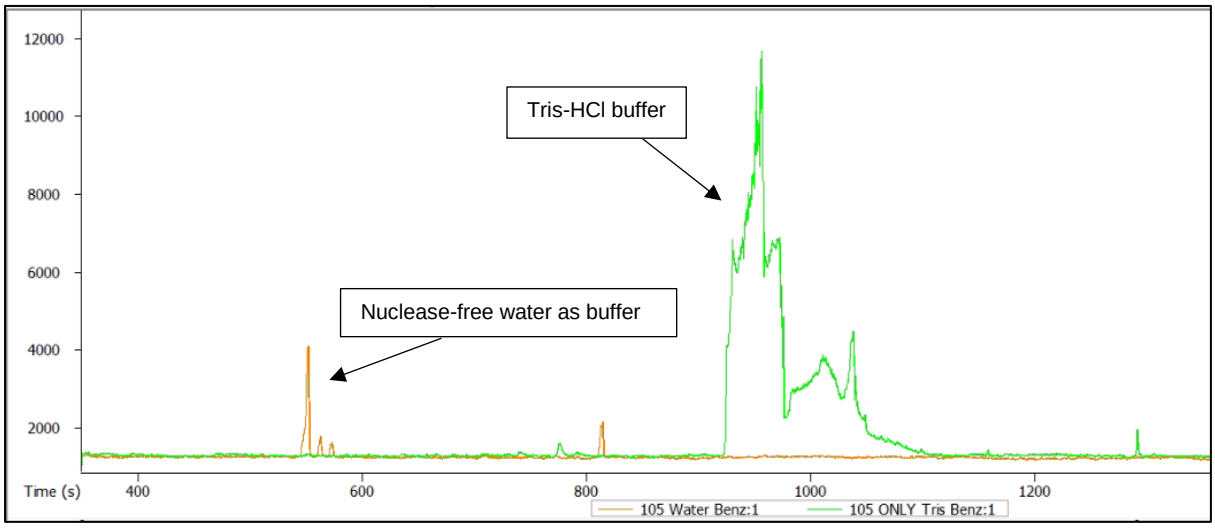


Figure 57: Analysis of Tris-HCl (pH 8.0) as good reaction and storage buffer.
The orange EIC represents the reactions without Tris-HCl (pH 8.0), whereas the green EIC represents the reactions with only Tris-HCl (pH 8.0).

Tris-HCl (pH 8.0) is not a good storage and reaction buffer to be used for subsequent GC–MS analysis at high concentrations. This might be explained with regard to the drying step in sample preparation section 4.2.2.4. The Tris-HCl solution was concentrated 8 x-fold during the drying and derivatisation steps. This might then possibly explain the peak distortions seen in Figure 57. The peak distortion was unique to Tris-HCl (pH 8.0) when the concentrated proteins were analysed. To my knowledge, no studies could be found that used Tris-HCl (pH 8.0) as a reaction buffer for GC–MS analysis at such a high concentration and therefore recommended that a much lower concentration of Tris-HCl (pH 8.0) be used for sample storage and reaction setup.

However, water can possibly still be used as a reaction and storage buffer since the resolution of the peaks was better. Therefore, it is suggested that the buffer be rather changed to nuclease-free ddH_2O during the concentration steps of section 3.2.2.6 or another well suited protein storage or reaction buffer.

4.4 Section summary

The main goal of this chapter was to qualitatively observe if there are any changes in the substrate decrease (benzoyl-CoA, phenylacetyl-CoA and isovaleryl-CoA) and product increase (hippuric acid, benzoylglutamine, isovalerylglycine, isovalerylglutamine, phenylacetylglycine, and phenylacetylglutamine) while either GLYAT or GLYATL1 was present in the enzyme reactions. First, a pure substrate and some products were run with the established GC–MS method to determine the retention times and the suitability of the GC–MS

method to detect the pure substrates and the product. The substrate that was used was benzoyl-CoA, and the products were hippuric acid (benzoylglycine) and isovalerylglycine. Thereafter, negative controls of the soluble lysate were run on the GC–MS to determine if the matrix influenced the GC–MS analysis. Reactions were set up and run on the GC–MS to determine substrate decrease and product increase.

The GC–MS analysis was performed by running the samples on a general GC–MS configuration and analysing the resulting chromatograms to observe the respective substrate-to-product changes. The biotransformation of benzoate (hydrolysed form of benzoyl-CoA) proved to be the most successful, whereas the biotransformation of phenylacetate (hydrolysed form of phenylacetyl-CoA) did display some success. No changes in isovaleric acid (hydrolysed form of isovaleryl-CoA) could be measured. The isovaleryl-CoA reactions, therefore, need further investigation.

The results in this chapter, although qualitative, first agree with the studies of Kelley and Vessey (1994); Nandi *et al.* (1979); Webster *et al.* (1976) confirming that benzoyl-CoA and phenylacetyl-CoA were conjugated by two separate N-acyltransferases. Furthermore, it also agrees with Matsuo *et al.* (2012) that these two acyltransferases are possibly GLYATL1 that may prefer phenylacetyl-CoA conjugation with glutamine, and GLYAT may prefer benzoyl-CoA conjugation with glycine. A summary of the results obtained from the EIC analysis can be seen in Table 29.

Severe peak distortions could be observed when the matrix of the purified, concentrated GLYAT and GLYATL1 was analysed and could not be used for specific metabolite analysis. It was also determined that the chosen Tris-HCl (pH 8.0) is not a good storage or a good reaction buffer at high concentrations for use with a GC–MS system, and the exchange buffer should be changed to 25 mM.

Table 29: Summary of the results obtained from GC–MS analysis of GLYAT and GLYATL1.

Enzyme	Substrate	Amino acid	Worked or not	Substrate depletion	Product detected	Inhibition	Expression system
GLYAT	Benzoyl-CoA	Glycine	Yes	Yes	Hippurate	ND	pETDuet
GLYAT	Benzoyl-CoA	Glycine	Yes	Yes	Hippurate	ND	pHis17
GLYAT	Benzoyl-CoA/Isovaleryl-CoA	Glycine	Yes	Yes	Hippurate	Yes (Benzoyl-CoA elimination)	
GLYAT	Phenylacetyl-CoA	Glycine	Partially	Partial	ND	ND	
GLYAT	Phenylacetyl-CoA	Glutamine	No	No	ND	ND	
GLYATL1	Phenylacetyl-CoA	Glutamine	Yes	Yes	ND	ND	pET23a(+)
GLYATL1	Phenylacetyl-CoA/Benzoyl-CoA	Glutamine	No (Benzoyl-CoA)	No (Benzoyl-CoA)	ND	No	
GLYATL1	Phenylacetyl-CoA	Glycine	Yes	Yes	ND	ND	

* ND – Not detected

Chapter 5: Conclusion and future prospects

5.1 Introduction

The glycine conjugation pathway responsible for detoxifying xenobiotics is gradually gaining the interest it deserves since the daily consumption rate of xenobiotics is on an increasing trajectory.

An important aspect of the glycine conjugation pathway is, therefore, the two-phased elimination process of the increasing xenobiotic concentration within the human body. As discussed in sections 1.1-1.3.2, the two enzymes involved in the biphasic glycine conjugation pathway are ACSM2B (Phase I) and GLYAT (Phase II).

These two enzymes are responsible for activating the xenobiotic to form the xenobiotic-CoA molecule and then for conjugating glycine to the activated xenobiotic-CoA molecule. This conjugation then increases the hydrophilicity of the xenobiotic, and the end product is excreted in the urine. For example, a confirmed substrate for the glycine conjugation pathway, benzoate (sodium benzoate used as a preservative), is activated to form benzoyl-CoA in the presence of ATP and CoA as cofactors. Thereafter, glycine is conjugated to the benzoyl-CoA molecule and forms benzoylglycine (more commonly known as hippurate). The CoA molecule is released again in the second phase.

It has already been indicated in multiple animal studies that if the xenobiotic concentration increases either by ingestion or environmental exposure, dire consequences will follow (Nair, 2001; Noorafshan *et al.*, 2014; Oyewole *et al.*, 2012; Srinithi & Bhargava, 2018; Swenson, s.a.; Tsay *et al.*, 2007). If this was to be extrapolated to humans, the future does not seem too clear for mankind and even more so when there are variants of either *ACSM2B* or *GLYAT* that may decrease the enzyme activity. One of these variants has already been identified by (van der Sluis *et al.*, 2013), and the increased ingestion of xenobiotics has already been linked to some adverse effects in humans. For example, ADHD in children (Beezhold *et al.*, 2014; Kilic *et al.*, 2006).

Other substrates that may be involved in the glycine conjugation pathway should not be neglected. These alternative substrate entry points are naturally occurring OA-CoA metabolites in the human body. One of these OA-CoA metabolites is isovaleryl-CoA when an individual suffers from isovaleric acidaemia, which is caused by a defective isovaleryl-CoA dehydrogenase (IVD) enzyme in the leucine catabolic pathway. This then leads to the accumulation of isovaleric acid and isovaleryl-CoA (Krieger & Tanaka, 1976; Tanaka *et al.*, 1966). Previous studies identified isovalerylglycine and isovalerylglutamine in the urine of patients with IVA (Ensenauer *et al.*, 2004; Wysocki *et al.*, 1983), which may suggest the involvement of the glycine conjugation pathway in alleviating the bioaccumulation of isovaleric

acid and isovaleryl-CoA. GLYAT forms part of the second phase of the glycine conjugation pathway discussed in section 1.3. GLYATL1 is expressed in the kidneys (Zhang *et al.*, 2007). It is therefore important to verify the possible involvement of GLYAT and GLYATL1 in isovaleryl-CoA conjugation and GLYATL1 in the glycine conjugation pathway.

Therefore, it is critical to study the effect of genetic variation on the biphasic glycine conjugation pathway and to establish an enzymatic model that accurately represents the detoxification steps.

In pursuit of a better understanding of the effect of genetic variation on the glycine conjugation pathway and to aid in establishing an accurate enzymatic model for this pathway, this study aimed to establish a model system in which variants of *ACSM2B*, *GLYAT* and *GLYATL1* can be expressed to determine the xenobiotic biotransformation capabilities of the two-phased glycine conjugation pathway and its involvement, if any, in the detoxification of isovaleryl-CoA.

5.2 Conclusion

Coexpression of biologically active ACSM2B and GLYAT in an established model

ACSM2B and *GLYAT* were successfully cloned into the pETDuet-1 vector. *ACSM2B* was cloned in line with an N-terminal His-tag, and *GLYAT* was cloned in line with an N-terminal S-tag. The pETDuet-1_ACSM2B/GLYAT construct was then successfully transformed into the OverExpress™ C41(DE3)pLysS host. The C41(DE3)pLysS cells did not express any soluble ACSM2B. Therefore, the C41(DE3)pLysS host was not considered a good expression host for the expression of soluble ACSM2B. Another attempt was made to express ACSM2B but was not successful.

Soluble GLYAT was, however, overexpressed in the C41(DE3)pLysS host. Therefore, the C41(DE3)pLysS host was considered a successful host for expressing GLYAT. Due to the failure to express soluble ACSM2B but the success of overexpressing soluble GLYAT, the first objective was only partially reached.

Identification and optimisation of an expression system to express GLYAT and GLYATL1

Since GLYAT was successfully overexpressed in C41(DE3)pLysS cells, further focus was placed on the expression of GLYAT and one of its paralogues, GLYATL1, in C41(DE3)pLysS cells. Soluble GLYAT and GLYATL1 were both successfully expressed with a C-terminal His-tag in C41(DE3)pLysS cells. The expression conditions were then further optimised to have maximal protein yield by performing an IPTG and time gradient. Soluble GLYAT and GLYATL1 were recovered from the cleared cell lysate using nickel affinity chromatography. Both GLYAT and GLYATL1 were successfully concentrated, stored, and quantified in Tris-HCl (pH 8.0). A

total of 36.4 µg/mL soluble GLYAT and 49.9 µg/mL could be retrieved from a 100 mL cell culture in this expression system, even when considering the loss during protein purification. This objective was, therefore, reached successfully.

Evaluating the biotransformation of benzoate using the partially established pETDuet-1_ACSM2B/GLYAT system and GC–MS

To attempt to determine the biotransformation of benzoate, reactions were set up with the cleared lysate of the induced C41(DE3)pLysS cells and the cells that did not contain soluble GLYAT. It was observed early in the experimental phase that the lysis buffer interferes with the GC–MS sample preparation during cell lysis. Therefore, it was also determined that the vibration mill lysis technique proved to be a more efficient lysis technique for downstream GC–MS analysis.

Because only soluble GLYAT was expressed in the pETDuet-1_ACSM2B/GLYAT system, benzoate biotransformation was measured with benzoyl-CoA as a xenobiotic substrate and glycine as an amino acid substrate. There was a clear conversion reaction visible on the EIC in terms of decreasing benzoyl-CoA, while the hippuric acid consequently increased. This phenomenon was only observed in the cell lysates containing soluble GLYAT and, therefore, may suggest that the expressed GLYAT in the pETDuet-1_ACSM2B/GLYAT system was biologically active.

Evaluating the biotransformation of benzoate using the successfully established and optimised pHis17_GLYAT and pET23a(+)_GLYATL1 systems and GC–MS

For the same reason stated previously, xenobiotic-CoA (benzoyl-CoA and phenylacetyl-CoA) and OA-CoA (isovaleryl-CoA) were used in the reaction setup for the GC–MS system. Benzoyl-CoA conversion to hippuric acid was observed to be the most successful. Therefore, it confirmed that the soluble GLYAT expressed in this system was indeed biologically active. The conversion rate of benzoyl-CoA to hippuric acid was, however, immensely affected when the reaction setup contained isovaleryl-CoA. This decelerated conversion suggested that isovaleryl-CoA may exert an inhibitory effect on the second phase of the glycine conjugation pathway. Unfortunately, the conversion of isovaleryl-CoA could not be determined by GC–MS.

Only the substrate decrease of phenylacetyl-CoA could be determined by GC–MS. This, nonetheless, gave some insightful results. First, a clear preference for this xenobiotic substrate could be seen between soluble biologically active GLYAT and soluble GLYATL1. When phenylacetyl-CoA was added to the reactions containing soluble biologically active GLYAT, little to no conversion was observed. However, when phenylacetyl-CoA is added to the reactions containing soluble GLYATL1, the phenylacetyl-CoA substrate decreases

immensely. Thus, phenylacetyl-CoA is not a preferred substrate for GLYAT but is indeed for GLYATL1.

Another, less clear observation was the preference of amino acid substrates of soluble biologically active GLYAT and soluble GLYATL1. The difference further suggested that glycine is preferred by GLYAT and that glutamine is preferred by GLYATL1. This conclusion was made due to the observation of some phenylacetyl-CoA being converted when GLYAT's preferred amino acid substrate was present and when almost all the phenylacetyl-CoA was converted in the presence of GLYATL1 and glutamine.

Since phenylacetylglycine and phenylacetylglutamine could not be detected, the problem was to be sure that soluble GLYAT and GLYATL1 were responsible for the conversion reactions. This problem was addressed by setting up a reaction with the cleared lysate of the C41(DE3)pLysS cell that did not contain either recombinant enzyme. The results, therefore, support the fact that GLYAT and GLYATL1 were responsible for the conversion of phenylacetyl-CoA since no decrease was determined in the reactions that did not contain either GLYAT or GLYATL1 and, therefore, suggested that GLYATL1 was also expressed in its biologically active form.

When both the results of GLYAT and GLYATL1 conversion reactions are taken into consideration, it suggests that GLYATL1 may not detoxify the same type of xenobiotics that GLYAT does since the structures of benzoyl-CoA and phenylacetyl-CoA are different by only an acetyl group. This may further suggest and confirm the existence of multiple detoxification enzymes that function to detoxify different types of molecules from the body since both are expressed in the kidneys.

At the end of the study, the capacity of Tris-HCl (pH 8.0) to be a storage and reaction buffer was tested. However, it was determined that Tris-HCl (pH 8.0) should not be used as a storage or reaction buffer for either protein at this high concentration.

5.3 Future prospects

Genetic variation is important when studying enzyme activity since it can lead to the expression of defective enzymes with lowered or even no enzyme activity. Although the occurrence is rare, the glycine conjugation pathway is not an exception to the influence of genetic variation since a gene variant of *GLYAT* has already been found, amongst many, that exerts low enzyme activity (van der Sluis *et al.*, 2013). Therefore, it remains crucial to establish an expression model that can coexpress ACSM2B and GLYAT and determine the influence of genetic variation on the detoxification rate of the glycine conjugation pathway.

Furthermore, since paralogues of *GLYAT exist,* e.g., *GLYATL1* and *GLYATL2,* it will be insightful to first determine their role in overall metabolism and whether they are related to the glycine conjugation pathway. Second, if it has a role in the glycine conjugation pathway, what the specific role of these paralogues are and whether the expression of their respective enzymes contributes to the alleviation of the pressure placed on the glycine conjugation pathway. It might also be sensible to determine the amino acid preference of the GLYATL1 and GLYATL2 enzymes to observe whether other amino acids are preferred other than glycine and glutamine. These studies will decrease the immense void currently present in the research and knowledge of the glycine conjugation pathway. Furthermore, more attention and troubleshooting can be given to express ACSM2B within this host by following the suggestions of Saida *et al.* (2006). If none of the amendments yield any results, it may be better to work with unicellular eukaryotic expression models to establish a coexpression system (Ahmad *et al.*, 2014; Owczarek *et al.*, 2019; Puetz & Wurm, 2019; Tripathi & Shrivastava, 2019; Yin *et al.*, 2007).

This study did not focus on quantification but on observing whether the biotransformation could be determined using a GC–MS system, it would be suggested that a quantification approach be followed next. This will, therefore, yield more specific results in decreasing substrates and increasing products, which will give a better understanding of the influence of genetic variations when the quantifiable results are then compared between variants. This prospect is much like the study of van der Sluis *et al.* (2017); however, the focus should be placed on quantifying different substrates and increased products.

A problem specific to this study to be addressed in the near future could be the establishment of a protein quantification method that measures one specific protein in a mixture of proteins. This method could then be used to determine the enzyme concentration (e.g., GLYAT) in the cleared lysate of the sample. This will then shorten the research output time immensely by avoiding the dependability on protein purification and concentration for protein quantification. This could be done by establishing a standard protein with a fluorescent tag attached to it. A standard curve can then be established for proteins with that specific fluorescent tag. Thus, if proteins are expressed with that fluorescent tag, they could be compared to the standard curve, and the protein concentration in the mixed sample will be known. This, however, is just a simple explanation of such a procedure, as it will take much research to establish such a method. One such assay that might be useful to detect GLYAT and GLYATL1 in the cleared supernatant may be ELISA (enzyme-linked immunosorbent assay). This assay can detect specific peptides in solution and can be quantifiable (Crowther, 2009).

Since isovaleryl-CoA substrates and products could not be measured in the soluble lysate of the cells containing GLYAT or GLYATL1 by this system, it is important to find a system in which the effects of isovaleric acid and isovaleryl-CoA on the glycine conjugation pathway can be measured since it was already suggested by this study that the presence of isovaleryl-CoA has an inhibitory effect on benzoyl-CoA elimination. This will especially contribute to evaluating the effects of isovaleric acidemia on the glycine conjugation pathway. A study such as this could also determine whether glycine treatment for isovaleric acidemia patients is truly the best treatment or if it adds to the patient's symptoms.

It is also suggested that further research be put into finding an appropriate storage and reaction buffer other than or a lower concentration of Tris-HCl (pH 8.0). At high concentrations, the buffer caused peak distortions; therefore, lowering it to the buffer concentration used in previous studies could pose a solution to this problem. This will then allow for the samples to be analysed on the established GC–MS system if the study requires the proteins to be purified and concentrated. Finding an appropriate storage and reaction buffer for purified GLYAT and GLYATL1 could also contribute to the measurement of isovaleryl-CoA elimination since it eliminates the carboxylation reaction of isovaleryl-CoA (Forster-Fromme & Jendrossek, 2008). The substrates (isovaleric acid, glycine and glutamine) and the products (isovalerylglycine or isovalerylglutamine) would elucidate which enzyme (GLYAT or GLYATL1) is involved in isovaleryl-CoA elimination and elucidate which amino acid is preferred to be conjugated to isovaleryl-CoA.

5.4 Study limitations

A great limitation of this study is the restriction to the biotransformation of GLYAT and GLYATL1 substrates only due to the inability to express soluble ACSM2B. Another limitation is the restriction on working with bacterial cells. Both of these limitations might be addressed with the same solution. It is, therefore, suggested that this study continues in mammalian cells that have various advantages when expressing a eukaryotic enzyme.

Moreover, another limitation of this study was that the Tris-HCl (pH 8.0) buffer was at a very high concentration. This then caused the inability to measure the conversion reactions on the GC–MS of the concentrated GLYAT and GLYATL1.

5.5 Study output

The pETDuet-1_ACSM2B/GLYAT system was presented as a poster at the Online 2022 SASBMB conference.

Appendices

Appendix A – Reagents, equipment and software used in this study

Table 30: List of chemicals, reagents and equipment needed for the expression protocols.

Chemicals/Reagents/Kits	Manufacturer
1% BSTFA	Sigma–Aldrich
1% Caesine in 1xTBS (Tris-buffered saline) blocking buffer	Bio-Rad
10 X Buffer Tango	Fermentas
2 x Laemmli buffer	Bio-Rad
20 x TBS Tween 20 Buffer	Thermo Scientific
30% Acrylamide mix	Sigma–Aldrich
AEC solution	Sigma–Aldrich
Agarose	Melford
Ammonium Persulfate	Sigma–Aldrich
Benzoyl-CoA	Sigma–Aldrich
Buffer 2.1	New England Biolabs® Inc.
Carbenicillin	Glentham Life Sciences
Chloramphenicol	Sigma–Aldrich
CloneAmp™ HiFi PCR Premix	TakaraBio
Coomassie Brilliant Blue R-250	Merck
Cutsmart® buffer	New England Biolabs® Inc.
DH5α cells	IOL
DMSO	Merck
EDTA	Fluka
Ethanol (Absolute)	Sigma–Aldrich
Ethidium Bromide	Sigma–Aldrich
Gel loading Dye Purple 6x	New England Biolabs
GenElute™ Plasmid Midiprep Kit	Sigma–Aldrich
Glacial acetic acid	Merck
Glucose	Melford
Glutamine	Sigma–Aldrich
Glycerol	Sigma–Aldrich
Glycine	Melford
HRP-labelled Monoclonal Mouse Anti-His-tag antibody	R&D systems®

HRP-labelled Polyclonal Goat Anti-S-tag antibody	Novus™Bio
Hydrogen peroxide	Clicks
InFusion® HD EcoDry™ Cloning kit	TakaraBio
IPTG	Melford
Isopropanol	Merck
Isovaleryl-CoA	Sigma–Aldrich
JM109 cells	IOL
Liquid nitrogen	IOL
Luria Bertani media and agar	Melford
Lysonase™ Bioprocessing Reagent	Novagen®
Methanol	Merck
MOX	Sigma–Aldrich
Nitrogen gas	IOL
Nuclease-free water	GE Healthcare Life Sciences
NucleoSpin® Gel and PCR clean-up kit	Macherey-Nagel
NucleoSpin® Plasmid Purification kit	Macherey-Nagel
O'GeneRuler™ 1 kb DNA ladder, ready-to-use	Thermo Scientific™
OverExpress®C41(DE3)pLysS chemically competent cells	Lucigen
Pantothenic acid	Sigma–Aldrich
PEG	Merck
pETDuet-1 vector	IOL
Phenylacetyl-CoA	Sigma–Aldrich
pHis17_ACSM2B	IOL
pHis17_GLYAT	IOL
Pierce® Universal Nuclease for cell lysis (5kU)	Thermo Scientific
Precision Plus Protein™ Dual Colour Standards 10-250 kDa	Bio-Rad
Primers	Inqaba biotec™ and integrated DNA technologies®
pUC 19 linear vector and insert	TakaraBio
pUC19 control DNA	TakaraBio

Pyridine	Sigma–Aldrich
Qubit™ Protein and Protein BR (Broad Range) Assay kit	Thermo Fischer
Restriction enzyme (BamHI-HF)	New England Biolabs® Inc.
Restriction Enzyme (NdeI)	New England Biolabs® Inc.
Restriction enzyme (XbaI)	Fermentas®
Restriction Enzyme (XmaJI (AvrII))	Fermentas®
SOC media	Sigma–Aldrich
Sodium acetate	Melford
Sodium Chloride (NaCl)	Merck
Sodium dodecyl sulfate	Melford
Sodium phosphate (NaH_2PO_4)	Sigma–Aldrich
Stellar™ competent cells	TakaraBio
TEMED	Merck
Towbin blocking buffer	Bio-Rad
Tris-HCl	Melford
Triton X-100	Merck
Tryptone	Merck
Yeast extract	Pronadisa
β-Mercapto ethanol	Sigma–Aldrich
Equipment	
`Centrifuge Tubes (15/50 mL)	Cellstar®
0.8/0.2 µm Acrodisc PF Syringe Filter	Pall Corporation
14-mL Round bottom tubes	Falcon®
22-G needles and 50 mL syringes	Neoject
Agarose gel running chamber	Bio-Rad
Autosampler 7693	Agilent technologies
Centrifuge Tubes (0.5 mL)	QIAGEN
Centrifuge Tubes (1.5/2.0 mL)	Biopointe
ChemiDoc™MP Imaging system	Bio-Rad
Clear GC–MS vials	Agilent technologies
Dark Reader® Transilluminator	Clare Chemical Research
Fixed angle rotor	Thermo Scientific
Freezer (-20°C)	Thermo Scientific
Freezer (-80°C)	Thermo Scientific

GC System 7890A	Agilent technologies
GC–MS vial inserts	Agilent technologies
General glassware	IOL
Glass Beads	Retch
Heating block	Grant
Incubator	HT Infors
NanoDrop One	Inqaba Biotech™
Nitrocellulose membrane	Bio-Rad
Pegasus HT-TOFMS	LECO
Pierce™ Protein Concentrator PES, 30 MWCO	Thermo Scientific
Protino® Ni-TED Packed Columns	Macherey-Nagel
Qubit™ 2.0 Fluorometer	Thermo Scientific
Refrigerator (4°C)	Bauer
Screw cap vials	Agilent technologies
SDS–PAGE gel running chamber	Bio-Rad
Spectrophotometer	Biochrom
SpeedVac Concentrator	Thermo Fischer
Sterile Blade	Lawton
Swinging bucket rotor	Thermo Scientific
Thermocycler	Bio-Rad
Transblot® Turbo™ Transfer system	Bio-Rad
Vibration Mill (MM400)	Retch
Water bath	Grant
Wattmann Blotting paper (MN 218 B)	Macherey-Nagel
Western blot sponges	IOL
Software	
MEGAX (Version 10.2.2.)	From service providers as software
ImageLab (Version 5.2.1.)	
FinchTV (Version 1.4.0.)	
LECO® ChromaTOF® Software (Version 4.50.8)	
NIST MS Search (Version 2.0)	
SnapGene® (Version 3.2.1)	
Facilities Used	

CAF: Stellenbosch (South Africa)	Stellenbosch University

*IOL-Was readily available in our lab

Appendix B – Quick reference of the conversion of µM to ppm

Molars to ppm (ug/mL)					10 ppm = 10 µg/ml
					Add X mg amount of substance to X mL distilled water to ge 1 mg/mL solution.
Molar mass (g/mol)	Isovalerylglycine 159,18	Benzoylglycine 179,17	Phenyacetylglycine 193,19		Use C1V1=C2V2
					Add 500 µL of stock (1 mg/mL) solution to 49.5 mL dd water. result is 10 ppm solution
	Molars (µM)				PPM
	6	6	5		1
	31	28	26		5
	63	56	52		10
	94	84	78		15
	126	112	104		20
	157	140	129		25
	188	167	155		30
	220	195	181		35
	251	223	207		40
	283	251	233		45
	314	279	259		50
	346	307	285		55
	377	335	311		60
	408	363	336		65
	440	391	362		70
	471	419	388		75
	503	447	414		80
	534	474	440		85
	565	502	466		90
	597	530	492		95
	628	558	518		100

Figure 58: Conversion of µM to ppm.

References

Ahmad, M., Hirz, M., Pichler, H. & Schwab, H. 2014. Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. *Appl Microbiol Biotechnol*, 98(12):5301-5317. 10.1007/s00253-014-5732-5

Altenschmidt, U., Oswald, B., Steiner, E., Herrmann, H. & Fuchs, G. 1993. New aerobic benzoate oxidation pathway via benzoyl-coenzyme A and 3-hydroxybenzoyl-coenzyme A in a denitrifying *Pseudomonas sp. Journal of Bacteriology*:4851-4858.

Amsel, L.P. & Levy, G. 1969. Drug biotransformation interactions in man. II. A pharmacokinetic study of the simultaneous conjugation of benzoic and salicylic acids with glycine. *J Pharm Sci*, 58(3):321-326. 10.1002/jps.2600580307

Assur, Z., Hendrickson, W.A. & Mancia, F. 2012. Tools for coproducing multiple proteins in mammalian cells. *Methods Mol Biol*, 801:173-187. 10.1007/978-1-61779-352-3_12

Azevedo, M.S., Seraglio, S.K.T., Rocha, G., Balderas, C.B., Piovezan, M., Gonzaga, L.V., ... Costa, A.C.O. 2017. Free amino acid determination by GC-MS combined with a chemometric approach for geographical classification of bracatinga honeydew honey (Mimosa scabrella Bentham). *Food Control*, 78:383-392. 10.1016/j.foodcont.2017.03.008

Badenhorst, C.P., van der Sluis, R. & Erasmus, E. 2013. Glycine conjugation: Importance in metabolism, the role of glycine N-acyltransferase, and factors that influence interindividual variation. *Expert Opin Drug Metab Toxicol*, 9(9):1139-1153.

Badenhorst, C.P., 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 Metab Rev*, 46(3):343-361. 10.3109/03602532.2014.908903

Bale, N.J., Koenen, M., Yadav, S., Hopmans, E.C., Villanueva, L., Sinninghe Damsté, J.S. & Schouten, S. 2020. Diagnostic amide products of amino lipids detected in the microaerophilic bacteria Lutibacter during routine fatty acid analysis using gas chromatography. *Organic Geochemistry*, 144, 10.1016/j.orggeochem.2020.104027

Bartlett, K. & Gompertz, D. 1974. The specificity of of glycine-*N*-acylase and acylglycine excretion in the organicacidaemias. *Biochemical Medicine and Metabolic Biology*, 10(1):1139-1153.

Beale, D.J., Pinu, F.R., Kouremenos, K.A., Poojary, M.M., Narayana, V.K., Boughton, B.A., ... Dias, D.A. 2018. Review of recent developments in GC-MS approaches to metabolomics-based research. *Metabolomics*, 14(11):152. 10.1007/s11306-018-1449-2

Beezhold, B.L., Johnston, C.S. & Nochta, K.A. 2014. Sodium benzoate-rich beverage consumption is associated with increased reporting of ADHD symptoms in college students: a pilot investigation. *J Atten Disord*, 18(3):236-241. 10.1177/1087054712443156

Beyoglu, D. & Idle, J.R. 2012. The glycine deportation system and its pharmacological consequences. *Pharmacol Ther*, 135(2):151-167. 10.1016/j.pharmthera.2012.05.003

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(10):1331-1339. 10.1016/j.bcp.2011.12.031

Bisswanger, H. 2014. Enzyme assays. *Perspectives in Science*, 1(1-6):41-55. 10.1016/j.pisc.2014.02.005

Boomgaarden, I., Vock, C., Klapper, M. & Doring, F. 2009. Comparative analyses of disease risk genes belonging to the acyl-CoA synthetase medium-chain (ACSM) family in human liver and cell lines. *Biochem Genet*, 47(9-10):739-748. 10.1007/s10528-009-9273-z

Brass, E.P. 1994. Overview of coenzyme A metabolism and its role in cellular toxicity. *Chemico-Biological Interactions*, 90:203-214.

Bremer, J. 1983. Carnitine-Metabolism and functions. *Physiological Reviews*, 63:1420-1480.

BRENDA. 2019. *Information on EC 2.3.1.13 - glycine N-acyltransferase*. https://www.brenda-enzymes.org/enzyme.php?ecno=2.3.1.13#DISEASE Date of access: 23 Nov.

Cardenas, C.L., Bourgine, J., Cauffiez, C., Allorge, D., Lo-Guidice, J.M., Broly, F. & Chevalier, D. 2010. Genetic polymorphisms of glycine N-acyltransferase (GLYAT) in a French Caucasian population. *Xenobiotica*, 40(12):853-861. 10.3109/00498254.2010.519407

Catherine, C., Lee, K.H., Oh, S.J. & Kim, D.M. 2013. Cell-free platforms for flexible expression and screening of enzymes. *Biotechnol Adv*, 31(6):797-803. 10.1016/j.biotechadv.2013.04.009
Chen, D.W., Zhang, Y., Jiang, C.Y. & Liu, S.J. 2014. Benzoate metabolism intermediate benzoyl coenzyme A affects gentisate pathway regulation in Comamonas testosteroni. *Appl Environ Microbiol*, 80(13):4051-4062. 10.1128/AEM.01146-14
Chung, C.T., Niemela, S.L. & Miller, R.H. 1989. One-step preparation of competent *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences of the United States of America*, 86:2172-2175.
Conti, A. & Bickel, M.H. 1977. History of Drug Metabolism: Discoveries of the Major Pathways in the 19th Century. *Drug Metabolism review*, 6(1):1-50. 10.3109/03602537708993764
CrawfordScientific. 2022. *Troubleshooting GC peak shapes*. https://www.crawfordscientific.com/chromatography-blog/post/gc-troubleshooting-peak-shapes Date of access:
Crowther, J.R. 2009. *Methods in molecular biology: The ELISA guidebook.* 2. NY: Humana Press.
Cruz, E.R., Nguyen, H., Nguyen, T. & Wallace, I.S. 2019. Functional analysis tools for post-translational modification: a post-translational modification database for analysis of proteins and metabolic pathways. *Plant J*, 99(5):1003-1013. 10.1111/tpj.14372
Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I.M., ... Flicek, P. 2019. Ensembl 2019. *Nucleic Acids Res*, 47(D1):D745-D751. 10.1093/nar/gky1113
Davidson, P.M., Sofos, J.N. & Branen, A.L. 2005. *Antimicrobials in Food*. 145. Taylor & Francis Group.
de Sousa, C., Chalmers, R.A., Stacey, T.E., Tracey, B.M., Weaver, C.M. & Bradley, D. 1986. The response to L-carnitine and glycine therapy in isovaleric acidaemia. *European Journal of Pediatrics*, 5(144):451-456.
Del Olmo, A., Calzada, J. & Nunez, M. 2017. Benzoic acid and its derivatives as naturally occurring compounds in foods and as additives: Uses, exposure, and controversy. *Crit Rev Food Sci Nutr*, 57(14):3084-3103. 10.1080/10408398.2015.1087964
Dempsey, D.R., Bond, J.D., Carpenter, A.M., Rodriguez Ospina, S. & Merkler, D.J. 2014. Expression, purification, and characterization of mouse glycine N-acyltransferase in Escherichia coli. *Protein Expr Purif*, 97:23-28. 10.1016/j.pep.2014.02.007
Dercksen, M., Duran, M., Ijlst, L., Mienie, L.J., Reinecke, C.J., Ruiter, J.P., ... Wanders, R.J. 2012. Clinical variability of isovaleric acidemia in a genetically homogeneous population. *J Inherit Metab Dis*, 35(6):1021-1029. 10.1007/s10545-012-9457-2
Diaz, E., Ferrandez, A. & Garcia, J.L. 2000. Bacterial promoters triggering biodegradation of aromatic pollutants. *Curr Opin Biotechnol*, 11:467-475.
Diaz, E., Ferrandez, A., Prieto, M.A. & Garcia, J.L. 2001. Biodegradation of aromatic compounds by Escherichia coli. *Microbiol Mol Biol Rev*, 65(4):523-569, table of contents. 10.1128/MMBR.65.4.523-569.2001
Dopp, J.L. & Reuel, N.F. 2018. Process optimization for scalable E. coli extract preparation for cell-free protein synthesis. *Biochemical Engineering Journal*, 138:21-28. 10.1016/j.bej.2018.06.021
Ensenauer, R., Vockley, J., Willard, J., Huey, J.C., Sass, J.O., Edland, S.D., ... Matern, D. 2004. A common mutation is associated with a mild, potentially asymptomatic phenotype in patients with isovaleric acidemia diagnosed by newborn screening. *American Journal of Human Genetics*, 75:1136-1142.
Erasmus, D., Mels, C.M.C., Louw, R., Lindeque, J.Z. & Kruger, R. 2019. Urinary Metabolites and Their Link with Premature Arterial Stiffness in Black Boys: The ASOS Study. *Pulse (Basel)*, 6(3-4):144-153. 10.1159/000492155
Flath, B., Rolinski, B. & Roscher, A.A. 1997. Simple high-preformance liquid chromatographic method for the detection of phenylpropionylglycine in urine as a diagnostic tool in inherited medium-chain acyl CoA dehydrogenase deficiency. *Journal of chromatography B: Biomedical Sciences and Applications*, 694(1):227-232.

Forster-Fromme, K. & Jendrossek, D. 2008. Biochemical characterization of isovaleryl-CoA dehydrogenase (LiuA) of Pseudomonas aeruginosa and the importance of liu genes fora functional catabolic pathway of methyl-branched compounds. *FEMS Microbiol Lett*, 286(1):78-84. 10.1111/j.1574-6968.2008.01266.x
Gagoski, D., Polinkovsky, M.E., Mureev, S., Kunert, A., Johnston, W., Gambin, Y. & Alexandrov, K. 2016. Performance benchmarking of four cell-free protein expression systems. *Biotechnol Bioeng*, 113(2):292-300. 10.1002/bit.25814
Gillingham, M.B., Heitner, S.B., Martin, J., Rose, S., Goldstein, A., El-Gharbawy, A.H., ... Vockley, J. 2017. Triheptanoin versus trioctanoin for long-chain fatty acid oxidation disorders: a double blinded, randomized controlled trial. *J Inherit Metab Dis*, 40(6):831-843. 10.1007/s10545-017-0085-8
Glasgow, J. 2006. Reye's syndrome: The case for a casual link with aspirin. *Drug Safety*, 29(12):1111-1121.
Gleba, Y., Klimyuk, V. & Marillonnet, S. 2007. Viral vectors for the expression of proteins in plants. *Curr Opin Biotechnol*, 18(2):134-141. 10.1016/j.copbio.2007.03.002
Goetzman, E.S., Bharathi, S.S., Zhang, Y., Zhao, X.J., Dobrowolski, S.F., Peasley, K., ... Monga, S.P. 2020. Impaired mitochondrial medium-chain fatty acid oxidation drives periportal macrovesicular steatosis in sirtuin-5 knockout mice. *Sci Rep*, 10(1):18367. 10.1038/s41598-020-75615-3
Gregersen, N., Kølvraa, S. & mortensen, P.B. 1986. Acyl-CoA: Glycine N-acyltransferase: In vitro studies on the glycine conjugation of straight- and branched-chain acylCoA esters in human liver. *Biochemical Medicine and Metabolic Biology*, 35(2):210-218.
Gregus, Z., Fekete, T., Varga, F. & Klaassen, C.D. 1992. Availibility of glycine and coenzyme A limits glycine conjugation *In vivo*. *Drug Metabolism and Pharmacokinetics*, 20(2):234-240.
Gundersen, R.Y., Vaagenes, P., Breivik, T., Fonnum, F. & Opstad, P.K. 2005. Glycine--an important neurotransmitter and cytoprotective agent. *Acta Anaesthesiol Scand*, 49(8):1108-1116. 10.1111/j.1399-6576.2005.00786.x
Higgins, N.P., Vologodskii, A.V., Tolmasky, M. & Alonso, J.C. 2015. Topological Behavior of Plasmid DNA. *Microbiology Spectrum*, 3(2), 10.1128/microbiolspec.PLAS-0036-2014
Houten, S.M. & Wanders, R.J. 2010. A general introduction to the biochemistry of mitochondrial fatty acid beta-oxidation. *J Inherit Metab Dis*, 33(5):469-477. 10.1007/s10545-010-9061-2
Huang, C.J., Lin, H. & Yang, X. 2012. Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. *J Ind Microbiol Biotechnol*, 39(3):383-399. 10.1007/s10295-011-1082-9
IHSMarkit. 2017. *Benzoic acid*. https://ihsmarkit.com/products/benzoic-acid-chemical-economics-handbook.html Date of access: 17 Mar. 2020.
Kaldalu, N., Toots, U., de Lorenzo, V. & Ustav, M. 2000. Functional domains of the TOL plasmid transcription factor XylS. *Journal of Bacteriology*:1118-1126.
Katzen, F., Chang, G. & Kudlicki, W. 2005. The past, present and future of cell-free protein synthesis. *Trends Biotechnol*, 23(3):150-156. 10.1016/j.tibtech.2005.01.003
Keller, W. 1842. On the conversion of benzoic acid into hippuric acid. *Retrospect of the medical science*, 4(13),
Kelley, M. & Vessey, D.A. 1994. Characterization of the Acyl-CoA:Amino Acid NAcyltransferases from Primate Liver Mitochondria. *Journal of Biochemical and Molecular Toxicology*, 9(3):153-158.
Kilic, N., Balkan, E., Akgoz, S., Sen, N. & Dogruyol, H. 2006. Comparison of the effectiveness and side-effects of tolterodine and oxybutynin in children with detrusor instability. *International journal of Urology*, 13:105-108.
Kiseleva, O., Kurbatov, I., Ilgisonis, E. & Poverennaya, E. 2021. Defining Blood Plasma and Serum Metabolome by GC-MS. *Metabolites*, 12(1), 10.3390/metabo12010015
Knights, K. & Miners, J. 2012. Amino Acid Conjugation: A novel route of xenobiotic carboxylic acid metabolism in man. In: Lyubimov, A.V., ed. *Encyclopedia of Drug Metabolism and Interactions*. 1: John Wiley & sons, Inc. pp. 595-610. 10.1002/9780470921920.edm020

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. 10.1517/17425255.3.2.159
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-coenzyme A synthetase ACSM2A. *J Mol Biol*, 388(5):997-1008. 10.1016/j.jmb.2009.03.064
Koek, M.M., Jellema, R.H., van der Greef, J., Tas, A.C. & Hankemeier, T. 2011. Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. *Metabolomics*, 7(3):307-328. 10.1007/s11306-010-0254-3
Koshiba, S., Motoike, I.N., Saigusa, D., Inoue, J., Aoki, Y., Tadaka, S., ... Yamamoto, M. 2020. Identification of critical genetic variants associated with metabolic phenotypes of the Japanese population. *Commun Biol*, 3(1):662. 10.1038/s42003-020-01383-5
Krieger, I. & Tanaka, K. 1976. Theraputic effects of glycine in isovaleric acidaemia. *Pediatric Research*, 10:25-29.
Kroll, J., Klinter, S., Schneider, C., Voss, I. & Steinbuchel, A. 2010. Plasmid addiction systems: perspectives and applications in biotechnology. *Microb Biotechnol*, 3(6):634-657. 10.1111/j.1751-7915.2010.00170.x
Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227:680-685.
Lehnert, W. & Niederhoff, H. 1981. 4-hydroxyisovaleric acid: A new metabolite in isovaleric acidaemia. *European Journal of Pediatrics*, 136:281-283.
Lennerz, B.S., Vafai, S.B., Delaney, N.F., Clish, C.B., Deik, A.A., Pierce, K.A., ... Mootha, V.K. 2015. Effects of sodium benzoate, a widely used food preservative, on glucose homeostasis and metabolic profiles in humans. *Mol Genet Metab*, 114(1):73-79. 10.1016/j.ymgme.2014.11.010
Leung, K.S.-Y. & Fong, B.M.-W. 2013. LC–MS/MS in the routine clinical laboratory: has its time come? *Analytical and Bioanalytical Chemistry*, 406(9-10):2289-2301. 10.1007/s00216-013-7542-5
Levy, G. 1979. Pharmacokinetics of salicylate in man. *Drug Metab Rev*, 9(1):3-19. 10.3109/03602537909046431
Levy, G. & Amsel, L.P. 1966. Kinetics of competitive inhibition of salicylic acid conjugation with glycine in man. *Biochemical Pharmacology*, 15:1033-1038.
López-Corcuera, B., Geerlings, A. & Aragón, C. 2009. Glycine neurotransmitter transporters: an update. *Molecular Membrane Biology*, 18(1):13-20. 10.1080/09687680010028762
Lorenz, T.C. 2012. Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *J Vis Exp*, (63):e3998. 10.3791/3998
Malhotra, A. 2009. Chapter 16 Tagging for Protein Expression. In. *Guide to Protein Purification, 2nd Edition*. pp. 239-258.
Marsh, K.J., Wallis, I.R. & Foley, W.J. 2005. Detoxification rates constrain feeding in common brushtail possums (Trichosurus vulpecula). *Ecology*, 86(11):2946-2954.
Mastrangelo, A., Ferrarini, A., Rey-Stolle, F., Garcia, A. & Barbas, C. 2015. From sample treatment to biomarker discovery: A tutorial for untargeted metabolomics based on GC-(EI)-Q-MS. *Anal Chim Acta*, 900:21-35. 10.1016/j.aca.2015.10.001
Matsuo, M., Terai, K., Kameda, N., Matsumoto, A., Kurokawa, Y., Funase, Y., ... Kishimoto, T. 2012. Designation of enzyme activity of glycine-N-acyltransferase family genes and depression of glycine-N-acyltransferase in human hepatocellular carcinoma. *Biochem Biophys Res Commun*, 420(4):901-906. 10.1016/j.bbrc.2012.03.099
Mawal, Y. & Qureshi, I. 1994. Purification to homogeneity of mitochondrial acyl CoA: Glycine N-acyltransferase from human liver. *Biochemical and Biophysical Research Communications*, 205(2):1373-1379.
Mawal, Y., Paradis, K. & Qureshi, I. 1997. Developmental profile of mitochondrial glycine N-acyltransferase in humar liver. *Journal of Pediatrics*, 130(6):1003-1007.
McKenzie, E.A. & Abbott, W.M. 2018. Expression of recombinant proteins in insect and mammalian cells. *Methods*, 147:40-49. 10.1016/j.ymeth.2018.05.013

Merlin, M., Gecchele, E., Capaldi, S., Pezzotti, M. & Avesani, L. 2014. Comparative evaluation of recombinant protein production in different biofactories: the green perspective. *Biomed Res Int*, 2014:136419. 10.1155/2014/136419
Mitchell, G.A., Gauthier, N., Lesimple, A., Wang, S.P., Mamer, O. & Qureshi, I. 2008. Hereditary and acquired diseases of acyl-coenzyme A metabolism. *Mol Genet Metab*, 94(1):4-15. 10.1016/j.ymgme.2007.12.005
Moldoveanu, S.C. & David, V. 2018. Derivatization methods in GC and GC/MS. In: Kusch, P., ed. *Gas Chromatography*. London: IntechOpen.
Nair, B. 2001. Final report on the safety assessment of benzyl alcohol, benzoic acid, and sodium benzoate. *International journal of toxicology*, 20:23-50.
Nakajima, T., Horiuchi, M., Yamanaka, H., Kizaki, Z., Inoue, F., Kodo, N., ... Sawada, T. 1997. The effect of carnitine on ketogenesis in perfused livers from juvenile visceral steatosis mice with systematic carnitine deficiency. *Pediatric Research*, 42:108-113.
Nandi, D.L., Lucas, S.V. & Webster, L.T.J. 1979. Benzoyl-coenzyme A:glycine N-acyltransferase and phenylacetyl-coenzyme A:glycine N-acyltransferase from bovine liver mitochondria purification and characterization. *Journal of Biological Chemistry*, 254(15):7230-7237.
Noorafshan, A., Erfanizadeh, M. & Karbalay-Doust, S. 2014. Sodium benzoate, a food preservative, induces anxiety and motor impairment in rats. *Neurosciences*, 19(1):24-28.
Novy, R. & Morris, B. s.a. Use of glucose to control basal expression in the pET system. *inNovations*, 13:8-10.
Owczarek, B., Gerszberg, A. & Hnatuszko-Konka, K. 2019. A Brief Reminder of Systems of Production and Chromatography-Based Recovery of Recombinant Protein Biopharmaceuticals. *Biomed Res Int*, 2019:4216060. 10.1155/2019/4216060
Oyewole, O., Dere, A.D. & Okoro, O.E. 2012. Sodium benzoate mediated hepatorenal toxicity in wistar rat: Modulatory effects of *Azadirachta indica* (Neem) leaf. *European journal of medicinal plants*, 2(1):11-18.
Pereira Braga, C. & Adamec, J. 2019. Metabolome Analysis. In. *Encyclopedia of Bioinformatics and Computational Biology*. pp. 463-475.
Perez-Palacios, T., Barroso, M.A., Ruiz, J. & Antequera, T. 2015. A Rapid and Accurate Extraction Procedure for Analysing Free Amino Acids in Meat Samples by GC-MS. *Int J Anal Chem*, 2015:209214. 10.1155/2015/209214
Piper, J.D. & Piper, P.W. 2017. Benzoate and Sorbate Salts: A Systematic Review of the Potential Hazards of These Invaluable Preservatives and the Expanding Spectrum of Clinical Uses for Sodium Benzoate. *Comprehensive Reviews in Food Science and Food Safety*, 16(5):868-880. 10.1111/1541-4337.12284
Prabhu, A.V., Luu, W. & Brown, A.J. 2017. Measuring Activity of Cholesterol Synthesis Enzymes Using Gas Chromatography/Mass Spectrometry. *Methods Mol Biol*, 1583:211-219. 10.1007/978-1-4939-6875-6_16
Puetz, J. & Wurm, F.M. 2019. Recombinant Proteins for Industrial versus Pharmaceutical Purposes: A Review of Process and Pricing. *Processes*, 7(8), 10.3390/pr7080476
Rechner, A.R., Kuhnle, G. & Bremmer, P. 2002. The metabolic fate of dietary polypohenols in humans. *Free radical*, 46(3):343-361.
Reinecke, C.J., Koekemoer, G., van der Westhuizen, F.H., Louw, R., Lindeque, J.Z., Mienie, L.J. & Smuts, I. 2012. Metabolomics of urinary organic acids in respiratory chain deficiencies in children. *Metabolomics*, 8(2):264.
Rodig, S.J. 2019. Detecting Horseradish Peroxidase-Labeled Cells. *Cold Spring Harb Protoc*, 2019(4), 10.1101/pdb.prot099713
Rohwer, J.M., Schutte, C. & van der Sluis, R. 2021. Functional Characterisation of Three Glycine N-Acyltransferase Variants and the Effect on Glycine Conjugation to Benzoyl-CoA. *Int J Mol Sci*, 22(6), 10.3390/ijms22063129
Rosano, G.L. & Ceccarelli, E.A. 2014. Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol*, 5:172. 10.3389/fmicb.2014.00172
Rosenblum, G. & Cooperman, B.S. 2014. Engine out of the chassis: cell-free protein synthesis and its uses. *FEBS Lett*, 588(2):261-268. 10.1016/j.febslet.2013.10.016

Sachdev, D. & Chirgwin, J.M. 1998. Solubility of proteins isolated from inclusion bodies is enhanced by fusion to maltose-binding protein or thioredoxin. *Protein Expr Purif*, 12:122-132.
Saida, F., Uzan, M., Odaert, B. & Bontems, F. 2006. Expression of highly toxic genes in *E. coli*: Special strategies and genetic tools. *Current Protein and Peptide Science*, 7:47-56.
Sambrook, J. & Russell, D.W. 2001. *Molecular cloning*. 3. 3. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. (Molecular cloning. A laboratory manual.).
Schachter, D. & Taggart, J.V. 1954. Product inhibition in the glycine N-acylase reaction. *Journal of Biological Chemistry*, 211(1):271-278.
Schinn, S.M., Broadbent, A., Bradley, W.T. & Bundy, B.C. 2016. Protein synthesis directly from PCR: progress and applications of cell-free protein synthesis with linear DNA. *N Biotechnol*, 33(4):480-487. 10.1016/j.nbt.2016.04.002
Schutte, C., van der Sluis, R. & Erasmus, E. 2019. Characterization of the enzyme activty and substrate specificity of GLYAT haplotypes.
Scopes, R.K. 2002. Enzyme activity and assays. In: *ENCYCLOPEDIA OF LIFE SCIENCES*: Macmillan Publishers Ltd. pp. 1-6.
Silvestro, L., Tarcomnicu, I. & Rizea, S. 2013. Matrix Effects in Mass Spectrometry Combined with Separation Methods — Comparison HPLC, GC and Discussion on Methods to Control these Effects. In. *Tandem Mass Spectrometry - Molecular Characterization*.
Simantini, D., Mohapatra, S.C. & Hsu, J.T. 1999. Studies on primer-dimers formation in polymerase chain reaction (PCR). *Biotechnology Techniques*, 13:643-646.
Srinithi, P. & Bhargava, A. 2018. *Effects of sodium benzoate using zebrafish animal model.* India: Indian Institute of Technology Hyderabad.
Stanojevic, D.C., L.; Stefanovic, O.; Solujic-Sukdolak, S.L. 2009. ANTIMICROBIAL EFFECTS OF SODIUM BENZOATE, SODIUM NITRITE AND POTASSIUM SORBATE AND THEIR SYNERGISTIC ACTION IN VITRO. *Bulgarian Journal of Agricultural Science*, 15(4):307-311.
Swenson, D. s.a. The effect of sodium benzoate on the survival and hatch rates of zebrafish embryos.,
TakaraBio. In-Fusion® HD EcoDry™ Cloning Kit User Manual. In: Inc., T.B.U.,ed. pp. 1-14.
Tanaka, K., Budd, M.A., Efron, M.L. & Isselbacher, K.J. 1966. Isovaleric acidaemia: A new genetic defect of leucine metabolism. *Biochemistry*, 56:236-242.
Teufel, R., Mascaraque, V., Ismail, W., Voss, M., Perera, J., Eisenreich, W., ... Fuchs, G. 2010. Bacterial phenylalanine and phenylacetate catabolic pathway revealed. *Proc Natl Acad Sci U S A*, 107(32):14390-14395. 10.1073/pnas.1005399107
ThermoFisher. 2002. *EnzChek Phosphate Assay Kit (E-6646)*. https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2Fmp06646.pdf&title=RW56Q2hlayBQaG9zcGhhdGUgQXNzYXkgS2l0 Date of access: 06 Feb. 2020.
ThermoFisher. 2020. *Overview of post-translational modifications*. https://www.thermofisher.com/za/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-post-translational-modification.html Date of access: 23 Jan.
Tolia, N.H. & Joshua-Tor, L. 2006. Strategies for protein coexpression in Escherichia coli. *Nat Methods*, 3(1):55-64. 10.1038/nmeth0106-55
Tripathi, N.K. & Shrivastava, A. 2019. Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Front Bioeng Biotechnol*, 7:420. 10.3389/fbioe.2019.00420
Tsay, H.J., Wang, Y.H., Chen, W.L., Huang, M.Y. & Chen, Y.H. 2007. Treatment with sodium benzoate leads to malformation of zebrafish larvae. *Neurotoxicol Teratol*, 29(5):562-569. 10.1016/j.ntt.2007.05.001
Ure, A. 1841. On gouty concretions, with a new method of treatment. *Medical-Cirurgical Transactions*, 24(24):30-35. 10.1177/095952874102400105
Vadali, R.V., Bennett, G.N. & San, K.Y. 2004. Cofactor engineering of intracellular CoA/acetyl-CoA and its effect on metabolic flux redistribution in Escherichia coli. *Metab Eng*, 6(2):133-139. 10.1016/j.ymben.2004.02.001

van der Sluis, R. 2018. Analyses of the genetic diversity and protein expression variation of the acyl: CoA medium-chain ligases, ACSM2A and ACSM2B. *Mol Genet Genomics*, 293(5):1279-1292. 10.1007/s00438-018-1460-3
van der Sluis, R. & Erasmus, E. 2016. Xenobiotic/medium chain fatty acid: CoA ligase - a critical review on its role in fatty acid metabolism and the detoxification of benzoic acid and aspirin. *Expert Opin Drug Metab Toxicol*, 12(10):1169-1179. 10.1080/17425255.2016.1206888
van der Sluis, R., Badenhorst, C.P.S., Van Der Westhuizen, F.H. & Van Dijk, A.A. 2013. Characterisation of the influence of genetic variations on the enzyme activity of a recombinant human glycine N-acyltransferase. *Gene*, 515(2):447-453. 10.1016/j.gene.2012.12.003
van der Sluis, R., Ungerer, V., Nortje, C., A, A.v.D. & Erasmus, E. 2017. New insights into the catalytic mechanism of human glycine N-acyltransferase. *J Biochem Mol Toxicol*, 31(11), 10.1002/jbt.21963
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(1):126-134. 10.1016/j.gene.2015.06.081
van der Westhuizen, F.H., Pretorious, P.J. & Erasmus, E. 2000. The Utilization of Alanine, Glutamic Acid, and Serine as Amino Acid Substrates for Glycine N-Acyltransferase. *Journal of Biochemical and Molecular Toxicology*, 14(2):102-109.
Van Hove, J.L., Vande Kerckhove, K., Hennermann, J.B., Mahieu, V., Declercq, P., Mertens, S., ... Jaeken, J. 2005. Benzoate treatment and the glycine index in nonketotic hyperglycinaemia. *J Inherit Metab Dis*, 28(5):651-663. 10.1007/s10545-005-0033-x
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. *Biochemica et Biophysica Acta*, 1428:455-462.
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. 10.1002/jbt.10056
Villas-Bôas, S.G., Mas, S., Åkesson, M., Smedsgaard, J. & Nielsen, J. 2005. Mass spectrometry in metabolome analysis. *Mass Spectrometry Reviews*, 24(5):613-646. 10.1002/mas.20032
Vivekanandan-Giri, A., Byun, J. & Pennathur, S. 2011. Quantitative analysis of amino Acid oxidation markers by tandem mass spectrometry. *Methods Enzymol*, 491:73-89. 10.1016/B978-0-12-385928-0.00005-5
Wang, W., Wu, Z., Dai, Z., Yang, Y., Wang, J. & Wu, G. 2013. Glycine metabolism in animals and humans: implications for nutrition and health. *Amino Acids*, 45(3):463-477. 10.1007/s00726-013-1493-1
Watanabe, H., Paxton, R.L., Tolerico, M.R., Nagalakshmi, V.K., Tanaka, S., Okusa, M.D., ... Gomez, R.A. 2020. Expression of Acsm2, a kidney-specific gene, parallels the function and maturation of proximal tubular cells. *Am J Physiol Renal Physiol*, 319(4):F603-F611. 10.1152/ajprenal.00348.2020
Watkins, P. 1997. Fatty acid activation. *Progress in Lipid Research*, 36(1):55-83.
Watkins, P.A., Maiguel, D., Jia, Z. & Pevsner, J. 2007. Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome. *J Lipid Res*, 48(12):2736-2750. 10.1194/jlr.M700378-JLR200
Webster, L.T., Siddiqui, U.A., Lucas, S.V., Strong, J.M. & Mieyal, J.J. 1976. Identification of separate acyl- CoA:glycine and acyl-CoA:L-glutamine N-acyltransferase activities in mitochondrial fractions from liver of rhesus monkey and man. *J Biol Chem*, 251(11):3352-3358.
Wibbertmann, A., Kielhorn, J., Koennecker, G., Mangelsdorf, I. & Melber, C. 2005. *Benzoic acid and sodium benzoate*. https://apps.who.int/iris/handle/10665/42310
Wu, J. 2008. Post-translational modifications of proteins. *Methods Mol Biol*, 446(Issue).
Wysocki, S.J., French, N.P. & Grauaug, A. 1983. organic aciduria associated with isovaleric acidaemia. *Clinical chemistry*, 29:1002-1003.

Yamamoto, A., Nonen, S., Fukuda, T., Yamazaki, H. & Azuma, J. 2009. Genetic polymorphisms of glycine N-acyltransferase in Japanese individuals. *Drug Metab Pharmacokinet*, 24(1):114-117. 10.2133/dmpk.24.114
Yin, J., Li, G., Ren, X. & Herrler, G. 2007. Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *J Biotechnol*, 127(3):335-347. 10.1016/j.jbiotec.2006.07.012
Zaar, A., Eisenreich, W., Bacher, A. & Fuchs, G. 2001. A novel pathway of aerobic benzoate catabolism in the bacteria Azoarcus evansii and Bacillus stearothermophilus. *J Biol Chem*, 276(27):24997-25004. 10.1074/jbc.M100291200
Zhang, A., Sun, H., Wang, P., Han, Y. & Wang, X. 2012. Modern analytical techniques in metabolomics analysis. *Analyst*, 137(2):293-300. 10.1039/c1an15605e
Zhang, H., Lang, Q., Li, J., Zhong, Z., Xie, F., Ye, G., ... Yu, L. 2007. Molecular Cloning and Characterization of a Novel Human Glycine-N-acyltransferase Gene GLYATL1, Which Activates Transcriptional Activity of HSE Pathway. *International Journal of Molecular Sciences*, 8(5):433-444. 10.3390/i8050433