The metabolism of plant glucosinolates by gut bacteria

Fatma Cebeci

A thesis submitted for the degree of Doctor of Philosophy to the University of East Anglia

Institute of Food Research

April, 2017

PhD Thesis 2017 Fatma Cebeci

Metabolism of Plant Glucosinolates by Gut Bacteria

ABSTRACT

Glucosinolates found in cruciferous vegetables are degraded by plant myrosinases into bioactive isothiocyanates (ITCs) which have been recognised as potent anticancer compounds. During cooking, plant myrosinases are heat inactivated so ITC production is dependent on the myrosinase-like enzymes produced by the gut bacteria. This study is focused on investigating glucosinolate metabolism by the human gut bacteria and identifying the enzymes that play a crucial role.

Human gut bacteria that were previously reported to metabolise glucosinolates were investigated in this study. In addition, 98 more human gut strains were isolated using a glucoraphanin enrichment method. It was hypothesised that bacterial myrosinases are β-glucosidases with specificity for glucosinolates. To identify the first bacterial myrosinase from the human gut, four putative β-glucosidases from *Enterococcus casseliflavus* CP1 and *Escherichia coli* FI10944 were cloned and heterologously expressed in *E. coli*. An alternative approach using a combination of ion exchange chromatography and gel filtration was also carried out to identify the bacterial myrosinase of *E. coli* FI109444.

It has been reported that some gut bacteria require a reduction step to metabolise methylsulfinylalkyl glucosinolates (such as glucoraphanin) that converts them into methylthioalkyl glucosinolates (such as glucoerucin) to produce ITCs. To identify the responsible reductase, candidate reductase genes were cloned and expressed in *E. coli*. Methionine sulfoxide reductase B (MsrB) from *Escherichia coli* VL8 and *Lactobacillus agilis* R16 was found to reduce glucoraphanin to glucoerucin under the conditions tested.

A bacterial myrosinase of *Citrobacter* WYE1 of soil origin was previously identified and myrosinase activity of this enzyme was characterised using cell-free extracts. In this study the myrosinase gene was heterologously expressed in *E. coli* to allow purification and characterisation. The recombinant enzyme showed activity against several glucosinolate substrates and protein was produced for crystallographic studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

OUTPUTS FROM THIS PROJECT

Publications

- Albaser A., Kazana E., Bennett M. H., Cebeci F., Luang-In V., Spanu P.D., Rossiter J.T. *Discovery of a Bacterial Glycoside Hydrolase Family 3 (GH3) β-Glucosidase with Myrosinase Activity from a Citrobacter Strain Isolated from Soil.* Journal of Agricultural and Food Chemistry. **2016, *64 (***7), p: 1520–1527.
- Luang-In, V., Narbad, A., Cebeci, F., Bennett, M., Rossiter, JT. *Identification of Proteins Possibly Involved in Glucosinolate Metabolism in L. agilis R16 and E. coli VL8.* Protein J. 2015 Apr; **34**(2): p.135-46.

Presentations

- Cebeci F., Mayer M., Rossiter J. T., Mithen R., Narbad A. Biotransformation of glucosinolates by human gut bacteria to Bioactive Isothiocyanates. *Food bioactives & Health*, 13-15 September 2016, Norwich, UK (Oral Presentation).
- Cebeci F., Mayer M., Rossiter J. T., Mithen R., Narbad A. Bacterial enzymes involved in glucosinolate metabolism. *IFR Science Symposium*, 2015, Norwich, UK (Oral presentation).
- Cebeci, F., Mayer, M., Rossiter J. T., Mithen, R. and Narbad, A. (2014). Glucosinolate Metabolism by bacterial enzymes. *Glucosinolates and Beyond.* 12-15 October, Wageningen, Netherlands (Poster Presentation).
- Cebeci, F., Pigram A., Mayer, M., Rossiter J. T., Mithen, R. and Narbad, A. (2014). Glucosinolate Metabolism by bacterial myrosinase-like enzymes., *Rowett-INRA 2014. Gut Microbiology: From Sequence to function*, 16-19 June 2014, Aberdeen, Scotland, UK (Poster Presentation).

ABBREVIATIONS

6XHis-tag	Hexa-histidine tag
ABTS	2-2'-azino-bis (3-ethylbenzthiazoline-6-sulphonic acid)
AITC	Allyl ITC
BCAT4	Branched-chain aminotransferase4
BHI	Brain heart infusion
BITC	Benzyl ITC
bp	Base pair
BSA	Bovine serum albumin
CE	Competitive exclusion
CFE	Cell-free extract
C-terminal	Carboxy-terminal
DEAE	Diethylaminoethyl
DGGE	Denaturing gradient gel electrophoresis
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGFR1	Epidermal growth factor receptor 1
ESP	Epithiospecifier protein
FISH	Fluorescence *in situ* hybridisation
FPLC	Fast protein liquid chromatography
GC-MS	Gas chromatography-mass spectrometry
gDNA	Genomic DNA
GF	Gel filtration
GH1	Glycosyl hydrolase 1
GH3	Glycosyl hydrolase 3
GRP	Glucoraphanin
GSH	Glutathione
GST	Glutathione-S-transferase
HER1	Human epidermal growth factor receptor 2
HPLC	High performance liquid chromatography
IAA	Indole-3-acetic acid
IARC	International Agency for Research on Cancer
IEX	Ion Exchange
IPMI	Isopropyl isomerase
IPM-DH	Isopropylmalate dehydrogenase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothiocyanate
K_m	Michaelis-Menten constant
L	Luria
LCMS	Liquid chromatography-mass spectrometry
LCMS/MS	Liquid chromatography mass spectrometry-mass spectrometry
LDL-C	Low density lipoprotein cholesterol
MAM	Methylthioalkylmalate
MetSO	Methionine sulphoxide
$MetSO_2$	Methionine sulphone
MOPS	3 [N-Morpholino]propanesulfonic acid

MQ	Milli-Q
MRS	de Man, Rogosa and Sharpe
Msr	Methionine sulphoxide reductase
NAC	N-acetyl cysteine
NADPH	Nicotinamide adenine dinucleotide phosphate
NB	Nutrient broth
NFκB	Nuclear factor kappa B
Ni-NTA	Nickel nitrolotriacetic acid
NMR	Nuclear magnetic resonance
Nrf2-ARE	Nuclear factor (erythroid-derived 2)-like 2-Antioxidant Response Element
NSP	Nitrile specifier protein
N-terminal	Amino-terminal
OD	Optical density
PCR	Polymerase Chain Reaction
PDP	Precipitated dialysed protein
pI	Isoelectric point
pKa	Ionisability constant
p-NP	p-nitrophenol
p-NPG	4-nitrophenyl β-D-glucopyranoside
PP	Precipitated protein
PTS	Phoshotransferase system
PVDF	Polyvinylidene difluoride
Q-PCR	Quantitative PCR
QR-1	Quinone reductase-1
QTL	Quantitative trait loci
RDP	Ribosomal Database Project
RNA	Ribonucleic acid
ROS	Reactive oxygen species
rRNA	Ribosomal ribonucleic acid
SCFA	Short chain fatty acid
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Scanning Electron Microscopy
S-GT	S-glucosyltransferase
SNG	Sinigrin
SNP	Single-nucleotide polymorphism
SOC	Super Optimal broth with catabolite repression
ST	Sulfotransferase
TFP	Thiocyanate forming protein
TGGE	Temperature gradient gel electrophoresis
T_m	Melting temperature
T-RFLP	Terminal restriction fragment length polymorphism
UV	Ultraviolet
V_{max}	Maximum velocity

Abbreviation	Amino acid	Symbol
Ala	Alanine	A
Arg	Arginine	R
Asn	Asparagine	N
Asp	Aspartic acid	D
Cys	Cysteine	C
Gln	Glutamine	Q
Glu	Glutamic acid	E
His	Histidine	H
Ile	Isoleucine	I
Leu	Leucine	L
Lys	Lysine	K
Met	Methionine	M
Phe	Phenylalanine	F
Thr	Threonine	T
Trp	Tryptophan	W
Tyr	Tyrosine	Y
Val	Valine	V
Ser	Serine	S
Pro	Proline	P
Gly	Glycine	G

ACKNOWLEDGEMENTS

First of all, I would like to thank to my supervisors Prof. Arjan Narbad, Dr. Melinda Mayer, Prof. Richard Mithen and Dr. John T. Rossiter for their support, training and guidance throughout my PhD. I was lucky to benefit from broad expertise of my supervisory team. My sincere thanks go to Prof. Arjan Narbad and Dr. MeIinda Mayer for their patience during my writing up process and especially to Dr. Melinda Mayer for being there when I needed someone to talk to.

I owe a big thank-you to Dr. Shikha Saha and Dr. Neil Rigby for their help about HPLC and FPLC and to Dr. Vijitra Luang-In for providing the bacterial strains. I would like to acknowledge Dr. Arnoud Van Vliet, Dr. Fran Mulholland and Rachael Stanley for their contributions to this study. I would like to thank to Turkish Ministry of Education for funding my PhD education. In addition, my special thanks go to my colleagues and friends in IFR and Norwich who turned this PhD journey into a great experience to me.

I am very grateful to my family for their love and invaluable encouragement. This was a long adventure away from home but I learnt not only about science but also about myself and the whole world. I am planning to continue my career as an academic and to contribute science as much I can for the rest of my life. As our great leader Mustafa Kemal Atatürk said that “Our true mentor in life is science”.

CHAPTER ONE

1 GENERAL INTRODUCTION

1.1 GLUCOSINOLATES

Glucosinolates are found in cruciferous vegetables such as broccoli and are known to confer health promoting effects when consumed; notably, reduction in prostate, pancreatic and bladder cancer has been reported. This was attributed to chemopreventation effect of them and their degradation products. Plant myrosinases are responsible for transforming glucosinolates into the bioactive isothiocyanates (ITCs) that are known to have chemopreventative effects. During cooking, plant myrosinases are heat inactivated so ITC production is dependent on the myrosinase-like enzymes produced by the gut bacteria. This study is focused on glucosinolate metabolism by the human gut bacteria and identifying the bacterial enzymes and mechanisms that play a crucial role.

1.1.1 General Structure and Biological Importance of Glucosinolates

Glucosinolates are sulphur containing glycosides found in cruciferous vegetables mainly belonging to the *Brassicaceae* family. They are mostly present in the Brassicaceae, Capparaceae and Caricaceae families [1]. These include *Brassica oleracea* var *italica* (broccoli), *Raphanus sativus* (radish), *Brassica oleracea* var *capitata* (cabbage) or *Brassica oleracea* var *gemmifera* (Brussels sprouts), *Brassica oleracea* var *botrytis* (cauliflower) and *Brassica rapa* subsp rapa (turnip). Cabbage, broccoli and Brussels sprouts are the main source of glucosinolates for the human diet and they are frequently consumed vegetables in Western and Eastern countries [2].

The general structure of a glucosinolate is shown in Figure 1.1. Glucosinolates are β-thioglucosides and (*Z*)-*cis*-*N*-hydroximinosulfate esters with a side chain R and sulfur-linked d-glucopyranose moiety [2]. The chemical and biological nature of glucosinolates are defined by the side chain in the structure. So far glucosinolates are known to have little biological activity while their degradation products serve as a defence against herbivores including insects, birds, aphids and mammals [3]. Glucosinolate hydrolysis in plant is avoided by storing glucosinolate and the degradation enzyme in different plant compartments (such as glucosinolates in S-cells and myrosinases in myrosin cells)[4-7] or by separating them at subcellular level [8].

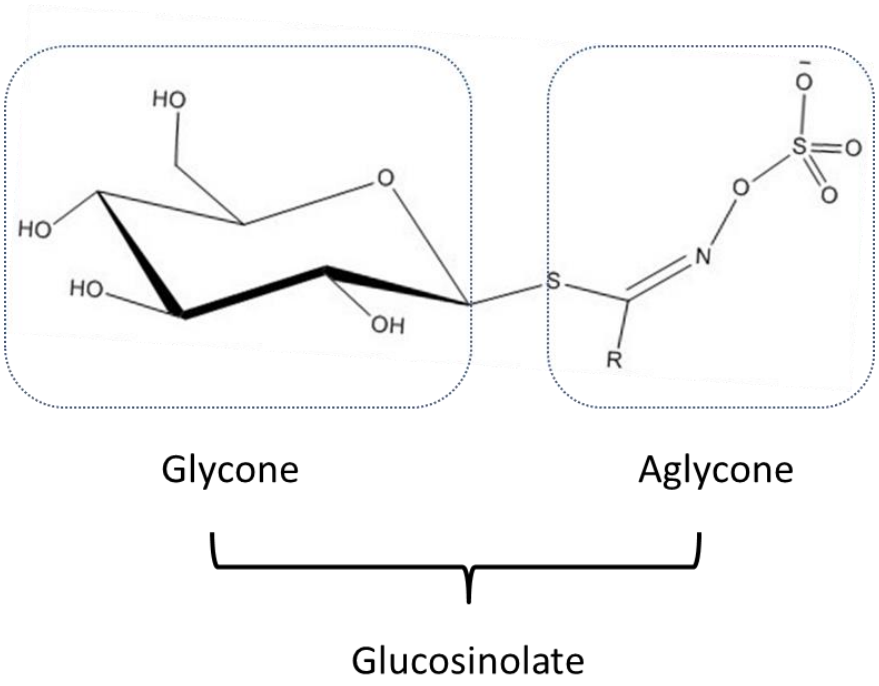


Figure 1.1 General structure of a glucosinolate. The image was adapted from [8]. R; variable side chain.

There are more than 120 different glucosinolates identified but only a few have been investigated deeply. They are known as sulphur and glucose contaning compounds and some of the benzyl glucosinolates may have an additional sugar moiety such as rhamnose and arabinose bound to aromatic ring [1]. Some of the glucosinolates found in the plants are given in Table 1.1. Amino acids are the precursors of glucosinolates. Depending on which amino acid acts as a precursor for glucosinolate formation, they can be divided into three groups. Aliphatic glucosinolates are derived from alanine (Ala), leucine (Leu), isoleucine (Ile), valine (Val), and methionine (Met). Aromatic glucosinolates are derived from phenylalanine (Phe) and tyrosine (Tyr) and indole glucosinolates are derived from tryptophan (Trp) [9]. As an example, methionine is the precursor in glucoraphanin biosynthesis.

Common Name	Structure of the Glucosinolate
Aliphatic Glucosinolates	
Sinigrin	2-Propenyl
Glucoraphanin	4-methylsulfinylbutyl
Glucoraphenin	4-methylsulfinyl-3-butenyl
Glucoiberin	3-methylsulfinylpropyl
Glucoiberverin	3-methylthiopropyl
Glucoerucin	4-methylthiobutyl
Progoitrin	2-hydroxy-3-butenyl
Glucoalyssin	5-methylsulfinylpentyl
Glucocapparin	Methyl
Glucobrassicanapin	4-pentenyl
Gluconapin	3-Butenyl
Aromatic Glucosinolates	
Glucotropaeolin	Benzyl
Glucosinalbin	p-hydroxybenzyl
Gluconasturtiin	2-phenethyl
IndoleGlucosinolates	
Glucobrassicin	3-Indolylmethyl
Neobrassicin	1-methoxy-3-indolylmethyl
4-Hydroxyglucobrassicin	4-hydroxy-3-indolylmethyl
4-Methoxyglucobrassicin	4-methoxy-3-indolylmethyl

Table 1.1 Some of the glucosinolates found in Brassica vegetables. Adapted from [1] and [2]).

Glucosinolates are found in almost every part of the plant varying in concentration throughout the plant development. Their content in plant is quite variable but they make up almost 1% of the dry weight in some tissues of Brassica [1]. Normally, up to 4 different glucosinolates dominate the glucosinolate occurance in the plant within a single species [2]. Seeds or young sprouts of broccoli (*Brassica oleracea* var *italica*) were reported to contain 70 - 100 µmol of total glucosinolates per g fresh weight and it was found that the content was entirely dominated by aliphatic glucosinolates; glucoraphanin, glucoerucin and glucoiberin [10]. The same cultivar was reported to have around 1 - 4 µmol glucosinolates per g fresh weight at late stages of growth possessing aliphatic and indolic glucosinolates at equal levels [10]. The concentration, variation and distribution of glucosinolates in the Brassica family depends on a number of criteria including the species [11], varieties [12], season [13] environmental conditions such as soil fertility [14], presence of pest infestation [2], plant organ [11, 15], plant age [10, 15] or abiotic stress factors such as salinity or drought [16]. The glucosinolate accumulation of the model plant *Arabidopsis thaliana* among different organs and developmental stages was studied and it was found that dormant and germinating seeds had highest amount (2.5 - 3.3% per dry weight) followed by the fruits, leaves and the roots. In the same study, younger leaves were reported to have higher glucosinolate concentrations than the older leaves [15].

Glucosinolates and their degradation products are responsible for the characteristic flavour and odour of *Brassica* vegetables. When cell-integrity is damaged, ITCs are responsible for the pungent and bitter flavour [17]. ITCs from progoitrin and gluconasturtiin have been found to relate bitterness in cabbage and Brussel sprouts.

Glucosinolates and their degradation products have also gained attention as pest control agents. Many studies are focused on their biofumigant properties [18-20]. Even if glucosinolate-myrosinase system is a defence mechanims for plants [21], herbivores are able to avoid from this by biochemical mechanisms or creating their own myrosinase-glucosinolate system against their predators [22, 23]. For instance, the cabbage aphids (*Brevicoryne brassicae*) can do this by producing their own myrosinases and storing glucosinolates obtained from plants [23].

Aqueous solubility, ionisability (p*K*a) and lipophilicity (octanol-water partition, log P) determine the dissolution of glucosinolates. Especially, log P is an important factor which

affects membrane permeability and inversely related to solubility. In other words, higher log P values mean enhanced permeability but reduced solubility. Glucosinolates have negative log P values and it is unlikely that they can cross cell membranes so active transport through aqueous pores is needed [24]. Glucosinolates have strongly acidic properties due to their sulfate group. Sulfate group and thioglucose moiety gives them water-solubility [24]. The potential roles of glucosinolates in plants can be listed as follows:

1) Glucosinolates have roles in plant growth regulation. The plant hormone indole-3-acetic acid (IAA) was proposed to arise from indole glucosinolates. Indole glucosinolates are proposed to be hydrolyzed to indole acetonitriles then nitrilase can convert indole acetonitriles to IAAs [25, 26]. Later, indole-3-acetaldoxime was found as a precursor of IAA. *Arabidopsis* mutants (the genes involved in late steps of biosynthesis of indole glucosinolates were mutated) were found to have high levels of IAA and resulted in dwarf phenotype. These results showed that disruption of the conversion of indole-3-acetaldoxime to indole glucosinolates resulted in increased flux into IAA [27, 28].

2) Glucosinolates are involved in plant-insect/herbivore interactions. They are known as part of defensive system against their predators. Plant feeding insects can be categorised as generalists and specialists. Glucosinolates may be a general poison for generalists and a feeding attractant for specialists [29]. They show growth inhibition or feeding deterrence against general herbivores such as birds, slugs and generalist insects [29].

3) Glucosinolates have roles in plant/pathogen interactions. It was reported that glucosinolates and their hydrolysis products have antifungal and antibacterial properties *in vitro* [30-33]. The hydrolysis products of aromatic glucosinolates (aromatic ITCs) were found to be more toxic than aliphatic ones and aliphatic ITCs showed decreased toxicity due to increasing length of side chain [34, 35]. As an example, 4-methylsulfinylbutyl ITC was identified with a wide range of antimicrobial activity against fungi and bacteria, it caused 50% growth inhibition at concentration of 28 μM for the most sensitive organism tested, *Pseudomonas syringae* [36].

1.1.2 Glucosinolate Biosynthesis

Amino acids were suggested to be precursors of the aglycone motiety of glucosinolates then this hypothesis was tested and stages of biosynthesis of glucosinolates were identified using labelled compounds (^{14}C, ^{15}N, ^{35}S) [37-39]. The biosynthesis of glucosinolates occurs in three steps. In the first step, chain elongation of aliphatic or aromatic amino acids occurs by inserting methylene groups into their side chains (Figure 1.2). The second step is metabolic modification of the amino acid via an aldoxime intermediate then glucosinolate core structure is formed. The pathway for biosynthesis of glucosinolate core structure is shown in Figure 1.3. Lastly, the initial glucosinolate is modified by many secondary transformations. For instance, the methylthioalkyl glucosinolates are produced directly from the unmodified elongated methionine derivatives [2, 8, 40]. The detailed explanation of biosynthesis steps is given as follows:

Initially, the parent amino acid is deaminated branched-chain aminotransferase4 (BCAT4) to form 2-oxo acid. The 2-oxo acid enters a cycle of 3 transformations. First, it condenses with acetyl-coA in a reaction catalysed by methylthioalkylmalate (MAM) and yields a 2-malate derivative (1). The 2-malate derivative isomerased by isopropyl isomerase (IPMI) to form 3-malate derivative (2). This derivative undergoes oxidation-decarboxylation by the action of isopropylmalate dehydrogenase (IPM-DH) to form an elongated 2-oxo acid which has one more methylene group than the starting 2-oxo acid (3). Then elongated 2-oxo acid can undergo deamination to form homomethionine or enter another round of chain elongation [8, 41] (See Figure 1.2).

The chain elongated amino acids are converted to aldoximes by cytochromes P450 belonging to the CYP79 family. CYP79F1 can metabolise all chain-elongated methionine derivatives, CYP79F2 only converts long-chained methionine derivates. While CYP79B2 and CYP79B3 can use tryptophan, CYP79A2 metabolises phenylalanine. After aldoxime formation, aldoximes are oxidized to either nitrile oxides or aci-nitro compounds by cytochromes P450 belonging to the CYP83 family. CYP83A1 converts aliphatic aldoximes and CYP83B1 converts both tryptophan and phenylalanine derived acetaldoximes. Then conjugation with a sulfur donor produces *S*-alkyl-thiohydroximates. This step is proposed to be catalysed by glutathione-S-transferases [41] and the sulfur donor was found to be glutathione [42]. The *C*-*S*-lyase converts *S*-alkyl-thiohydroximates to thiohydroximates. Thiohydroximates undergo glycosylation by the action of glucosyltransferases of the UGT74 family and forms

desulfoglucosinolates then desulfoglucosinolates are sulfated by sulfotransferases to produce the core glucosinolate structure [8, 41] (See Figure 1.3).

Secondary modifications of R group including oxidation, desaturation, hydroxylation, removal of a methylsulfinyl group and addition of methoxy group can occur to form final glucosinolate such as glucoiberin, sinigrin or 4-methoxyglucobrassicin [8, 41].

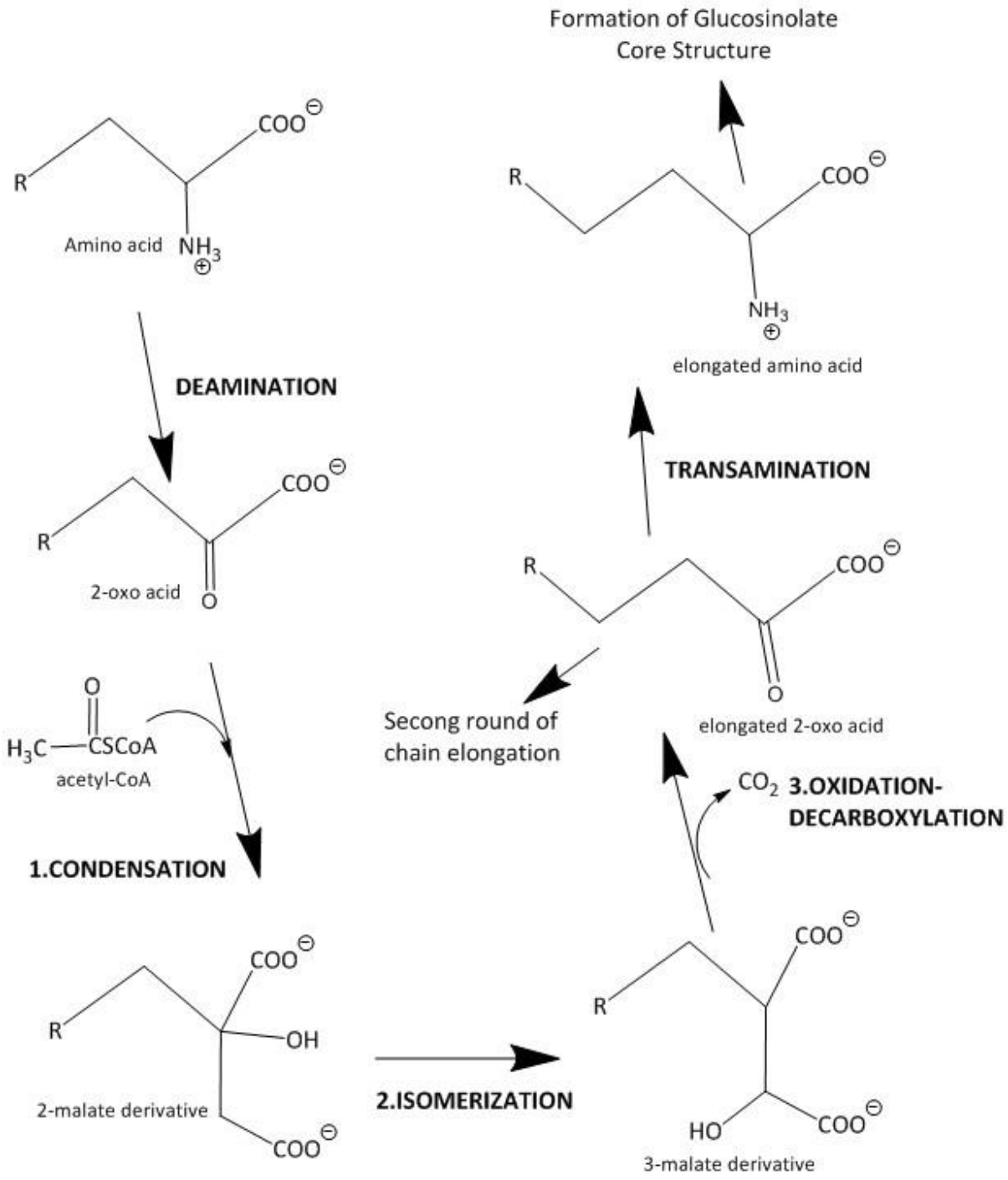


Figure 1.2 Amino acid chain elongation cycle for glucosinolate biosynthesis. Adapted from [8]. Three mains steps include condensation with acetyl-CoA (1), isomerization (2) and oxidation-decarboxylation (3).

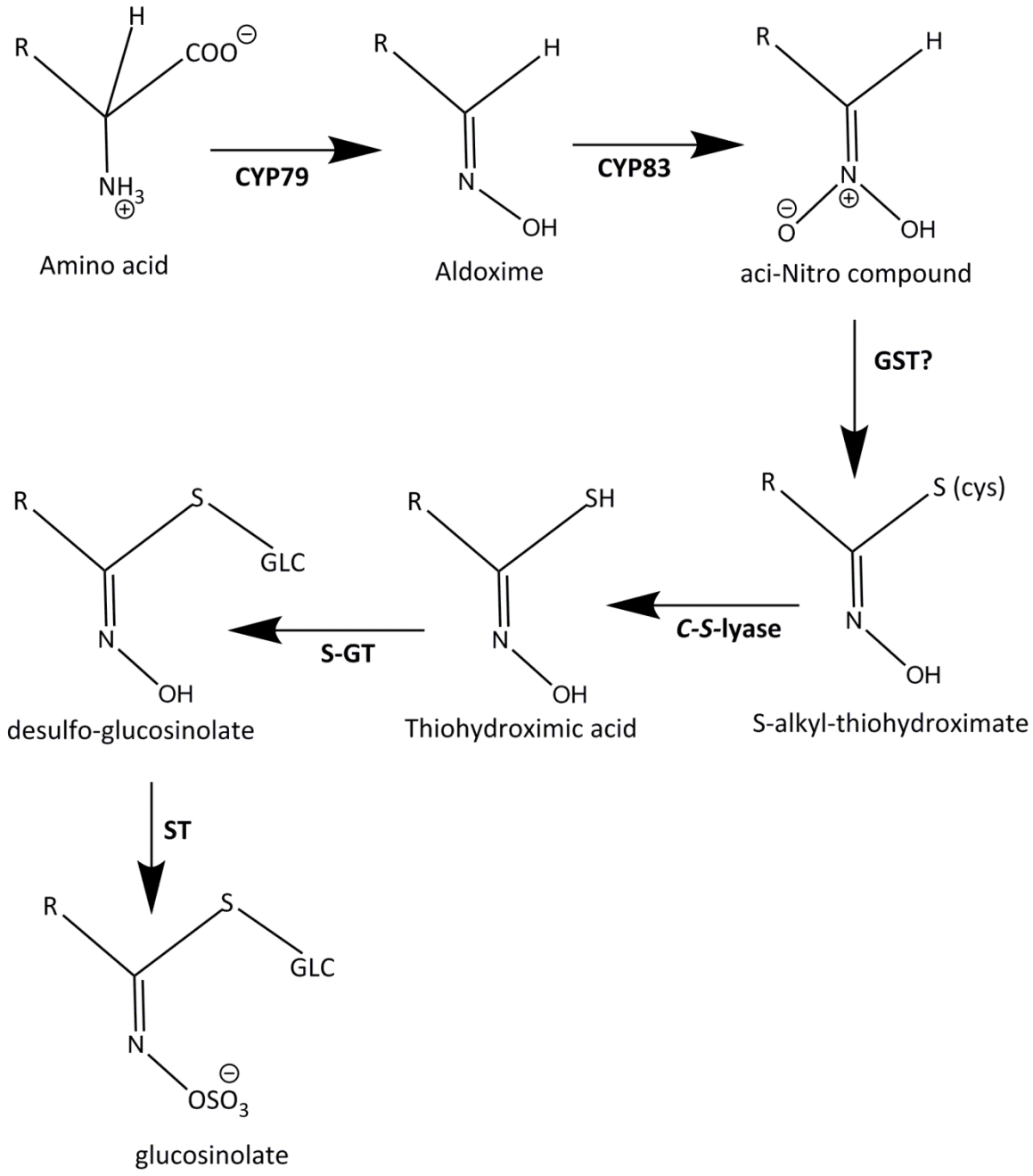


Figure 1.3 Biosynthesis of glucosinolate core structure (adapted from [8]). CYP79 enzymes are responsible for conversion of elongated amino acids to aldoximes then CYP83 enzymes act on aldoximes to form aci-Nitro compounds. Glutathione-S-transferase (GST) was proposed to be the sulfur-donating enzyme. R; variable side chain, S-GT; S-glucosyltransferase, ST; sulfotransferase.

There are studies focusing on increasing the glucosinolate content of plant by manipulating the genetic pathways [43, 44]. Increasing the glucosinolate content to maximize the health benefits of glucosinolates is a target for plant breeders [45]. For instance, *Brassica villosa* wild species were used to develop 3 high-glucoraphanin broccoli hybrids which have 2.5-3 times

the glucoraphanin content of standard hybrids. Two of these high-glucoraphanin hybrids were commercialised as Beneforté® broccoli [44].

Different thermal and non-thermal processes also can affect the glucosinolate content [46, 47]. A study investigated the effect of gamma-irradiation on genes involved in glucosinolate biosynthesis pathway and it was found that many genes are up-regulated and sinigrin content was increased by 41% in the irradiated cabbage [46].

Epidemiological evidence has shown that there is an inverse association between the consumption of cruciferous vegetables and the risk of cancer. The intake of cruciferous vegetables has been linked to a reduced risk of certain cancer types such as lung cancer, colorectal cancer, prostate cancer in many epidemiological studies [48-50]. This effect is attributed to the ITCs, the hydrolysis products of glucosinolates. Glucosinolate research started with toxicological aspects of glucosinolates such as progoitrin and methods for removing them from dietary sources and animal feed but nowadays, most of the studies are focused on their potential health-promoting effects [1, 2].

1.2 GLUCOSINOLATE HYDROLYSIS AND DEGRADATION PRODUCTS

The hydrolysis of glucosinolates is mediated by the thioglucosidase enzyme myrosinase (Figure 1.4). Glucosinolate containing vegetables also contain myrosinase in all parts of the plant including leaves, roots, stems, seeds and flowers [51]. Myrosinase is found separately from glucosinolates, in idioblasts or myrosin cells of the parenchymatous tissue. Cutting, chewing or food processing causes tissue damage so the integrity of cells is disrupted and glucosinolates and myrosinase come into contact. When glucosinolate degradation is initiated and enzymatic hydrolysis results in the formation of glucosinolate hydrolysis products such as ITCs, thiocyanates or nitriles [24, 51] (Table 1.2).

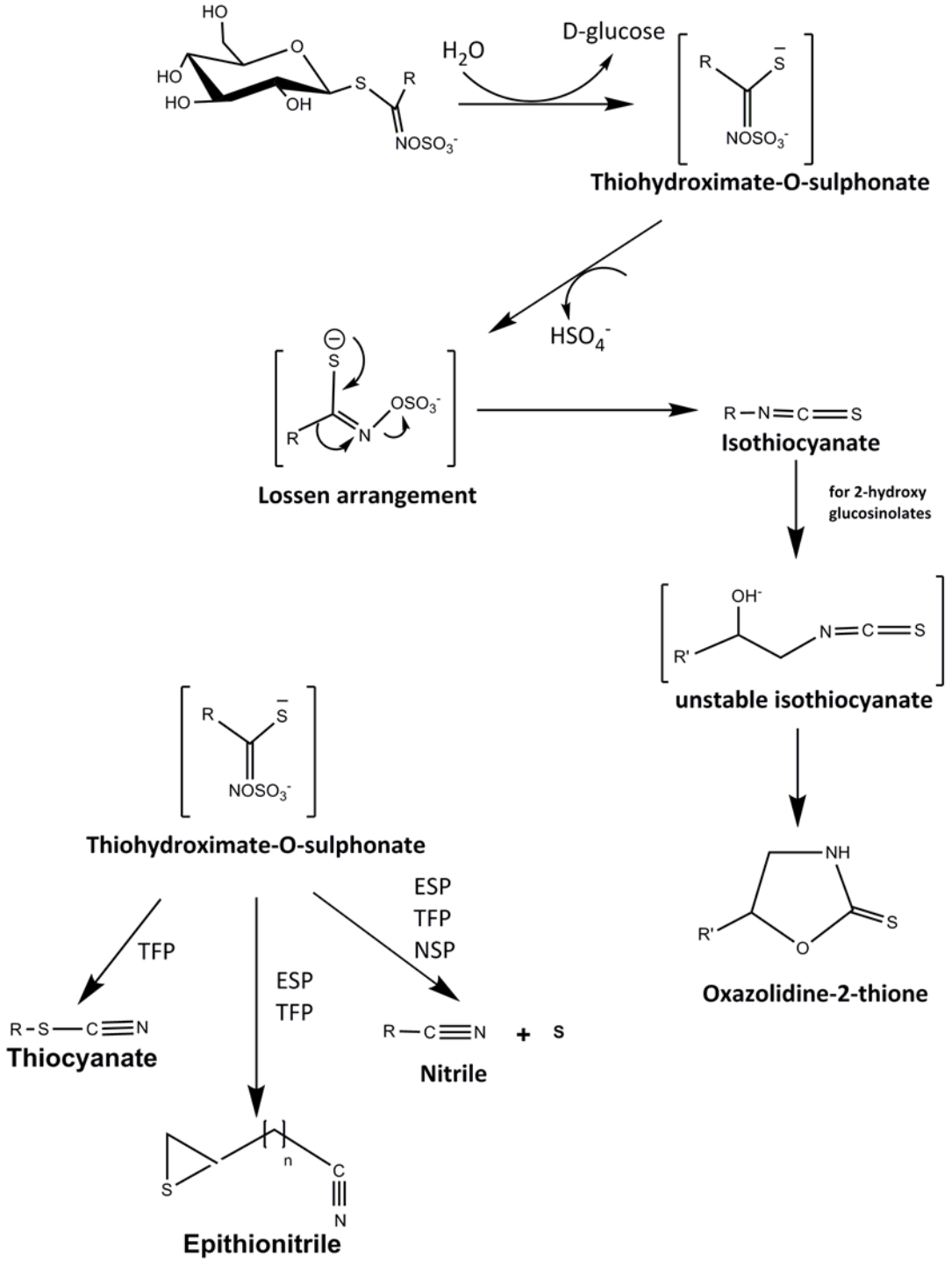


Figure 1.4 Hydrolysis of glucosinolates by myrosinase. First, myrosinase catalysed degradation produces glucose and unstable aglycone. The aglycone rapidly rearranges to ITC (a). In the presence of specifier proteins (ESP, TFP and NSP), other degradation products are formed (b). ESP; epithiospecifier protein, TFP; thiocyanate forming protein, NSP; nitrile specifier protein. The presence of a terminal double bond in the side chain is needed for epithionitrile formation. Thiocyanate formation has only been reported from sinigrin, glucoerucin and glucotropaeolin hydrolysis. The image was adapted from [52].

Glucosinolates are converted to an unstable aglycone (thiohydroximate-*O*-sulfonate) by myrosinase and undergo a 'Lossen' re-arrangement to give an ITC. Thiocyanates, epithionitriles and nitriles formation is dependant on pH, metal ions and specifier proteins [2, 53, 54].

The hydrolysis products of glucosinolates can be unstable. For instance, β-Hydroxy-ITCs are unstable and undergo rapid cyclisation to give oxazolidine-2-thiones [2]. Some of the ITC and nitrile products of glucosinolates are shown in Table 1.2.

Glucosinolate	ITC Product	Nitrile Product
Sinigrin	Allyl ITC	Allyl nitrile
Glucoerucin	Erucin	Erucin nitrile
Glucoiberin	Iberin	Iberin nitrile
Glucoraphanin	Sulforaphane	Sulforaphane nitrile
Glucotropaeolin	Benzyl ITC	Benzyl nitrile
Gluconasturtiin	Phenethyl ITC	Phenethyl nitrile

Table 1.2 Some glucosinolate hydrolysis products catalyzed by plant myrosinase

1.2.1 ITC Formation

ITCs are important and valued hydrolysis products of glucosinolates owing to their chemopreventative effects. They are known to be produced at pH range of 5 – 7 and nitriles are known as degradation products in acidic conditions [2, 55]. A recent study showed that ITC and nitrile production can occur in a more broader pH range of 3.7-7.6 during bacterial fermentation [56] but it was not reported how much ITC formed at specific pH. Unlike glucosinolates, ITCs were reported to be unstable and easily degraded in media, buffers and water [57].

1.2.2 Nitrile and Epithionitrile Formation

Nitrile production was originally thought to be dependent on ferrous ions, pH and enzyme independent. Then epithiospecifier protein (ESP) was identified and partially characterised as the responsiple factor for epithionitrile formation in Crambe abyssinica seeds [58]. Later, ESP was purified independently by two diferent study groups and characterised [59, 60]. Moreover, thiocyanate forming proteins (TFP) were also identified to form nitrile and epithionitriles from aliphatic glucosinolates [61].

A new group of specifier proteins, nitrile-specifier, was identified and reported to produce nitriles [62]. The nitrile specifier proteins (NSP) were reported to require the presence of a myrosinase for its nitrile specifier activity [62]. When bioactivity of sulforaphane and sulforaphane nitrile was studied using mouse and rats hepatoma cells, sulforaphane was found to be more bioactive to induce phase II enzymes than sulforaphane nitrile [63].

In one study, fresh broccoli florets or broccoli sprouts were used to examine the effects of heating on sulforaphane and sulforaphane nitrile formation. It was concluded that heating fresh broccoli florets and broccoli sprouts to 60 °C decreased sulforaphane nitrile formation while increasing the sulforaphane formation. It was suggested that this was the effect of different thermal labilities of myrosinase and ESP. While myrosinase is respectively heat stable, ESP is not [64].

1.2.3 Thiocyanate Formation

Thiocyanates are also degradation products of glucosinolates. After breakdown of β-thioglucosidic bond, thiohydroxymate-O-sulfonate can be transformed into thiocyanates in presence of thiocyanate forming proteins (TFP). Thiocyanates are normally found in extracellular mammalian fluids such as saliva, nasal lining fluid, milk, tears and thought to be originated from diet [65]. For the formation of thiocyanates, two factors are suggested to be important; a myrosinase and TFP. The thiocyanate forming enzyme were identified and well studied [61, 66, 67]. Thiocyanate formation has only been reported from sinigrin, glucoerucin and glucotropaeolin hydrolysis [52]. There is no evidence for bacterial or aphid originated thiocyanate forming enzyme so far.

Thiocyanates were suggested to be useful therapeutic agents as possesing antioxidant properties. It was found that they can protect cells against oxidizing agents like hypochlorous

acid [65]. However, thiocyanates are reported to inhibit the uptake of iodide by the thyroid [68].

1.2.4 Other Products

β-hydroxy ITCs are reported to be unstable and cyclising to oxazolidine-2-thiones that can cause goiter [8]. Indole glucosinolates produce unstable ITCs and undergo on further hydrolysis to form 3-indolemethanol, 3-indoleacetonitrile or 3,3`-diindolylmethane then give dimers, trimers or tetramers [2, 24]. The unstable ITC, 3-(thiocyanatomethyl)-1H indole, is thought to be the origin of 3-indolemethanol (Indole-3-carbinol) which go through solvolysis with water [53].

1.3 THE IMPORTANCE OF GLUCOSINOLATES AND THEIR DEGRADATION PRODUCTS

1.3.1 Chemopreventative effects of Glucosinolates and ITCs

Diets that are rich in fruits and vegetables are reported to be associated with reduced risk of many cancer types in many studies. Cruciferous vegetables are becoming known for their health promoting effects against various cancers [50, 69-72].

Cancer is known as unrestricted division and proliferation of cells. It is a global burden on human health. According to the International Agency for Research on Cancer (IARC), there were 14.1 million new cancer cases and 8.2 million cancer deaths worldwide in 2012. By 2030, the cancer incidents worldwide are estimated to rise up to 22 million (http://www.who.int/mediacentre/factsheets/fs297/en/). Cancer may be caused by internal conditions such as genetic mutations, hormones, immune conditions or external conditions such as tobacco, infectious organisms and unhealthy diet. The chemopreventation of cancers by bioactive compounds in the diet is one of the ways to combat cancer.

Cancer mechanism is very complicated and many theories exist about its cause [73]. According to the mutation theory, it is not raised because of a single mutation. For instance, several hundred human genes were reported to drive the unrestricted cell division and growth process which results in pancreatic cancers [74]. Genetic or epigenetic factors induce cancer by changing pathways that control cell proliferation, apoptosis and differentiation. There are two types of enzymes which are effective on this process, phase I and phase II enzymes. Phase I enzymes primarily consist of cytochrome P450 enzymes and act as bioactivators for chemical carcinogens to induce carcinogenesis. Cytochrome P450 is a large

family of transferases and located in endoplasmic reticulum and mitochondria of the cells [75]. On the other hand, phase II enzymes such as quinone reductase-1 , glutathione-S-transferase (GST), thioredoxin reductase, UDP-glucuronosyltransferases and γ-glutamylcysteine synthetases act as protectors for cells/tissues against carcinogens [2, 76-78]. They may perform many reactions including glucuronidation, sulfation, methylation, acetylation, glutathione and amino acid conjugation. Generally, they do create conjugates more hydrophilic than the parental compounds. They can locate in different compartments of the cells for example UDP-glucuronosyltransferases are membrane bound proteins, whereas soluble GSTs are located mainly in cytoplasm but can be found in nucleus, mitochondria and peroxisomes as well [79].

ITCs derived from glucosinolates are able to modulate Phase I and Phase II enzymes, they can directly inhibit or down-regulate Phase I enzymes. They can reduce carcinogen activation by inhibiting P450 monooxygenases and activate nuclear factor (erythroid-derived 2)-like 2-Antioxidant Response Element (Nrf2-ARE) signalling pathway to induce the detoxifying enzymes (Phase II enzymes). Carcinogen detoxifying enzymes in human cells such as quinone reductase-1 (QR-1) are induced at transcriptional level and ARE were suggested to mediate the process [80].

ITCs were reported to activate caspase-3, -8, or -12 to increase apoptosis [80]. ITCs have the ability to induce cell cycle arrest and they are able to inhibit cell-proliferation and invasive metastasis [81]. It was reported that most of the ITCs are absorbed by passive diffusion by the cells [78, 82, 83] and intracellular concentrations can be up to 100 - 200 times higher than extracellular concentrations [78].

1.3.2 Protective Effects of Glucosinolates and ITCs against Diseases

In addition to their chemopreventive effects against cancer, glucosinolates were studied for their protection of cardiovascular and central nervous system. Rat model study fed hypertensive stroke-prone rats with sulforaphane by gavage and concluded that sulforaphane rectified pathological abnormalities in spontenously hypertensive stroke-prone rats and significantly improved blood pressure [84]. Moreover, consumption of high glucoraphanin broccoli was associated with a reduction in plasma low density lipoprotein cholesterol (LDL-C) [85]. Similar to the cardiovascular system, sulforaphane was determined to have positive effects on the central nervous system. Administration of sulforaphane was found to decrease infarct ratio [86], brain edema [87, 88], cortical apoptosis [88]. These effects were mostly attributed to activation of Nrf2 and upregulation of its target genes [89].

ITCs have antioxidant, immunostimulatory, anti-inflammatory, antiviral and antibacterial properties [81]. ITCs have a electrophilic nature and may undergo reactions with N- O- or S-based nucleophiles. The –N=C=S group of ITCs can show direct reaction with the cysteine sulfhydryl groups of glutathione [82] and proteins [90].

Different ITCs have been tested for their potential bioactivity so far. For instance, Benzyl ITC was suggested to have anticancer effects by inhibiting initiation, growth and metastasis of human cancers in mouse models. The targets of benzyl ITC in the inhibition of pancreatic tumour growth are overexpressed proteins such as protein kinase B (AKT), signal transducers and activators of transcription 3 and nuclear factor kappa B (NFκB) [78, 91]. Phenethyl ITC has been identified for its anticancer activity and function in cell-cycle arrest and apoptosis induction [78].

Sulforaphane, derived from glucoraphanin, is the best characterised ITC due to its cancer preventive and therapeutic properties [2, 76, 78, 92]. Sulforaphane shows its chemopreventative ability by inhibiting phase I enzymes and inducing phase II enzymes. Sulforaphane targets certain molecules like survivin and NFκB that are essential for cancer cell survival [93]. It was also reported that sulforaphane inhibits important proteins for breast cancer cells like estrogen receptor, epidermal growth factor receptor 1 (EGFR1) and human epidermal growth factor receptor 2 (HER2) [94] and a study showed that sulforaphane has efficacy against tumor growth and metastasis of breast cancer in vivo [95].

Allyl ITC (3-isothiocyanato-1-propene or 2-propenyl ITC) was studied for its effects on health and was shown to have an inhibitory effect on bladder cancer development and progression in rats [96].

Erucin, an enzymatic hydrolysis product of glucoerucin, was suggested as a potential therapeutic agent against prostate cancer cells [97]. Erucin was investigated for its inducing capacity on heme oxygenase gene expression *in vitro* (in cultured cells) and *in vivo* (in mice). The study reported significant increase in mRNA and protein levels of heme oxygenase when HT-29 cells were treated with erucin [98].

In addition to ITCs, glucosinolates were also studied for their health promoting effects. A study reported the potential neuroprotective role of glucoraphanin on Parkinson's disease using a mouse model. Two different sets of experiment were used, an acute (2 injections, 40 mg/kg1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine) and a sub-acute (5 injections, 20 mg/kg 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine) model of Parkinson disease. The study revealed that glucoraphanin is able to reduce dopamine transporter degradation, tyrosine hydroxylase expression, interleukin-1-beta release, the triggering of neuronal apoptotic death pathway and the production of radical species; results focused on nitrotyrosine, Nrf2 and glial fibrillary acidic protein immunolocalization and concluded that glucoraphanin can protect neurons against neurotoxicity [99].

1.3.3 Toxicity of Glucosinolates and Their Degradation Products

Historically glucosinolates were studied for their toxicological effects. Many Brassica vegetables are used in animal feed and the main goal of studies was to remove glucosinolates from food and animal feed because of decreased palatability effect of glucosinolates [2]. Sinigrin and progoitrin are associated with the bitter taste. Especially degradation product of progoitrin, goitrin, is known as a very bitter product [100]. Goitrin can easily be nitrosated if it in contact with nitrites and form the mutagenic compound N-nitroso-oxazolidine [101]. Degradation products of glucosinolates were reported to be prone to disrupt iodine bioavailability resulting in goitrogenic effects [100, 102, 103].

Some of the ITCs such as allyl ITC, benzyl ITC and phenethyl ITC were reported to have genotoxic effects. These isothiocyanates have electrophilic reactivity and can form adducts with DNA. This can lead inducing gene mutations and chromosomal aberrations. It was also noted that the doses used in animal studies are several times higher than normal dietary

exposure doses. If there is a potential health risk about ITCs due to their genotoxicity, it should not be excluded. Therefore, further *in vivo* studies using the normal dietary exposure doses are needed to confirm and verify the *in vitro* findings [104].

1.4 BIOAVAILABILITY OF GLUCOSINOLATES AND THEIR DEGRADATION PRODUCTS

The bioavailability term consists of several sequential steps. It includes the availability for absorption, metabolism, tissue distribution, and bioactivity. The bioactivity measures the biologic activity of components on specific organs or tissues. The bioavailability and bioaccessibility terms can often be confused but the bioaccessibility term only refers to the proportion that is being available to be absorbed from intestinal tract. However, the bioavailability term includes transport of components to cells and bioactivity term as well [105] (See Table 1.3). It is not easy to determine the bioactivity so bioavailability term is usually used in a narrow sense. It is referred as the part that enters the blood stream from a consumed dose of a nutrient or the metabolites [105]. For instance, the importance of the mercapturic acid pathway in the metabolism of ITCs was reported and ITC mercapturic acids in urine were used as a biomarker to study ITC bioavailability [106, 107].

BIOAVAILABILITY	
BIOACCESSIBILTY	BIOACTIVITY
• Events that take place during food digestion	• Transport and assimilation by the target tissue
• Absorption/assimilation through epithelial tissue	• Interactions with biomolecules
• Pre-systemic metabolism	• Post-absorptive metabolism or biotransformation
	• Physiological response

Table 1.3 Bioaccessibility and bioactivity steps . Adapted from [105]).

Following this terminology, the glucosinolate content of the raw or processed cruciferous vegetables is the first step for bioavailability. Food processing has effects on glucosinolate content of the Brassica vegetables [108]. For instance, the effect of boiling, low pressure steaming and microwaving on total glucosinolate content was studied and it was reported

that low pressure steaming and microwaving was responsible for a lower glucosinolate loss than boiling [109]. Partially or complete inactivaction of enzymes involved in glucosinolate degradation by heating affect the final product formation. Myrosinase and ESP proteins were reported to show different thermal labilities. While mild heating (60 -70°C) was enough to inactivate ESP, myrosinase activity can be retained and favour the ITC formation [64].

After food processing, mastication plays a key role in the breakdown of the plant cell wall and enables the myrosinase to come in contact with glucosinolates. It was also reported that differences in chewing might cause interindiviual differences in the bioavailability of ITCs [107]. No additional myrosinase activity was found in the saliva so degradation of glucosinolates depends on the myrosinase in the plant. During digestion, four main processes could happen; acidic hydrolysis of glucosinolates (1), unspecific binding to biomolecules (2) and hydrolysis by plant myrosinase (3) and microbial breakdown by gut bacteria (4) [24]. The glucosinolate content from rapeseed meal and some glucosinolate degradation products were investigated during *in vitro* incubation with pepsin-HCl or with contents of porcine small intestine and caecum [110]. It was found that individual glucosinolates were differently affected from peptic and intestinal conditions. Under peptic conditions, the loss was between 3-23% and it was 7-28% under intestinal conditions [110]. The degradation of glucosinolates by the acidic environment in the stomach may favour nitrile formation over ITC formation. Maskell et al. (1994) failed to identify ITCs or goitrin in their samples. This phenomenon might have occured because of extensive ITC binding [110]. Non-specific binding of glucosinolates to dietary fibres was reported by Michaelson et al. (1994) [111]. The microbial breakdown of glucosinolates will be discussed in later section 1.9.

Absorption of the glucosinolates and their hydrolysis products occur after completion of digestion. Intact glucosinolates were suggested to pass through gut epithelium without hydrolysis [112]. As intact glucosinolates are hydrophilic compounds with negative log P values, they are supposed to be absorbed by active transport or paracellular route (through tight junctions) [24]. However, Michaelsen et al. (1994) proposed that glucosinolates can pass cell membranes by forming ion pairs with lipophilic cations in the matrix [111]. Owing to a relatively lipophilic nature and low molecular weight, degradation products of glucosinolates have potential to be absorbed into cell mebranes by passive diffusion [24]. ITCs can rapidly form conjugates with cellular glutathione (GSH) and this conjugation is the diriving force of accumulation of ITC within the cells and passive diffusion [24]. The excretion of ITCs was

revealed to be closely related with the non-enzymatic ITC-GSH conjugation reaction rate and cellular GSH levels [113]. It was also found that glutathione-S-transferase promoted the conjugation reaction [113]. There is limited data about the absorption of other degradation products [24]. In a study, epithionitrile 1-Cyano-2,3-epithiopropane was rapidly taken up by experimental animals showing the peak values after 1 h [114].

Transport of glucosinolates or degradation products involves the transfer of compounds between the blood and the body tissues. Transport of these compounds might be affected by blood flow, membrane barriers, tissue affinity, ion trapping and plasma protein binding (albumin and glycoproteins) [24]. The same mechanism for intestinal absorption applies to the assimilation by the target tissue in general [24]. Following the absorption into the intestinal epithelium, ITCs were reported to pass into the systemic circulation and metabolised by the mercapturic acid pathway [106]. The liver and kidney are key organs in ITC post-absorptive metabolism where ITCs are conjugated with glutathione in the liver then with N-acetyl cysteine (NAC) in the kidney [115].

The distribution of ITC metabolites in tissues was studied by Briecker et al. (2014) using a mice model. The mice were fed with broccoli sprout powders or pure sulforaphane for a week then ITC metabolite profile in different tissues were determined. ITC metabolites accumulated in specific sites especially bladder followed by the liver and kidney. Lower concentrations were detected in plasma, skin and lung tissues, these concentrations were almost 100-200 times less compared to their levels in the bladder. In addition, interconversion between sulforaphane and erucin was determined *in vivo* [115]. A study reported that complete inactivation of myrosinase in broccoli products resulted in 10% and 29% of sulforaphane and iberin recovery in the urine respectively which indicates the role of human gut microbiota in glucosinolate metabolism [116]. The interaction with biomolecules and physicological response were discussed in section 1.3. The metabolic fate of glucosinolates and ITCs in the human body is illustrated in Figure 1.5.

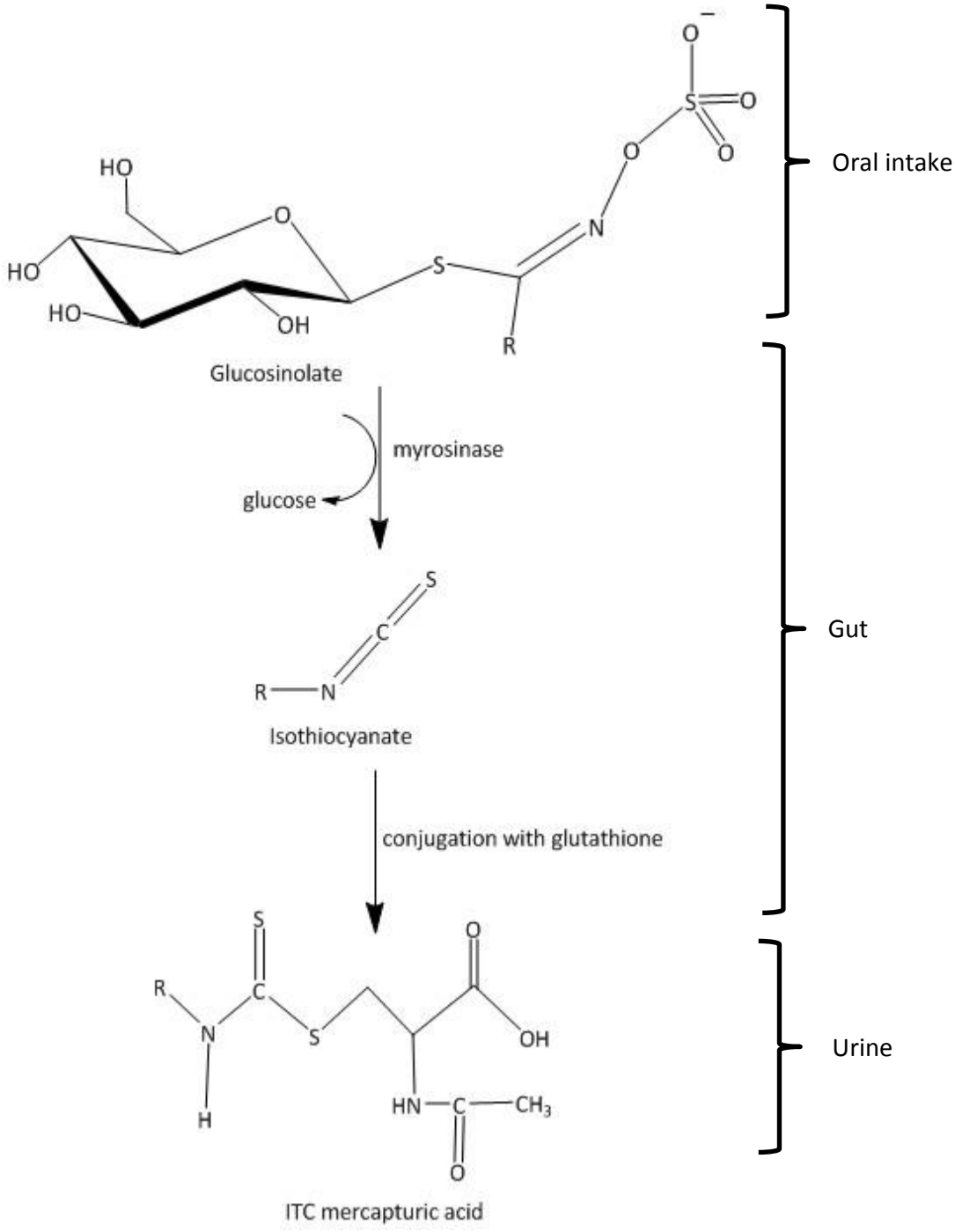


Figure 1.5 Expected metabolic fate of glucosinolates in the human body. Oral intake: During mastication, ITC formation can also occur. Gut: Glucosinolates are hydrolysed into ITCs then perform conjugation with glutathione and ITC-glutathione conjugate can pass through cell membrane. Urine: ITC is metabolised by the mercapturic acid pathway and excreted with the urine. Figure was adapted from [106].

To maximize the health benefits of ITCs, bioavailability of glucosinolates and ITCs should be fully adressed. In this context, the human gut bacteria involved in glucosinolate metabolism and the enzymes responsiple for this metabolism in the gut have a key role.

1.5 MYROSINASES

Myrosinase (β-thioglucoside hydrolase, EC 3.2.3.147) is an enzyme responsible for hydrolysis of glucosinolates [117, 118]. Plants that contain glucosinolates have also myrosinase stored separately from their substrate glucosinolates [4-8]. The plant myrosinases belong to Glycosyl Hydrolase 1 (GH1) family. They are well studied and characterised [117, 119-121].

The β-glucosidases (EC 3.2.1.21) catalyses the breakdown of the glucosidic bond of a variety of saccharides and glucosides. They are classified into different Glycosyl Hydrolase families such as GH1, GH3, GH9 or GH30 based on amino acid sequence and structural similarity [122]. Mostly, there is a catalytic triad motif located at active sites of glucosidases. Two carboxylic residues act as acid/base and as nucleophile for double displacement at the anomeric center. Generally, O-β-glucosidases have an acid/base catalyst like glutamate at their active sites. However, plant myrosinases do not contain glutamate, instead they have a glutamine at this position. Once the glycosyl-myrosinase intermediate is formed, this glutamine residue enables placement of the water molecule at correct position, without deprotonation. This correct placement of water molecule is essential for hydrolysis of the glycosyl-myrosinase intermediate and production of breakdown products [123, 124]. This myrosinase catalysed step was found to be rate-limiting and ascorbate can work as a cofactor here. It can place the water molecule to the correct position for activation. The glucosinolate degradation in presence of ascorbate by *Sinapis alba* myrosinase (Figure 1.6) occurs in this way: First, the glucosinolate binds to the active site and forms myrosinase-glucosinolate complex but this complex has a short life. The bound glucosinolate can be cleaved by nucleophile Glu-409. After formation of the glucosyl-myrosinase intermediate, ascorbate binds to the active site and releases a proton from a water molecule then nucleophilic attack at the anomeric center takes place. Eventually, glucose and ascorbate are unbound from the active site [124]. Breakdown products can be D-glucose, hydrogen sulphate, H+, nitriles, thiocyanates, amines, epithionitriles and ITC depending on the substrate or the pH [24].

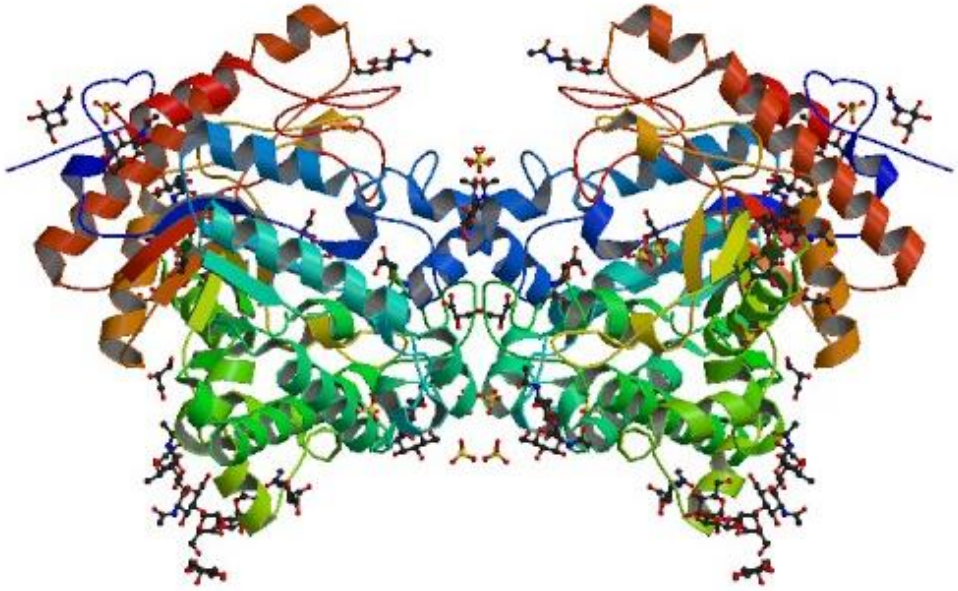


Figure 1.6 The overall structure of *S. alba* myrosinase. It exists as a dimer bound with a zinc atom. This image is taken from pdb database (1E4M) [124].

Myrosinases can be substrate specific for glucosinolates. For instance, two myrosinase isoenzyme isolated by James and Rossiter (1991) showed different degradation rates on different glucosinolates [125]. The study showed that both enzymes used in the study are more specific for aliphatic glucosinolates rather than indole glucosinolates [125], while another study reported the existence of an myrosinase that was highly specific for epi-progoitrin [126]. The specificity of enzymes is also affected by other factors like epithiospecifier or myrosinase binding proteins [2]. Plant type, organ, stage of development, seasonal factors or climatic conditions are the factors causing variation in the myrosinase activity [2, 123].

Myrosinase activity is affected by food processing conditions. Several studies tested the effect of processing conditions on glucosinolate content, bioavailability and myrosinase activity potential of Brassica vegetables [109, 116, 127-130]. Even if glucosinolates are reported to be heat-stable at mild temperatures [130], glucosinolate content might be affected by leaching to the medium during heating or enzymatic and thermal breakdown. In a study, the effect of boiling, low pressure steaming and microwaving on total glucosinolate content was investigated and it was reported that low pressure steaming and microwaving was responsible for a lower glucosinolate loss than boiling [109]. Oliviero et al. (2014) studied the potential bioavailability of sulforaphane and iberin and their conjugates in the urine after consumption of broccoli products. It was reported that complete inactivation of myrosinase

resulted in 10% and 29% of sulforaphane and iberin recovery in the urine respectively and this amount was the lowest recovery compared to other broccoli products with higher myrosinase activity. The study emphasises the importance of even a 20% residual myrosinase activity after food processing results in an ITC bioavailability as high as minimally processed broccoli with 100% myrosinase activity [116].

Myrosinases are not only found in plants. There is myrosinase-like activity from different sources like fungus [131], aphids [132] and gut bacteria [56, 112]. While plant and aphid myrosinases are well characterised [118, 124, 132-136], gut bacterial myrosinases are not.

The diets of aphids, butterflies, caterpillars and moths can include cruciferous vegetables [20] so they can also possess myrosinases and use myrosinase-glucosinolate system as a defence mechanim [132]. A study revealed that *Brevicoryne brassicae* (cabbage aphid) and *Lipaphis erysimi* (turnip aphid) can store glucosinolates separately from myrosinases which compartmentalized into crystalline microbodies [22]. The aphids were examined for myrosinase like activity and a non-plant myrosinase from cabbage aphid, *Brevicoryne brassicae*, was purified and characterised. It was the first non-plant myrosinase to be purified and not activated by ascorbate unlike the plant myrosinases [132]. Further study revealed that cabbage aphid myrosinase shows significant amino acid sequence identity to plant myrosinases (35% for *Sinapis alba* myrosinase) and enzymes from glycosyl hydrolase family [133]. Glutamic acid residues (Glu-167 and Glu 374) were reported to act as proton donor and nucleophile while lysine (Lys-173) and arginine (Arg-312) residues were proposed to have significant importance in the breakdown of glucosinolates by these aphids. The study concluded that according to sequence similarity and phylogenetic comparison, aphid myrosinases are more closely related to animal β-O-glucosidases [133]. In a further study with *Brevicoryne brassicae* myrosinase, the crystal structure of this enzyme was achieved. The general structure of this aphid myrosinase shows similarity to white mustard (*Sinapis alba*) myrosinase and plant β-O-glucosidases but plant β-O-glucosidases are not able to degrade glucosinolates [134].

Myrosinase activity in human gut bacteria is attributed to the β-glucosidases which have an specificity for glucosinolates [137]. A bacterial myrosinase from human gut has not been characterised in terms of gene identification. A study investigated the induced proteins of *L. agilis* R16 and *E. coli* VL8 from human gut in presence of sinigrin using 2D gel electrophoresis. They revealed that induced proteins were mostly part of carbohydrate metabolism,

oxidoreduction system and sugar transport, some have roles in purine metabolism, hydrolysis and proteolysis [138]. Recently, a bacterial myrosinase from soil was identified and partially characterised. This myrosinase was a periplasmic β-glucosidase with a signal peptide from *Citrobacter* WYE1 strain. It was found that it was activated by ascorbate like plant myrosinases [118].

1.6 METHIONINE SULPHOXIDE REDUCTASES

The enzymatic oxidation of methionine occurs in proteins and this oxidation was reported to impair the enzyme activity. [139-142] (Figure 1.7). Free amino acids or aminoacid residues in the proteins are being oxidized by reactive oxygen species (ROS) like superoxide, hydrogen peroxide, hydroxyl radical or hypochlorite anion [142]. This oxidation was reported to be oxidized by NADPH-dependent flavin monooxygenase enzymes in rabbits and humans [141]. The enzyme peptide methionine sulphoxide reductases are responsible for the reduction of methionine sulphoxide (MetSO) proteins to methionine (Met) in many organisms including mammals [143-145], flies [146], plants [147], yeast [148], bacteria [142, 149-151]. Methionine sulphoxide reductase (Msr) enzymes can repair key proteins that will lead to maintance of the activity of other proteins and they enable the usage of these proteins as ROS quenchers [142]. While the oxidation of Met to MetSO is a reversible reaction, the oxidation of MetSO to methionine sulphone ($MetSO_2$)is not [152]. The catalytic mechanism for Msr activity consists of sulfenic intermediate formation followed by the release of the reduced methionine and reduction of Msr to its original form which is called recycling step (Figure 1.8)[153]. Various reducing agents can perform the recycling step but thioredoxin/thioredoxin reductase system is the physiological electron donor [154].

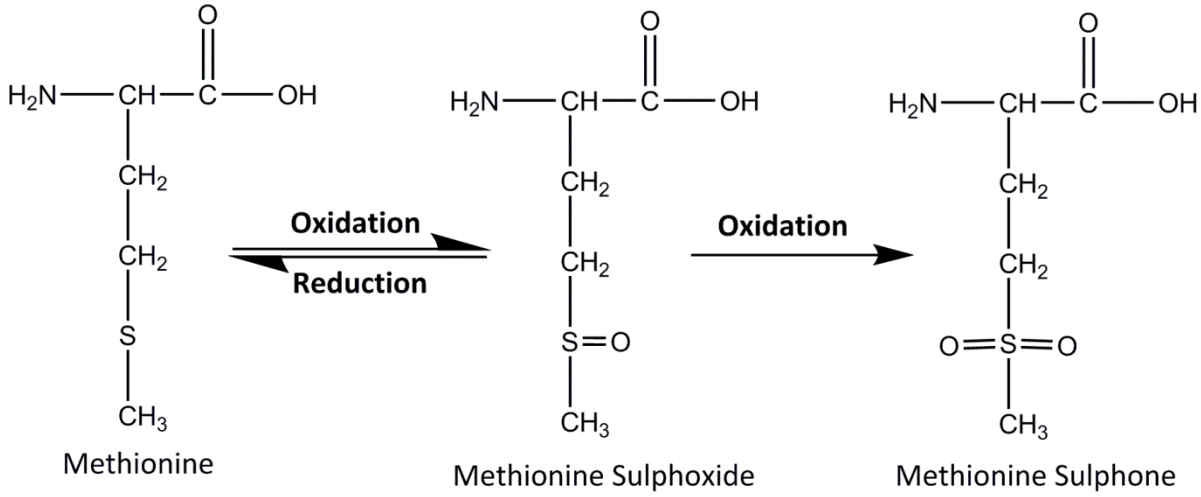


Figure 1.7 Structures of methionine and its oxidation products. Adapted from [152].

Msrs are divided in 3 main subclasses. First class, MsrAs are able to reduce free and protein-based Met(S)SO (S-epimers of MetSO); the second class, MsrBs can reduce Met(R)SO (R-epimers of MetSO) and the third class, free methionine sulphoxide reductases (fSMsr or

fRMsr) are responsible for catalysis of the reduction of free Met(S)SO and Met(R)SO [139, 140, 153, 155].

MsrA was knocked out in many organisms to evaluate their roles in oxidative stress and it was found to increase the susceptibilty to oxidative stress in mice [144], yeast [148] and bacteria [140]. The overexpression of MsrA in *Arabidopsis* [147], *Drosophila* [156], *Saccharomyces* and human T cells [145] was performed and it resulted in a better resistance to oxidative stress.

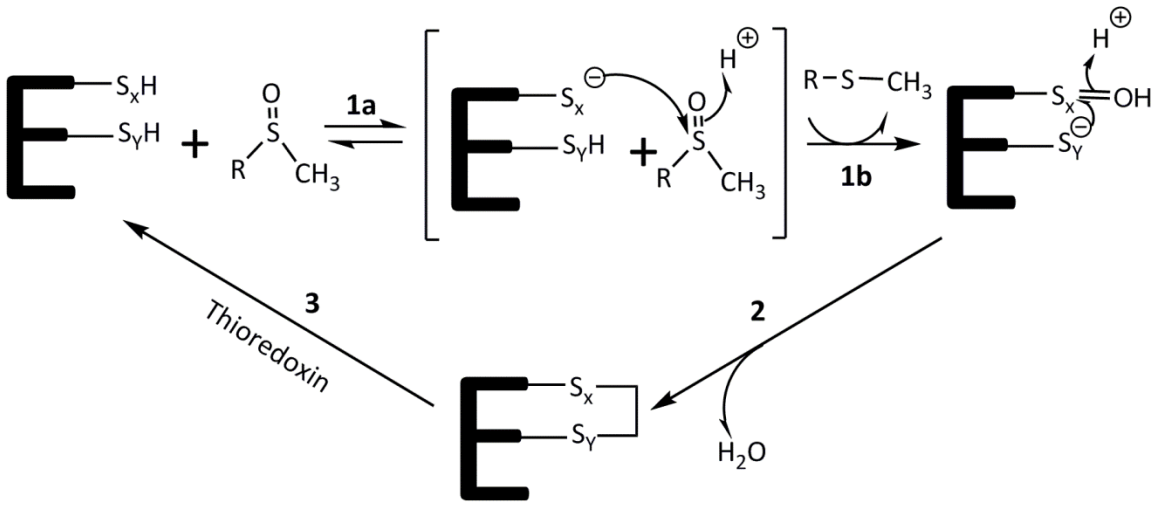


Figure 1.8 The action mechanism of Msrs (adapted from [153]). The action of Msrs initiates with the formation of Michaelis complex between Msr and MetSO (Step 1a). It proceeds by the nucleophilic attack of Cys-X on the sulphur atom of the sulphoxide group. This leads the release of reduced Met (Step 1b) and formation of a sulfenic acid intermediate. Then, nucleophilic attack of Cys-Y on the sulphur atom of sulfenic acid takes place resulting in release of a water molecule and formation of a Cys-X/Cys-Y disulphide bond. Later, reduction of Msr to its original form is performed by reduced thioredoxin. For *E. coli* MsrA, Cys-X is Cys 51 and Cys-Y is Cys 198.

Thioredoxin and thioredoxin reductase were reported to have role in reduction of methionine sulphoxide by transferring electrons from NADPH to Msr [141]. MsrA was discovered first in *E. coli* in 1980s [157] and there are many examples of crystallised structures from *E. coli* [158] (Figure 1.9), *Mycobacterium tuberculosis* [159], *Populus trichocarpa* (poplar) [160] and of bovine MsrA in the literature [161]. Later, MsrB of *Neisseria gnorroae* was cyrstallised in 2002 [151]. The free methionine sulphoxide reductases (fMsr) are reported to be missing in multicellular organisms like plants, mammals [155]. A different version of the Msr was identified in *Neiseria gnorrhoeae*. This enzyme is a fused form of MsrA and MsrB and called as MsrA/MsrB [151]. In addition to these Msr classes, the presence of membrane associated Msr for R and S epimers of MetSO was also reported [162].

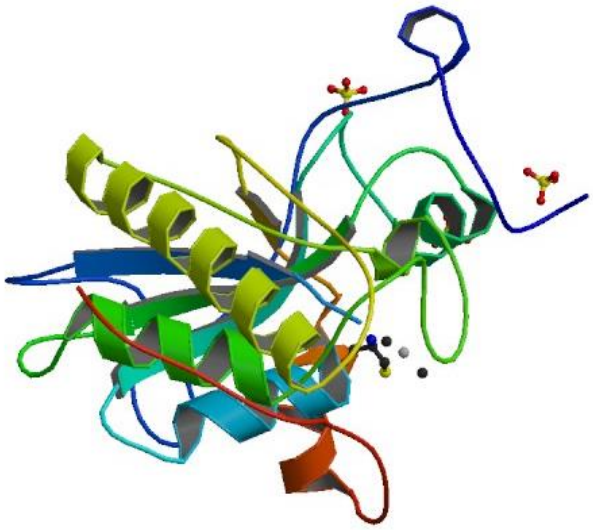


Figure 1.9 The structure of methionine sulfoxide reductase from *E. coli*. This image is taken from pdb database (1FF3) [158].

Bacteria might have different copy number of MsrA and MsrB. While *Staphylococccus* has 3 copies of MsrA and one MsrB, *Escherichia coli* has one copy for each [154].

Even though Msr is mainly known due to its potential role in repair system of proteins, it has been shown that it can be active towards wide variety of compounds with methyl sulphoxide groups [163]. Sulindac is an anti-inflammatory drug and was reported to be reduced by MsrA. The study also reported that MsrB was not efficient at reducing methyl sulphoxide of sulindac even if its amount was 200 times higher than MsrA [164]. Both MsrA and the membrane associated Msr (mem-*R, S*-Msr) are able to reduce sulindac [164].

MsrA was suggested to be responsible for reduction of methylsulfinylalkyl glucosinolates such as glucoraphanin into methylthioalkyl glucosinolates such as glucoerucin [55] (Figure 1.10). This reduction is an important step to fully understand the glucosinolate metabolism by gut bacteria. In one study [165], the glucosinolate bioconversion by human gut microbiota in a batch fermentation model using glucosinolates and faecal inoculum from a healthy volunteer was studied and conversion of glucoraphanin to glucoerucin did not occur with heat sterilized samples [165]. This suggested that this reduction might be enzymatically driven. Methionine sulphoxide reductases serve key roles for organisms by reduction of methionine sulphoxide but their role in glucosinolates metabolism is yet to be identified. In addition to reduction of glucosinolates, Msr might have role in reduction of ITCs such as reduction of sulforaphane to erucin. The interconversion between sulforaphane and erucin was reported *in vivo* previously but the detailed mechanism are not known [115].

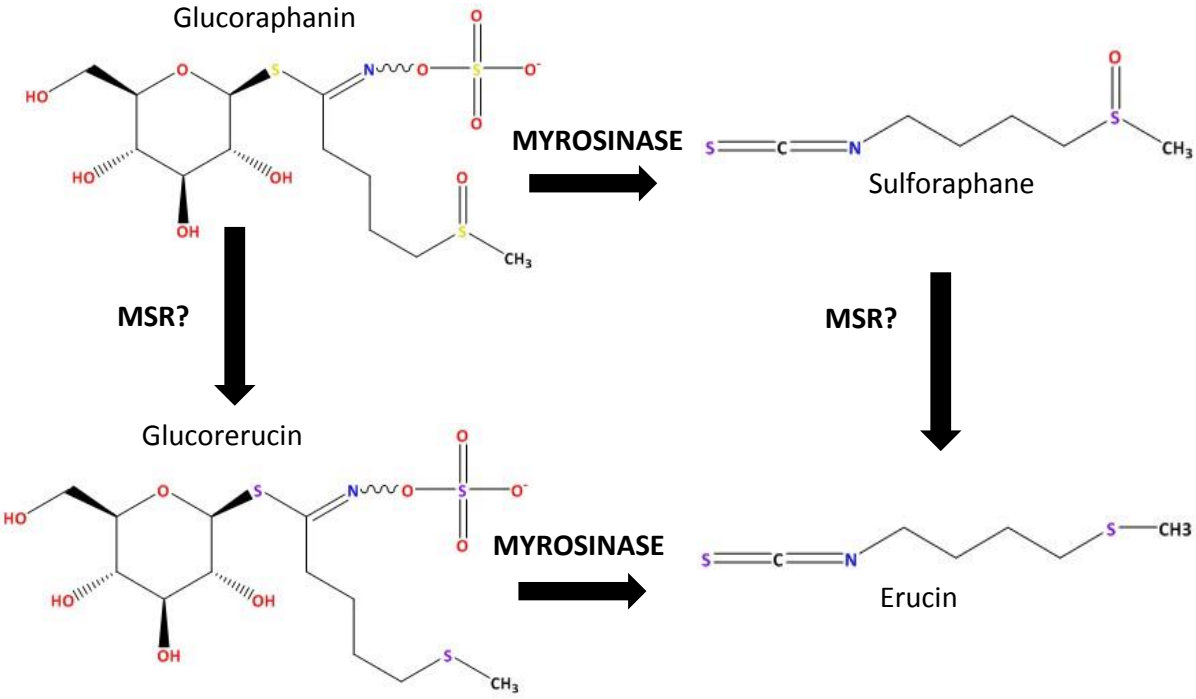


Figure 1.10 Proposed action mechanism of Msr enzymes on methylsulfinyl glucosinolates (adapted from[137]).

1.7 GUT MICROBIOTA

The human gastrointestinal (GI) tract (Figure 1.11) is a complex ecosystem and hosts a wide variety of microorganims which have important roles and various influences on the GI tract and host health [166, 167] (Figure 1.9). Although the microbiota makes up a relatively small amount of the human body composition, it is suggested that human gut microbiota has 100 times more gene capacity than the host [168].

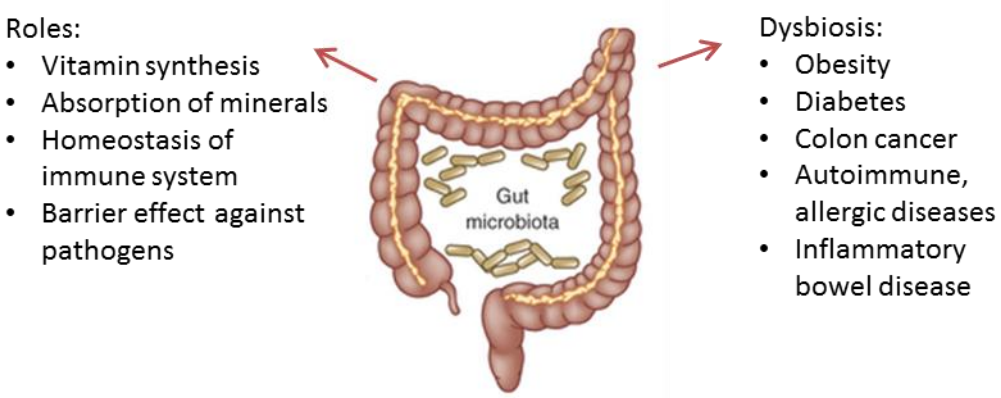


Figure 1.11 Illustration of the effect of human gut microbiota on human health and disease (adapted from [166, 167]).

The microbial content of GI tract can be variable and the microbiota plays an important role for host nutritional, physiological, immunological functions like food digestion, vitamin production, and disease pathogenesis [167]. The human intestinal system was initially thought to harbour 300-500 different species of bacteria but it has estimated recently to have more than 1200 different bacterial species [169]. This bacterial species refers to nearly 99% of microbes in the human gut, the rest (1%) consists of archaea, viruses and prokaryotes [169]. Only a small portion of this microbiota can be culturable but recent advances in culture-independent high-throughput sequencing has made it possible to identify more microbes in human gut [168, 170]. In addition, new methods combining sequencing and bacterial culturing approach are being developed and promising to culture and identify the substantial proportion of the gut microbiota [171]. The most dominant microbiota in human gut are from the phylum, *Firmicutes* and *Bacteroidetes*, they make up the >98% of all 16S rRNA sequences [172]. *Proteobacteria*, *Actinobacteria*, *Verrucobacteria* and *Fusobacteria* are also other abundant members of human gut microbiota [167, 172].

The diversity of microbiota in the infant gut is initially low and colonisation starts with aerotolerant species but then aerobes are replaced by anaerobes as typical microbiota of the adult gut [170]. By 2.5 years of age, the gut microbiota of an individual are established fully and once it has established, it remains constant until old age [170, 173]. Early colonised bacteria can modulate gene expression of host epithelial cells to provide a suitable habitat for themselves. Therefore the initial colonisation is quite important and influences the final composition of the adult gut microbiota [166]. Generally, microbiota composition remains stable but under certain conditions, it can show fluctuations. For instance, acute diarrhoeal diseases, diet, dysfunction of immune system or antibiotic intake and age can modulate the gut microbiota composition [166, 167].

Changes in diet have significant effect on the composition of gut microbiota [174]. A study about the effect of diet on the gut microbiota using humanized gnotobiotic mouse models showed that when mice diets are switched to a high in fat and sugar and low plant polysaccharide diet from a low in fat and rich in plant polysaccharide diet, there are differences in gut microbiota composition of mice just after 1 day [175]. Li et al. (2009) studied the effect of diet rich in cruciferous vegetables on fecal bacteria composition with a randomized, crossover, controlled feeding study and they concluded that human gut bacteria was altered by the cruciferous vegetables in the diet [176]. It was found that there was a

significant difference in gut bacterial composition of participants when they consumed a high cruciferous vegetable diet compared to consumption of a low-phytochemical, low-fiber diet. Specific bacteria such as *Eubacterium hallii*, *Phascolarctobacterium faecium*, *Alistipesputredinis*, and *Eggerthella* spp. were suggested to be linked to cruciferous vegetable intake. It was also reported that the response in the gut to cruciferous vegetables was individual-specific and there was no direct link between the response and amount of cruciferous vegetables taken [176].

One of the main factors that can modulate gut microbiota composition is antibiotic intake [170]. Antibiotic overuse can result in an increase in antibiotic-resistant pathogens. The antibiotic intake reshapes the gut microbiota and the effect of intake can cause long-term decrease in bacterial diversity [170].

The composition of the gut microbiota is quite variable between individuals depending on several factors such as host genetics [177, 178], diet [174, 175, 179] and environmental factors [169, 177]. The host genetics show significant effect on murine gut microbiota [177, 178]. A study investigating the gut microbiota of human infants from birth throughout the first year of their life using a microarray based analysis, found that the composition and temporal patterns of the intestinal microbiota were more similar in a pair of dizygotic twins than 12 unrelated infants [180]. Turnbaugh et al. (2009) investigated the effect of host genotype and adiposity on gut microbial communities of female monozygotic and dizygotic twins and their mothers. It was concluded that there were specific bacterial lineages for each individual and a comparable degree of co-variation between monozygotic and dizygotic twins [181]. Benson *et*. al (2010) [177] defined a core measurable microbiota of 64 conserved taxonomic groups found in across most animals in studied population showing variation quantitatively. The core measurable microbiota was tested with 530 fully informative single-nucleotide polymorphism (SNP) markers and 18 host quantitative trait loci (QTL) that have genome-wide linkage with microbial taxa were identified [177]. These QTLs were reported to affect the gut microbiota in different ways. Some loci have the control of individual microbial species or groups of related taxa and some loci are believed to have pleiotropic effects on groups of distantly related organisms [177].

The gut microbiota has a barrier effect against pathogens, and plays a crucial role in host homeostasis, vitamin synthesis and absorption of calcium, magnesium and iron [166]. It keeps homeostasis of the immune system by controlling epithelial cell proliferation and

differentiation. Ligation of toll-like receptors on intestinal epithelial cells by bacterial products induces epithelial cell proliferation and expression of antimicrobial peptides. The intestinal microbiota provides epithelial cell homeostatis and a repair system in the intestine [182]. The gut epithelium and immune system provide protection of the host against pathogens, but the intestinal microbiota also has a crucial role in the inhibition of the pathogens in the gut environment [166]. This protective effect has been called several names including bacterial antagonism, bacterial interference, barrier effect, colonisation resistance and competitive exclusion (CE) [183]. CE can be achieved through many mechanisms that are mostly bacteriostatic. Firstly, CE can be performed by creating a microecology which is hostile to pathogens, for example by production of volatile fatty acids. Secondly, CE can occur via competition between pathogens and native microbiota for limiting nutrients [184]. The third mechanism of CE is production of antibacterial substances such as bacteriocins and hydrogen peroxide [184-186].

Dysbiosis, a microbial imbalance of the gut, can be associated with several severe diseases like autoimmune and allergic diseases, obesity, iflammatory bowel diseases (IBD) or diabetes [167, 170]. Many studies have focused on search for the link between the gut bacteria and obesity [181] and diseases such as Crohn`s disease [187, 188], ulcerative colitis (UC) [188, 189] and colon cancer [190]. Obesity is associated with reduced bacterial diversity [181]. In a mouse model, obesity was observed to cause a decrease in Bacteroidetes and increase in Firmicutes [191]. The decrease in Bacteroidetes in gut was confirmed with a study on human twins but increase in Firmicutes was shifted to increase in *Actinobacterium* [181]. Dysbiosis is thought to be involved in the pathogenesis of IBD [192], the patients with IBD are reported to have more bacteria, from a varied genera, attached to their epithelial surfaces than healthy people [166]. The genera in gut such as *Lactobacillus* and *Bifidobacteria* showed prevention against tumorigenesis whereas *Bacteroides* and *Clostridium* can increase the incidence [166]. In a human study, the presence of two *Bacteroides* species, *Bacteriodes vulgatus* and *Bacteroides stercoris* were associated with high risk of colon cancer and *Lactobacillus* SO6, *Eubacterium aerofaciens* or total *Lactobacillus* concentration were associated with low risk of colon cancer [193].

Many techniques are used to study the human gut microbiota. Firstly, taxonomic profiling is used to identify the microbial community and determine the microbial content of that community. Second, metagenomics is used to provide data about abundance of genes and

gene transcripts. Third, the use of animal models provides us information about the effect of these genes on organisms and their functions [194]. Some of the techniques are listed below. They have both advantages and disadvantages. In the design of a study to assess the gut microbiota, cost and depth are key elements to select the method for taxonomic profiling. Sometimes, combination of a few techniques listed here is another way to study the human gut microbiota.

Culture techniques: Based on isolation of bacteria on selective media. It is fairly cheap but labour intensive. It gives a limited information about diversity because only a small portion of the gut microbiota have been cultured so far [195]. However, gut microbiota are not necessarily unculturable, a recent study using a combined metagenomic sequencing and bacterial culturing approach suggested that a substantial proportion of the intestinal bacteria is culturable [195].

Quantitative PCR (qPCR): Based on amplification and quantification of 16S rRNA. The techniques is similar to PCR but uses a compound that fluoresces when it binds to PCR product. It is a fast method and provides phylogenetic discrimination. However, PCR bias is possible and it is unable to identify novel species [195].

DNA Fingerprinting Techniques: Denaturing Gradient Gel Electrophoresis (DGGE) and Temperature Gradient Gel Electrophoresis (TGGE) were very popular inititally but they are now replaced by the sequencing methods. The 16S rRNA amplicons is separated by denaturant/temperature. The gels used can be excised and used in further analysis. It is a fast, semi-quantitative method but lacks phylogenetic identification. PCR bias may also be problem. These methods were mainly used for comparative analysis such as comparing the healthy state to a disease state [195].

Terminal restriction fragment length polymorphism (T-RFLP): Based on fragmentation of the 16S rDNA gene amplicons by restriction endonucleases with band visualization. It is fast, cheap and semi-quantitative but lacks phylogenetic identification. PCR bias may again be a problem. T-RFLP has been used to determine microbial community diversity in the gut microbiota [195].

Fluorescence *in situ* hybridisation (FISH): Based on hybridisation of fluorescently labelled oligonucleotide probes to target 16S rRNA sequences. It is a fast, semi-quantitative method

and provides phylogenetic identification. However, it lacks identification of unknown species [195].

DNA microarrays: Based on hybridisation of fluorescently labelled oligonucleotide probes to complementary nucleotide sequences. It is a fast, semi-quantitative method and provides phylogenetic identification. However there is a risk for cross-hybridisation (hybridisation of multiple probes to single targets). In addition, the detection of species which is at low levels in the gut microbiota can be missed out. DNA microarrays mainly used for comparison of the microbiota between different populations [195].

Sequencing of cloned 16S rRNA gene amplicons: Based on cloning of full-length 16S rRNA amplicon using the Sanger sequencing. It is good for quantification and phylogenetic identification of gut microbiota and enables the analysis of uncultured bacteria. It is laborious and expensive [195].

Direct Sequencing of 16S rRNA amplicons: Described as massive parallel sequencing of partial 16S rRNA gene amplicons. It may use beads (454 Pyrosequencing®), slides (Illumina®) and solid surfaces (SOLiD™). It is a quantitative method and provides phylogenetic identification with the chance of identification of unknown bacteria. However, it is labour intensive and expensive [195].

Microbiome Shotgun Sequencing: Based on massive parallel sequencing of the whole community DNA. It is a quantitative method, enables the phylogenetic identification. While other techniques analyse only genetic diversity, microbiome shotgun sequencing is based on both the genetic diversity and functions of the gut microbiota. It is an expensive method and needs an intense computational analysis of the data [195].

Animal models (in vivo): Based on rat, mouse, pig or zebrafish models. For instance, gnotobiotic rats can be used to study the link of one bacterial species or a group to a putative function or diversity of the gut microbiota but these models lack the full microbial community context [196].

In vitro colon models: These models are designed to mimic microbial processes throughout the gastrointestinal tract but they typically lack human cells. They are cheaper and offer an alternative to in vivo models for hypothesis testing [196].

1.8 THE ROLE OF GUT MICROBIOTA IN BIOTRANSFORMATION OF DIETARY COMPOUNDS

As diet can shape the human gut bacteria composition diversity and abundance [197, 198], the human gut bacteria also affects and directs the biosynthesis and metabolism of many compounds [176]. One of the main metabolic functions of the gut microbiota is fermentation of non-digestable dietary residues (such as non-digestable carbohydrates) and endogenous mucus [166, 199]. The microbial community of the gut provides a broad range of enzymes to the host which are different from its own resources. The main source of energy in the gut is the fermentation of carbohydrates. These carbohydrates include large polysaccharides, resistant starch, cellulose, hemicellulose, pectins and gums [166]. Short-chain fatty acids (SCFA) like acetate, propionate and butyrate are end products of fermentation by some members of the gut microbiota [199]. These fatty acids have crucial role in host physiology. For instance, butyrate is the major energy source of colonocytes [199]. Propionate [200] and acetate [166] are utilised by liver. These two fatty acids act as modulator in glucose metabolism so the intake of propionate and acetate are related with lower glycaemic responses [166]. A study showed that butyrate and propionate provides protection against high-fat diet induced weight gain and glucose intolerance and stimulate gut hormone secretion predominantly [200].

Some studies also suggest that gut bacteria are able to metabolise polyphenolic compounds in gut. This might result in with the formation of more biologically active compounds than the parent compounds [201]. For instance, certain health promoting gut bacterial species belonging to the genera *Bifidobacterium* and *Lactobacillus* are involved in phenolic acid metabolism in the gut [202] and five *Bifidobacteria* species are related with hydrolysis of soymilk isoflavones [203]. A study showed that *Eubacteruim limosum*, a strict anaerobe from the human intestinal tract, is able to metabolise some isoflavonoids [204]. An anaerobic bacterium strain isolated from mouse intestine was reported to be involved in the conversion of soy isoflavones [205]. Lignan metabolism is also performed by human gut bacteria. Plant lignans were reported to be metabolised by intestinal microbiota and β-glucosidases in the mucosal brush border were suggested to be responsible for this metabolism [206]. It was reported that 7-79% of hop derived isoxanthohumol was degraded to 8-prenylnaringenin after 72 h when fecal samples from female volunteers were used in a colon model [207]. A ten fold difference in 8-prenylnaringenin production was estimated between individuals

consuming beer moderately. Similarly, inter-individual differences in degradation of naringenin was reported [40]. In addition to other benefits of human gut bacteria, they are also involved in biotransformation of glucosinolates into their bioactive derivatives, ITCs.

1.9 GLUCOSINOLATE METABOLISM BY GUT BACTERIA

During cooking, plant myrosinases can be heat inactivated so biotransformation of glucosinolates becomes dependent on human gut bacteria. Many studies have reported bacteria with ability to metabolise glucosinolates. These studies are summarized in Table 1.4. In addition, some of the human gut bacterial strains including *Lactobacillus agilis* R16, *Enterococcus casseliflavus* CP1, *Escherichia coli* VL8 [56, 57] and *Bacteroides thetaiotaomicron* [112] were reported to show glucosinolate degrading ability.

Bacteria/faecal microbiota	Study Method	Reference
Enterobacter cloacae	*in vitro* incubation in media	[208]
Bacillus, *Pseudomonas* and *Lactobacillus* species	*in vitro* incubation in media	[209]
Lactobacillus agilis R16	*in vitro* incubation in media	[210]
Bacteroides thetaiotaomicron	Rat model	[112]
Human feacal microflora	*in vitro* incubation in media	[211]
Human feacal microflora	*in vitro* large-intestinal model	[212]
Human feacal microflora	Gnotobiotic rat model	[213]
Bifidobacterium species	*in vitro* incubation in media	[214]
Rat cecal microbiota	Rat model	[215]
Human feacal microflora	*in vivo* and *in vitro* incubation in media	[216]
Lactic acid bacteria and *Staphylococcus cornosus* UM 123M	*in vitro* incubation in media	[217]

Table 1.4 Studies reporting bacteria which have glucosinolate metabolising ability.

Bacteria/faecal microbiota	Study Method	Reference
Escherichia coli O157:H7 *Staphylococcus carnosus* *Staphylococcus auerus* *Pediococcus pentosaceus* *Salmonella typhimurium* *Listeria monocytogenes* *Enterococcus feacalis* *Pseudomonas fluorescens*	*in vitro* incubation in media	[218]
Human feacal microflora	a batch fermentation model	[165]
Lactococcus lactis subsp *lactis* KF147 *Lactobacillus plantarum* KW30 *Escherichia coli* Nissle 1917 *Enterobacter cloacae* ATCC13047	*in vitro* incubation in media	[55]
Listeria monocytogenes and *Salmonella*	*in vitro* incubation in media	[219]
Enterococcus casseliflavus CP1 *Escherichia coli* VL8	*in vitro* incubation in media	[56]

Table 1.4-Continued

A previous study investigated the sinigrin degrading activity of 42 lactobacilli and a strain called *Lactobacillus agilis* R16 showed the highest sinigrin transformation. The study showed this activity with sinigrin can be induced by the presence of the substrate sinigrin itself in culturing media. When *L. agilis* R16 was grown in media with sinigrin and glucose, the glucose was used immediately then sinigrin degradation started and resulted in 80% reduction of sinigrin in 4 h and total reduction in 24 h. Analysis of sinigrin degradation products showed the production of allyl ITC (AITC) and lactic acid [210]. The study also showed that cell-free extracts lack sinigrin degrading activity. This indicates that a cell-associated activity is involved. This phenomenon was further supported by another study by Luang-In et al. (2014) [56].

Bacteroides thetaiotaomicron, a human gut isolate, was studied for its ability to degrade sinigrin in a gnotobiotic rat model [112]. An oral dose of 50 µmol sinigrin was given to gnotobiotic rats harbouring *B. thetaiotaomicron*. Sinigrin was metabolised in the large bowel that host high level of *B. thetaiotaomicron*. By GC-MS analysis, allyl ITC, hydrolysis product of sinigrin, was detected in the digestive contents [112].

NMR spectroscopy was used to study the biotransformation of two dietary glucosinolates, sinigrin and glucotropaeolin by the human digestive microflora *in vitro*. It was shown that sinigrin and glucotropaeolin were transformed into allylamine and benzylamine, respectively and there was a low ITC formation (10-20% of the initial glucosinolate amount). Whether these amines are produced *in vivo* from dietary glucosinolates or not remains to be established. The effect of glucose in transformation of glucosinolates was also examined and the study showed that the presence of glucose did not modify the nature of the metabolites or the rate of transformation [211].

The dynamic *in vitro* large intestinal model was used to study the AITC production from sinigrin by human gut microbiota [212]. Sinigrin and AITC concentrations were determined in the lumen and dialysis fluid of the model using two different concentrations (1 mM and 15 mM) of sinigrin. AITC formation was observed and the percentage conversion of sinigrin to AITC was 1% and 4% in the lumen, respectively. The control experiment with inactivated microflora did not result in AITC formation. The concentration of AITC peaked between 9 and 12 h after addition of sinigrin. The study also suggested that different individual human microbiota result in significantly different conversion rate of sinigrin to AITC (9.6-30. 4% of sinigrin) [212].

The role of plant myrosinase and gut bacterial myrosinase on transformation of glucosinolates to ITCs were studied in a gnotobiotic rat model [213]. Glucosinolate biotransformation in gnotobiotic rats with a whole human fecal microbiota (Flora +) was compared to germ-free rats (Flora-). Active (Myro+) or inactive myrosinase (Myro-) were added to rat diet to determine the effect of plant myrosinase. The excretion of ITCs as mercapturic acid derivatives in urine was determined. The results indicated that highest excretion was performed by germ-free rats having active myrosinase in their diet (Flora-, Myro+ treatment) and a lower excretion was seen in rats with human flora when fed with inactive myrosinase (Flora+, Myro- treatment). The study suggested that the human microbiota can reduce the available ITC for intestinal absorption and this can be caused by

further metabolism of ITCs in the digestive tract. This study concluded that plant myrosinase and human gut microbiota both contributed to the degradion of glucosinolates into ITCs *in vivo* [213].

Cheng et al. studied *B. pseudocatenulatum*, *B. adolescentis* and *B. longum* for their ability to utilise sinigrin and glucotropaelin *in vitro* [214]. It was found that bacteria used in the study were hardly able to grow in basal medium without glucose in presence of 2.3% sinigrin but these bacterial strains used sinigrin as a glucose source. The study concluded that *Bifidobacterium* strains are able to utilize both sinigrin and glucotropaelin. However, GC-MS analysis results showed that the main products were 3-butene-nitrile. Phenylacetonitrile and AITC or benzyl isothicyanate (BITC) were hardly detectable. This result might be caused by rapid degradation of ITCs or transformation into non-volatile derivates [214].

The study examining the *ex vivo* degradation of glucoraphanin by rat cecal microbiota produced evidence for glucoraphanin hydrolysis to sulforaphane and its absorption across the cecal walls of rats [215]. Glucoraphanin (150 µmol/kg BW) was directly introduced into the cecum and ITC was detected in the mesenteric plasma by 120 min and plasma levels were stably maintained for an hour [215].

A study tested if individual differences in the human gut microbiota can result in different levels of glucosinolate metabolism. Twenty-three healthy individuals were fed with cooked broccoli and ITC levels in urine samples were measured. The highest and lowest 5 ITC excreters were selected for feacal sample collection. T-RFLP analysis of 16S rRNA gene revealed that there was no significant differences between high or low excreters' microbiota composition. However, high ITC excreters were able to degrade more glucoraphanin than low excreters. The study could not find a direct link between glucosinolate metabolism and specific bacterial species. It was suggested that this might be caused the complexity and functional redundancy of the gut microbiota [216].

Saha et al. (2012) studied the glucosinolate bioconversion by the human gut microbiota in a batch fermentation model using glucosinolates and fecal inoculum from a healthy volunteer. Glucoraphanin decreased after 4 h of incubation and mostly degraded after 24 h incubation. Glucoerucin was first detected at 8 h and further increased after 24 h. The study concluded that glucoraphanin was transformed into glucoerucin and erucin nitrile by the gut microbiota [165].

Luang-In et al. (2014) investigated the metabolism of glucoerucin, glucoiberin and glucoraphanin by gut bacterial strains *in vitro*. *Lactobacillus agilis* R16, *Enterococcus casseliflavus* CP1 and *Escherichia coli* VL8 were tested for the metabolism of glucosinolates supplied to their media. All three bacteria were able to degrade glucoerucin to erucin and erucin nitrile by 16 h. *E. coli* VL8 showed degradation of glucoiberin and glucoraphanin. The metabolism of glucoraphanin and glucoiberin by *E. coli* VL8 also resulted in erucin, erucin nitrile and iberverin, iberverin nitrile production respectively. A potential reductase enzyme has been proposed to be responsible for this bioconversion and the enzyme was reported to require both Mg^{2+} and NAD(P)H as cofactors. *E. casseliflavus* CP1 could consume 41 ± 7% of glucoiberin and 53 ± 8% of glucoraphanin and resulted in the formation of iberin and sulforaphane in low concentrations over a 24 h anaerobic fermentation. However, *L. agilis* R16 degraded only 11 ± 1% of glucoiberin and 10 ± 2% of glucoraphanin without giving any detectable products [56]. Following this study, desulfoglucosinolates were used to test their degradation by *L. agilis* R16, *E. casseliflavus* CP1 and *E. coli* VL8. All three metabolised desulfo-glucosinolates and gave pure nitriles products excluding *L. agilis* R16 which failed to produce phenethyl nitrile from gluconasturtiin [57].

Rats were found to be capable of metabolising glucosinolates in rapeseed meal after the myrosinase in the meal was inactivated [220]. This study was followed by more research and axenic rat and chickens were tested for glucosinolate metabolism in their guts. However, these studies did not result in identifying a responsible bacterial strain for this metabolism.

In brief, all these studies showed that several different phylogenetic families are involved in glucosinolate metabolism. It has not been possible to attribute glucosinolate degrading ability to only a specific group of bacteria so far.

1.10 SCOPE OF THE THESIS

As cancer is a very dreadful disease that is on the increase. It is important to take every single measure to fight with this cancer. The chemopreventation against cancer by bioactive compounds in the diet is one of the ways to combat the cancer. ITCs are very promising compounds for chemopreventation so it is essential to determine the glucosinolate metabolism and the effects of degradation products on human health. A better understanding of glucosinolate metabolism by human gut bacteria and identification of the

key enzymes in the metabolism are needed. So far, various bacteria from human gut have been reported to have glucosinolate degrading and ITC producing ability and their myrosinase activity was characterised. However, we lack the information about the enzymes responsible for glucosinolates metabolism in the human gut. In order to investigate these enzymes:

1. A glucoraphanin enrichment method will be used to isolate various bacteria possessing glucosinolates degrading ability from the human gut. Genomes of the isolated bacteria will be sequenced, annotated and assembled for interrogation in search for candidate myrosinase enzymes.
2. Selected candidate myrosinase genes from the bacteria will be cloned and expressed in *E. coli*. Enzyme activity of recombinant proteins will be examined and characterised in detail.
3. In addition to myrosinases, reductases which are proposed to play a key role in glucosinolates metabolism will be investigated. Different putative reductase genes from the genomes of human gut bacteria will be cloned and expressed in *E. coli*. The reductase activity of the recombinant reductases will be assessed.
4. Finally, a bacterial myrosinase of soil origin will be cloned and expressed in *E. coli* to achieve enzyme activity characterisation.

CHAPTER TWO

2 GENERAL MATERIALS AND METHODS

2.1 MICROBIOLOGY METHODS

All chemicals were purchased from Sigma Aldrich (Dorset, UK) unless stated otherwise.

2.1.1 Culture Media

All media used in the study are given below with their compositions.

Chemostat Media: 2 g peptone water (Oxoid), 2 g yeast extract (Oxoid), 0.1 g NaCl, 0.04 g K_2HPO_4, 0.04 g KH_2PO_4, 0.01 g $MgSO_4.7H_2O$, 0.01 g $CaCl_2.2H_20$, 2 g $NaHCO_3$, 2 ml Tween 80, 10 µl Vitamin K1, 0.5 g cysteine. HCl, 0.5 g bile salts (Oxoid), 0.02 g Hemin in 1 liter of water adjusted to pH 8.6 with 1 M NaOH.

M9 Minimal Media (without glucose): 6 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 1 g NH_4Cl, 0.147 g $CaCl_2$, 2.47 g $MgSO_4$, 1.5 ml of 10 mg/ml threonine stock, 11 ml of 10 mg/ml leucine stock, 0.23 g proline, 0.13 g arginine, 6.5 ml of 10 mg/ml histidine stock, 1.7 ml of 10 mg/ml thiamine stock in 1 liter of water.

Rogosa Agar: 10 g tryptone (Oxoid), 5 g yeast extract (Oxoid), 20 g glucose, 1 ml Tween 80, 6.0 g KH_2PO_4, 2.0 g ammonium citrate, 17.0 g anhydrous NaOAc, 0.575 g $MgSO_4$, 0.12 g $MnSO_4$, 0.034 g $FeSO_4$, 20 g agar (Oxoid) in 1 liter of water, pH was adjusted to 5.4 with glacial acetic acid.

Beerens Agar: 39 g Colombia agar base (Oxoid), 5 g glucose, 0.5 g cysteine. HCl, 5 g agar (Oxoid), 5ml propionic acid in 1 liter of water, pH was adjusted 5 with 1 M NaOH.

Wilkins Chalgren Media (without glucose): 10 g tryptone (Oxoid), 10 g peptone (Oxoid), 5 g yeast extract (Oxoid), 5 g NaCl, 10 g agar (Oxoid) 1 g L-Arginine, 1 g sodium pyruvate, 0.005 g Hemin, 0.0005 g Vitamin K in 1 liter of water (pH 6.5 ± 0.2). Solid media contains 10 g agar (Oxoid).

Bacteriodes Agar: 13 g agar, 10 g casein peptone, 1 g glucose, 10 g meat peptone (Oxoid), 5 g NaCl, 2 g yeast extract (BD), 10 ml of 0.5 mg/ml Hemin solution, 3 ml of 25 mg/ml Kanamycin, 7.5 ml of 1 mg/ml Vancomycin, 50 ml laked horse blood (Oxoid) in 1 liter of water (pH: 7.0 ± 0.2)

Clostridia Agar: 10 g tryptone (Oxoid), 10 g peptone (Oxoid), 5 g yeast extract (Oxoid), 5 g NaCl, 10 g agar (Oxoid), 1 g L-Arginine, 1 g sodium pyruvate, 0.005 g Hemin, 0.0005 g Vitamin K, 8 ml of 1 mg/ml Novobiocin solution, 8 ml of 1 mg/ml Colistin solution in 1 liter of water (pH 6.8 ± 0.2).

BHI Media: 12.5 g brain heart infusion solids (Oxoid), 5.0 g beef heart infusion solids, 10.0 g proteose peptone (Oxoid), 2.0 g glucose, 5.0 g NaCl, 2.5 g Na_2HPO_4 in 1 liter of water (pH 7.4 ± 0.2).

MRS Media (without glucose): 8 g Lab Lemco (Oxoid), 10 g peptone (Oxoid), 5 g Difco yeast extract (BD), 5 g $NaOAc.3H_2O$, 2 g K_2HPO_4, 2 g triammonium citrate, 1 ml Tween 20, 5 ml of 11.5% $MgSO_4.7H_20$ solution, 5 ml of 2.8% $MnSO_4.4H_2O$ solution in 1 liter of water (pH 6.5 ± 0.2). Solid media contains 20 g agar (Oxoid).

Nutrient Broth (without glucose): 5 g peptone (Oxoid), 1.5 g Lab Lemco (Oxoid), 1.5 g yeast extract (Oxoid), 5 g NaCl in 1 liter of water (pH 6.8 ± 0.2).

L Broth: 10 g Difco Bacto tryptone (BD), 5 g Difco Bacto yeast extract (BD), 10 g NaCl, 1 g glucose in 1 liter water (pH 7.0 ± 0.2).

SOC Broth: 20 g Difco Bacto tryptone (BD), 5 g Difco Bacto yeast extract (BD), 0.5 g NaCl in 1 liter of water.

2.1.2 Preparation of Antibiotic Stock Solutions

The antibiotic stock solutions were prepared according to laboratory cloning manual [221] then sterilised by filtration using 0.22 μm filters. The aliquots were stored at -20°C until usage. The stocks were listed below in Table 2.1.

Antibiotic	Concentration of the stocks	Final concentration
Ampicillin	50 mg/ml in H_2O	100 μg/ml
Kanamycin	10 mg/ml in H_2O	30 μg/ml

Table 2.1 Antibiotic stock solutions used in this study.

2.1.3 Isolation of Glucosinolate Degrading Bacteria by Glucoraphanin Enrichment

A faecal sample from a healthy volunteer (5 g) was homogenized in 45 ml of phosphate buffered saline (PBS) with a stomacher (Stomacher 400 Circulator, Seward) at 230 rpm for 2 min. The homogenate (100 μL) was used to inoculate the chemostat media or M9 minimal media both containing 8 mM glucoraphanin (900 μL) and these cultures were grown in an anaerobic cabinet (Whitley A95 anaerobic work station, Don Whitley Scientific) under an atmosphere of 5% CO_2, 10% H_2 and 85% N_2. Aliquots were diluted (10 fold) in corresponding media containing 8 mM glucoraphanin every two days. At day eleven, cultures were diluted in PBS and plated on 6 different media including Rogosa, Beerens, Wilkins Chalgren,

Bacteriodes, Clostridia and BHI. Individual colonies from these plates were selected and streaked on BHI agar twice for purity.

2.1.4 Bacterial Strains and Culture Conditions

Escherichia coli VL8, *Escherichia coli* FI10944 and *Enterococcus casseliflavus* were isolated from human gut by using sinigrin enrichment by Dr. Vijitra Luang-In at Institute of Food Research and *Lactobacillus agilis* R16 was provided by Dr. Llanos-Palop (Table 2.2).

Bacterial Strain	Enrichment Method	Growth Condition	Media
Lactobacillus agilis R16 [56]	*	37°C, 200 rpm	MRS Broth
Escherichia coli VL8 [56]	Sinigrin	37°C, 250 rpm	L Broth
Escherichia coli FI10944	Sinigrin	37°C, 250 rpm	L Broth
Enterococcus casseliflavus CP1 [56]	Sinigrin	37°C, 200 rpm	BHI Broth
Escherichia coli FC44	Glucoraphanin	37°C, 250 rpm	Nutrient Broth
Citrobacter freundii FC50	Glucoraphanin	37°C, 250 rpm	Nutrient Broth
Citrobacter WYE1 [118]	Sinigrin	37°C, 250 rpm	Nutrient Broth
Escherichia coli DH5α[a]	*	37°C, 250 rpm	L Broth
Escherichia coli BL21(DE3)[b]	*	37°C, 250 rpm	L Broth

Table 2.2 Bacterial strains used in the study. [a] *Escherichia* coli DH5α and [b] *Escherichia* coli BL21(DE3) were obtained from ThermoFisher Scientific. *These strains were not isolated by a glucosinolate enrichment method.

Bacterial isolates were grown overnight and cultures were used to prepare 1 ml of glycerol stocks (20% v/v) then stocks were frozen on dry ice and stored at -80°C.

Optical cell density (OD_{600}) was measured with a JENWAY 6715 UV/Vis Spectrophotometer (Staffordshire, UK) using bandwidth 1.5 nm wavelength 600 nm.

2.1.5 Bacterial Growth Analysis

Growth of strains was analysed over time using a Bioscreen C machine (Labsystems). Corresponding media for each strain was inoculated with 1% (v/v) concentration of overnight grown culture then inoculated media was transferred to Bioscreen plate wells (Honeycomb, ThermoFisher Scientific), 300 µl of each inoculated media was used per well. Growth was screened for 48 h at an appropriate temperature for each strain. Measurements were

performed at 600 nm with 10 min measurement intervals with shaking for 10 seconds before each measurement.

2.1.6 Scanning Electron Microscopy (SEM)

L. agilis R16, *E.* coli FI10944 and *E.* casseliflavus CP1 were grown overnight from bacterial stocks. These cultures (2 X 1 ml) were centrifuged at 3000 x g for 2 min at room temperature. The supernatants were removed and the pellets were resuspended in 1 ml PBS. To produce a final concentration of 2.5% (v/v) glutaraldehyde (Sigma, UK), 110 µl of 25% (v/v) glutaraldehyde was added to the tubes and left for 1 h to complete the fixation.

After fixation, each sample analysed by Rachael Stanley at Institute of Food Research. Scanning electron microscopy was performed using a Zeiss Supra 55 VP FEG SEM (Zeiss, Germany) operated at 3 kV.

2.2 MOLECULAR BIOLOGY

2.2.1 Polymerase Chain Reaction (PCR)

Bacterial genomic DNA was used as template to amplify the genes of interest using High Fidelity Phusion Polymerase (Finnnzymes) according to supplier`s instructions. Reactions were set up as in Table 2.3 and performed as in Table 2.4.

Components	Volume (µl)	Final Concentration
5X Phusion Buffer	10	1X
100 mM dNTPs	0.4	0.2 mM each
20 µM Forward Primer	1.25	0.5 µM
20 µM Reverse Primer	1.25	0.5 µM
Phusion Polymerase (2 U/µl)	0.4	
Template DNA	1	<50 ng/50 µl
Sterile MQ water	34.7	
Total Volume	50	

Table 2.3 Components of one PCR for amplification of the gene of interest.

The annealing temperatures and conditions were changed depending on primers. In general, the reactions were carried out as shown in Table 2.4 for 5 cycles using T_m of 100% sequence match part of primers first and then 20 cycles using the T_m of entire primers.

	Steps	Temperature (°C)	Time	Cycle(s)
1	Initial Denaturation	98	30 s	1
2	Denaturation	98	10 s	5
	Annealing	T_1	30 s	
	Extension	72	30 s/kb	
3	Denaturation	98	10 s	20
	Annealing	T_2	30 s	
	Extension	72	30 s/kb	
4	Final Extension	72	5min	1

Table 2.4 PCR conditions for amplification of the gene of interest.

2.2.2 Gel Electrophoresis

To check the sizes of DNA fragments, agarose gels were prepared at 0.6% (to check genomic DNA samples) or 1% (g w/v). 5-10 µl of DNA samples were loaded to the wells of agarose gel and run in 0.5X Tris Borate EDTA buffer (ThermoFisher Scientific). The gels were stained in 1 mg/ml ethidium bromide for 20-30 min then rinsed in deionized water. After electrophoresis, the gels were visualised with an Alphaimager (Alpha Innotech) using UV light. The size of products was determined by running a DNA marker for every electrophoresis (HyperLadder, Bioline).

2.2.3 Plasmid Preparation

For cloning and expression of the genes of interest, either pET15b or pET28b vectors (Novagen) were used. While the pET15b vector has an N-terminal His Tag sequence, the pET-28b vector also carries an optional C-terminal His Tag sequence (maps of the vectors are given in Appendix 1). The E.Z.N.A Plasmid Mini Kit I (Omega Bio-Tec) was used to purify the plasmids following the supplier`s protocol.

2.2.4 Enzyme Restrictions

Reactions were set up in 15-40 µL volumes following the supplier`s advice (New England Biolabs, UK). After digestion, enzymes were heat inactivated and the products were purified by SureClean (Bioline) to remove the restriction enzymes.

2.2.5 Dephosphorylation of Digested Plasmids

To prevent the recirculation of digested plasmids, Antarctic Phosphatase (New England Biolabs, UK) was used to remove 5' phosphate groups. The dephosphorylation reactions contained ~ 1 µg of the digested vector, 1X Antarctic Phosphatase Buffer, 1 µl Antarctic Phosphatase and incubated for 15 min at 37°C for 5' extensions or blunt ends or 60 min for 3' extensions. After incubation, enzymes were heat-inactivated by incubating at 65°C for 5 min.

2.2.6 Primer Design

Primers were ordered from Sigma Genosys. For subcloning, primers were designed with restriction sites on forward (shown as blue) and reverse (shown as green) site of the target gene (shown as red) (Figure 2.1).

Figure 2.1 Illustration of a PCR with forward and reverse primers.

Primers were designed to be between 18-24 nucleotide long with an annealing temperature between 45-65°C. It is advised to have a GC clamp at 3'-end of the primers and not to have the more than 4 repeats of the same nucleotide. The restriction enzymes need a certain amount of base pairs at the start of recognition site to cleave efficiently. The numbers of base pairs needed for successful recognition was obtained from the New England Biolabs website (https://www.neb.com/tools-and-resources/usage-guidelines/cleavage-close-to-the-end-of-dna-fragments) and these amount of base pairs were added to the primers. The primers also consisted of at least 15 bp of 100% complementary sequence to the gene of interest if introducing a new restriction site. The melting temperature of the entire primers and 100% match part were calculated by using the ThermoFisher Scientific Tm Calculator (https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular

biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html). Sigma OligoEvaluator tool was also used to assess primer design (http://www.sigmaaldrich.com/technical-documents/articles/biology/oligo-evaluator.html).

2.2.7 PCR Product Purification

The PCR products (5 µl) were analysed on 1% agarose gel and checked to have expected product sizes. The rest of the products were purified using SureClean (Bioline) and quantified using Nanodrop 2000 (ThermoFisher Scientific).

2.2.8 DNA Ligation

Ligations were performed using Fastlink DNA Ligase (Epicentre) following the supplier`s protocol with the given components in Table 2.5. The reactions were performed in a total volume of 15 µl. The volumes of insert and vector were varied depending on their concentrations but final concentration of vector in all ligations was kept as 10 ng/µl and the molar ratio between vector and insert was 1:3 (vector: insert). The ligations were incubated at room temperature overnight and then heat-inactived at 70°C for 15 min. The heat-inactivated ligations were used directly for transformation of *E. coli*.

Components	Volume
Double Digested Vector	3 µl (10 ng/ml, in 15 ml reaction)
Gene Insert	1 µl
10X Fast Link Ligation Buffer	1.5 µl
10 mM ATP	1.5 µl
Sterile MQ water	7 µl
Fast Link DNA Ligase (2 U/µl)	1 µl
Total Volume	15 µl

Table 2.5 Components of one DNA ligation reaction. The molar ratio of vector: insert was 1:3.

2.2.9 Preparation of Chemically Competent Cells

Transformation buffers were prepared as given in the Table 2.6. Buffer 1 was adjusted to pH 5.8 with 0.2 M acetic acid and Buffer 2 was adjusted to pH 6.4 with KOH then they were both filter sterilised and kept at 4°C. L Broth (5 ml) was inoculated with *E. coli* DH5α or BL21(DE3) strains from -80°C stocks and grown overnight at 37°C, 200 rpm with shaking. Next day,

overnight cultures were subcultured (1:100) into 50 ml of fresh L Broth. Cultures were grown at 37°C, 200 rpm shaking until OD_{600} reached 0.3-0.35. The cells were spun down for 7 min at 1,500 x g at 4°C then resuspended in 20 ml of Buffer 1 and were left on ice for 5 min. The cells were then spun down for 7 min at 1,500 *x* g at 4°C and resuspended in 2 ml of Buffer 2. The cell suspension was transferred to prechilled Eppendorf tubes as 110 µl stocks. The stocks were used as soon as possible or frozen immediately on dry ice for long term storage and kept at -80°C.

Buffer 1 composition	Stock solutions	V of stock solution for 50 ml TFbI
30 mM potassium acetate	300 mM	5 mL
100 mM rubidium chloride	500 mM	10 mL
10 mM calcium chloride	100 mM	5 mL
50 mM manganese chloride	200 mM	12.5 mL
15% glycerol (v/v)	100%	7.5 mL
Distilled water		10mL

Buffer 2 composition	Stock	50 ml
10 mM 3[N-Morpholino]Propanesulfonic Acid (MOPS)	100 mM	5 mL
75 mM calcium chloride	100 mM	37.5 mL
10 mM rubidium chloride	500 mM	1 mL
15% glycerol (v/v)		7.5 mL

Table 2.6 Components of buffers used in competent cell preparation.

2.2.10 Transformation of *E. coli*

Freshly prepared or stored *E. coli* DH5α or BL21 (DE3) chemically competent cells were thawed and incubated on ice for 10 min. The transformation was set up by adding ligation reactions (<100 ng DNA per 50 µl cells) to 50 µl of competent cells and incubated on ice for 30 min. After incubation, cells were exposed to heat shock at 42°C for 45 s and immediately put on ice again for 2 min then 250 µl of SOC broth was added. The cells were incubated at 37°C for at least 1 h shaking, 200 rpm. After incubation, 25 µl and 100 µl of cultures were plated on selective agar. The final antibiotic concentration was 100 µg/ml of ampicillin for pET15b and 30 µg/ml kanamycin for pET28b.

2.2.11 Selection of the Positive Transformants

L Agar plates were prepared with the final antibiotic concentration of 100 µg/ml ampicillin and 30 µg/ml kanamycin for pET15b and pET28 respectively. After transformation, the colonies grown on selective plates were selected randomly and colony PCR was performed to determine if these colonies have the right gene fragment of interest. Colony PCR was set up to screen the colonies for plasmids with right insert. The colonies were picked and resuspended in 10 µl of sterile Milli-Q water (1 µl of this was used as DNA template for PCR).

The products were amplified by PCR using the primers T7P2: 5': TGAGCGGATAACAATTCCC, Tm: 62.3 ^{o}C and T7_T: 5': GCTAGTTATTGCTCAGCGC, Tm: 60.1^{o}C (Sigma Genosys) and 1 µl of toothpicked colony suspension as DNA template. Colony suspension was replaced with sterile MQ water or empty vector control (pET15b or pET28b) for negative and positive controls. The reactions were set up as shown in Table 2.7.

a

Component	Quantity
DNA Template	1 µl
5X GoTaq Reaction Buffer (Promega)	10 µl
dNTP 25 mM each (Bioline)	0.4 µl
Primer F 20 mM	1 µl
Primer R 20 mM	1 µl
GoTaq DNA Polymerase (5 U/µl)	0.25 µl
Water	36.35 µl
Total	50 µl

b

PCR conditions		
Temp , time		cycles
95°C, 2 min		1
95°C, 30 s		25
55°C, 30 s		
72°C, 1min/kb		
72°C, 5min		1

Table 2.7 Components in one colony PCR (a) and thermal cycling conditions (b).

2.2.12 16S rDNA Sequencing

An overnight culture (150 µl) was centrifuged at 13000 x *g*, the pellet was washed and resuspended in 15 µl sterile Milli-Q water. The pellet was boiled at 95°C for 5 min, cooled on ice and 1 µl of this was used as DNA template in PCR. The gene was amplified using the primers AMP_F: 5'-GAG AGT TTG ATY CTG GCT CAG Tm: 60.5, AMP_R: 5'-AAG GAG GTG ATC

CAR CCG CA Tm: 68.9 [222] to give a product of approximately 1500 bp. PCR was set up as given in Table 2.8.

a

Component	Quantity
DNA Template	1 µl
5X GoTaq Reaction Buffer (Promega)	10 µl
dNTPs 25 mM each (Bioline)	0.4 µl
Primer F 20 µM	1 µl
Primer R 20 µM	1 µl
GoTaq DNA Polymerase (5 U/µl)	0.25 µl
Water	36.35 µl
Total	50 µl

b

PCR conditions	
Temp, time	cycles
95°C, 2 min	1
95°C, 30 s	20
55°C, 30 s	
72°C, 1min 30s	
72°C, 5min	1

Table 2.8 Components in one PCR reaction (a) and thermal cycling conditions (b) In addition, a negative control (without template) was prepared. PCR products (10 µl) were analysed by agarose gel electrophoresis to check the size of the products. PCR products were purified by QIAquick PCR purification kit (QIAGEN) and quantified by a Nanodrop 2000 spectrometer (ThermoFisher Scientific). Products were sequenced with both primers (Eurofinn Genomics). Finch TV was used to check chromatogram quality (sequence quality), the reliable sequences were selected and copied to EditSeq (DNAStar) and saved. Then these sequences were assembled by using Seqman (DNAStar) to create a contig and this contig was searched on RDP database (https://rdp.cme.msu.edu/seqmatch/seqmatch_intro.jsp).

2.2.13 Genomic DNA Extraction, Sequencing, Assembly and Annotation

Genomic DNA of strains was extracted by using the Genomic DNA extraction kit with Genomic Tip 20/G columns (QIAGEN) as advised by the supplier. To enhance the cell lysis of *L. agilis* R16, 50 Units of mutanolysin (Sigma) was added. Genomic DNA of *E. coli* FI10944, *E. casseliflavus* CP1, *L. agilis* R16 and *C.* WYE1 was sequenced by The Genome Analysis Centre (TGAC), Norwich, UK, using the Illumina MiSeq platform. DNA sequencing resulted in small reads of DNA with a read length of 251bp (2x251, paired end). These small reads were analysed using different bioinformatics tools step by step. Sickle [223] , Abyss [224] and CAP3 [225] were used to produce quality-trimmed extended contigs from these small reads.

Genomic DNA of Isolate 44 and *E. coli* VL8 was sequenced at University of Birmingham, MicrobesNG Unit, Birmingham, UK. The sequencing was performed using NexteraXT library prep on two MiSeq runs and produced 2x250bp paired-end reads. Genomic DNA assembly of *E. coli* FC44, *E. coli* VL8 and *C.* WYE1 was performed by using SPAdes. The reads were analysed and quality-trimmed extended contigs were produced by using the SPAdes program version 3.5.0 [226].

Genomic DNA of *C. freundii* FC50 was sequenced at the Animal and Plant Health Agency, NGS Services Unit, Surrey, UK. The sequencing was performed using the Ilumina MiSeq platform, producing 150bp paired end reads. Genomic DNA assembly was performed using SPAdes program version 3.7.1 [226]. Following assembly, contigs were analysed by ARTEMIS [227] and annotated by RAST [228].

2.2.14 Preparation of Phylogenetic Tree

The 251 bp length of DNA reads from Genomic DNA sequencing were analysed and quality-trimmed extended contigs were produced by using SPAdes program version 3.6.1 [226]. These contigs of *Citrobacter* were used to construct a phylogenetic tree. Draft genomes of different *Citrobacter* species were obtained from Genbank whole genome shotgun projects. Core genome SNPs in draft genomes and *C.* WYE1 were used to prepare the phylogenetic tree using the parSNP program Version 1.2 by Dr. Arnoud Van Vliet [229].

2.3 PROTEIN BIOCHEMISTRY

2.3.1 Induction of Protein Expression and Cell-Free Extract Preparation

Isopropyl-β-D-1-thiogalactopyranoside (IPTG) was used to induce protein expression in *E. coli* BL21(DE3) expressing the gene of interest. The cells were inoculated in L broth with antibiotic (10 ml) from -80°C stocks and grown overnight at 37°C, 250 rpm. Next day, 10 ml L broth with antibiotic medium was inoculated with 200 µl of the overnight culture and grown until midlog (OD_{600}=0.4) then cells were induced with 0.5 mM IPTG.

For each of the expressed protein, the effect of several time points and temperatures on protein expression was tested to determine the optimum protein expression conditions. *E. coli* BL21 (DE3), transformed with pET15b or pET28b (empty vector control) was also induced as a negative control. After induction, cells were harvested by centrifugation for 15 min at 3200 x *g* at 4°C, the supernatant was removed and the pellet was frozen at -20°C. Later, the

cell pellet was resuspended in 500 µl of 20 mM Tris base 50 mM NaCI pH 7.5 (resuspension buffer) unless a different buffer is stated. The cells were extracted by sonication (Soniprep 150, MSE) by sonicating for 15 s and were keeping on ice for 30 s to cool down. This step was repeated 7 times for each sample. After sonication, the suspensions were centrifuged for 25 min, 13000 x *g* at 4°C. The supernatants were transferred into new clean tubes (soluble extracts), 500 µl resuspension buffer was added to the cell pellet (insoluble extracts). The suspensions and extracts were always kept on ice.

2.3.2 Protein Quantification

The protein concentration of samples was determined using the Bradford assay (Bio-Rad, UK). The reactions were set up in 96 well plates (Greiner Bio-One, Austria) in triplicates. Bovine serum albumin (BSA) (New England Biolabs, UK) was used to prepare a calibration curve. BSA was prepared in different concentrations between the range of 0-200 µg/ml then the samples were diluted to ensure that protein concentration will fall within a 0-200 µg/ml range. Bradford assay concentrate was diluted 100X and added to samples and BSA solutions. The absorbance was measured at 600 nm using a ThermoMax Plate Reader (GMI, US). A calibration curve of BSA was prepared (Figure 2.2) and protein concentrations of samples were determined.

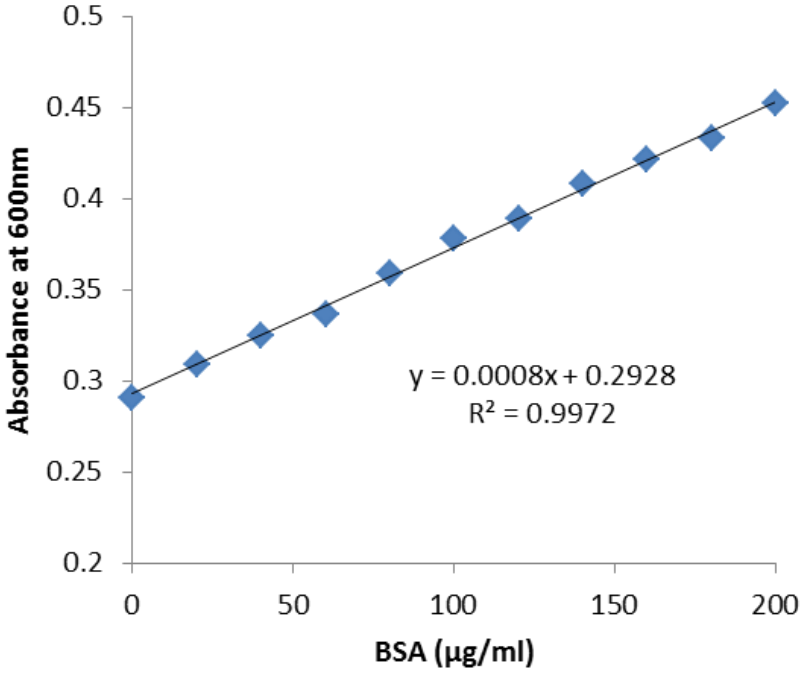


Figure 2.2 A reference calibration curve of Bradford assay. Different amounts of BSA were plotted against the absorbance at 600 nm.

2.3.3 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE materials were purchased from Invitrogen (ThermoFisher Scientific, UK). The proteins were mixed with LDS sample buffer and reducing agent (under reducing conditions) as advised by the supplier and then heat denaturated by heating at 70°C for 10 min. The proteins (mainly ~10 µg) were loaded on a NuPAGE®4-12% Bis-Tris protein gel in MES buffer and placed in a X-Cell Sure Lock electrophoresis cell.

The SeeBlue Plus2 protein marker (Figure 2.3) was used to determine the protein sizes of the samples. For reducing conditions, 500 µl of NuPAGE® antioxidant was added to the internal reservoir. The protein gel was run for 35 min at 200 V. Then the gel was removed from cell then either washed three times in MQ water and stained with Simply Blue SafeStain or directly stained by Instablue dye (Expedeon, US) for 1 h.

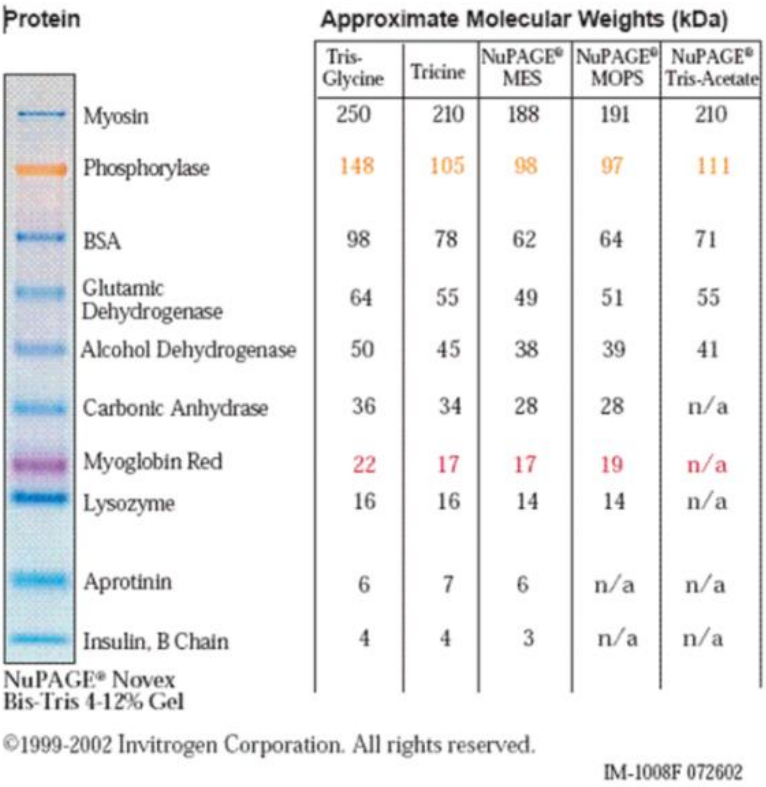


Figure 2.3 See Blue Plus2 protein marker (Invitrogen) on a 4-12% Bis-Tris gel with different buffers.

2.3.4 Western Blotting

Western Blot procedure was used to purify the 6XHis- tagged proteins of interest. SDS-PAGE gel was run as described previously (Section 2.3.3) and the proteins were transferred from protein gel to polyvinylidene difluoride (PVDF) membrane using a XCell II Blot module (Novagen, Germany). NuPAGE® Transfer Buffer (20X) buffer was prepared as advised by supplier and used in preparation of 1X transfer buffer (See Table 2.9).

Components	Volume (ml)
NuPAGE®Transfer Buffer (20X)	25
NuPAGE® Antioxidant	0.5
Methanol	50
Deionized water	424.5
Total	500.0

Table 2.9 Components in NuPAGE® Transfer Buffer (20X) buffer.

The blotting pads were soaked in transfer buffer at least 20 min before use. The PVDF membrane was pre-wetted in methanol and rinsed in deionised water and placed in 50 ml of

NuPAGE® Transfer Buffer before use. The filter papers were briefly soaked in NuPAGE® Transfer Buffer prior to use. The blot module was set up as shown in the Figure 2.4 within the order from cathode core; two blot pads, filter paper, protein gel, PVDF membrane, filter paper and two more blot pads. All components were set up to contact each other to perform the transfer efficiently and blot module was placed in buffer chamber. The blot module was filled with 1X NuPAGE® Transfer Buffer to cover the gel and membranes. Excessive amount of buffer was avoided for a successful transfer. The outer chamber was filled with ~650 ml of MQ water and transfer was performed at 12V for 3 h.

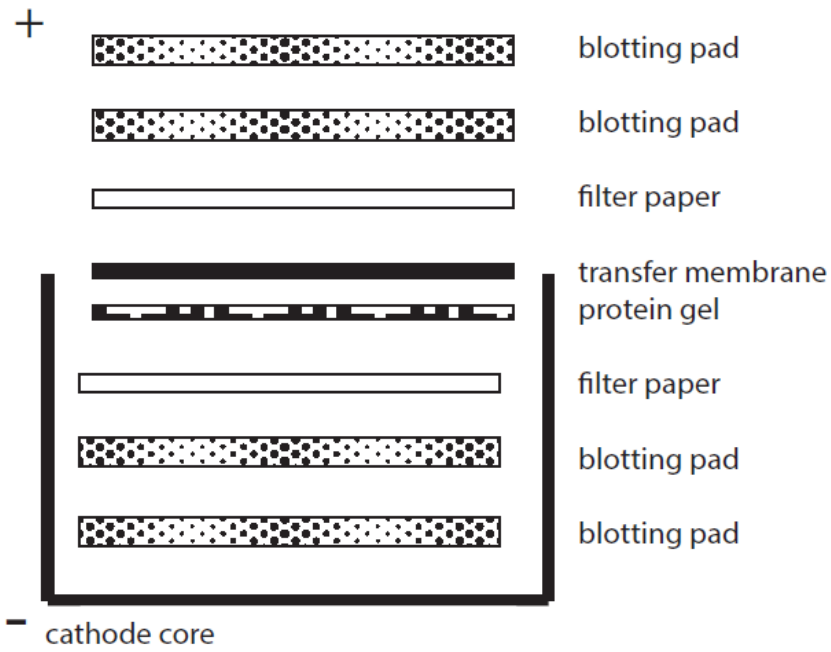


Figure 2.4 Western blot module set up order.

The transfer membrane was removed from the Blot module and placed in a clean container. The membrane was washed twice for 10 min with 10 mM Tris.Cl 150 mM NaCl pH:7.5 (TBS buffer) then incubated in 10 ml of 3% BSA in TBS buffer (blocking buffer) at room temperature by shaking. After blocking, membrane was washed twice with 50 ml of TBS-Tween/Triton buffer then once with TBS buffer and incubated with primary antibody, His-tag monoclonal antibody, (Novagen) for 1 h by shaking gently. The primary antibody was prepared in 5 ml of blocking buffer by adding 2.5 µl of antibody (200 ng/µl). The membrane was washed twice with TBS-Tween/Triton buffer to remove unbound antibody then washed once with TBS buffer for 10 min at each step. The membrane was incubated with the Anti-

Mouse IgG (Sigma, UK), Alkaline phosphatase produced in goat, as the second antibody for 1 h by shaking gently. The secondary antibody was prepared in 5 ml of blocking buffer by adding 2.5 µl of antibody. The membrane was washed 4X with TBS-Tween/Triton buffer and stained with BCIP/NBT solution (Sigma, UK) until clear visible bands were seen. The reaction was stopped by rinsing the membrane with distilled water. The membrane was dried and imaged. Washing and blocking steps were performed at room temperature.

2.3.5 Ni-NTA Purification

E. coli BL21 (DE3) cells expressing the protein of interest were grown overnight in 10 ml of L broth with antibiotic (final concentration of 100 µg/ml ampicillin for pET15b and 30 µg/ml kanamycin for pET28b). The overnight culture was used to inoculate prewarmed 250 ml of L broth with antibiotic at 37°C with 250 rpm shaking until an OD_{600} of 0.6 was reached. The culture was induced with IPTG at a final concentration of 1 mM IPTG. The cells were harvested by centrifugation for 20 min at 3200xg, 4°C. The supernatant was removed and the pellet was frozen overnight. The pellet was thawed on ice for 15 min, resuspended in 10 ml of 50 mM Tris pH 8.0 300 mM NaCl 1% Triton X-100 10 mM Imidazol 1mg/ml Lysozyme 25U/ml Benzonase nuclease (lysis buffer), incubated on ice for 30 min by swirling occasionally. The cell lysate was spun down for 30 min at 14000 x *g*, 4°C. The Ni-NTA column was prepared by adding 1 ml of Ni-NTA agarose resin slurry (QIAGEN, Germany) and letting resin to settle down on frit. The column was washed with 5 ml of 10 mM Tris pH 8.0 500 mM NaCl 10 mM Imidazol (wash buffer) without disturbing the resin. The cleared supernatant from centrifugation was removed and loaded on to the resin. All of the flow through was collected. The resin was washed with 5 ml of wash buffer twice and each wash was collected separately. Finally, the 6X His-tagged protein was eluted 3 times with 1 ml of 10 mM Tris pH 8.0 150 mM NaCl 200 mM Imidazol (elution buffer). The benzonase nuclease used in lysis buffer was purchased from Sigma-Aldrich.

2.3.6 Measurement of Myrosinase Activity by God-Perid Assay

God-Perid assay (Figure 2.5) was used to measure the glucose release upon glucosinolate or β-glucoside breakdown by myrosinase activity or β-O-glucosidase activity respectively [137]. The God-Perid reagent was prepared in 250 ml of Tris HCl buffer pH 7.2 (1.2 g/ml) consisting of 2-2'-azino-bis (3-ethylbenzthiazoline-6-sulphonic acid) (ABTS) (1 mg/ml), glucosidase

oxidase (8 U/ml) and peroxidase (0.35 U/ml). This reagent was stored in cold and dark place away from light until use. All chemicals were purchased from Sigma-Aldrich.

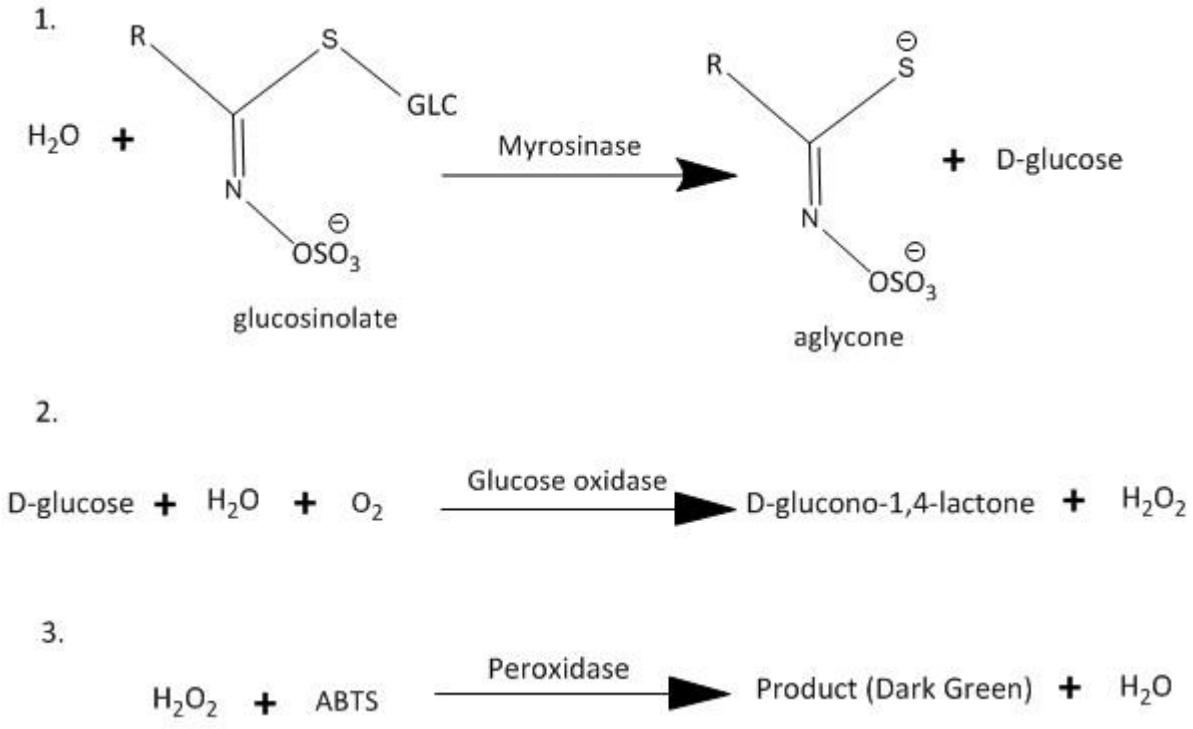


Figure 2.5 The principle of the God-Perid assay. The glucose (GLC) formed by myrosinase activity becomes the substrate of glucose oxidase and glucose oxidase releases H_2O_2 then H_2O_2 is catalysed by peroxidase forming a green product (ABTS radical cation). The green product gives an absorbance at 420 nm and can be measured by spectrophotometry.

The reaction mixture (300 µl), consisted of enzyme and 2 mM sinigrin in 0.1 M citrate phosphate buffer pH 7.0 (Buffer can change depending on enzyme used in the assay), was incubated aerobically for 30 min at 37°C. The mixture was boiled for 5 min to inactivate the enzyme and then 1 ml of God-Perid reagent added to the mixture and incubated for 15 min at 37°C. The mixture was transferred to a spectrophotometer cuvette and the absorbance at 420 nm was measured using 6715 UV/Vis Spectophotometer (JENWAY, UK). A calibration curve (Figure 2.6) was prepared using glucose (0.1 mg/ml-0.7 mg/ml) to quantify the glucose released by the enzyme activity.

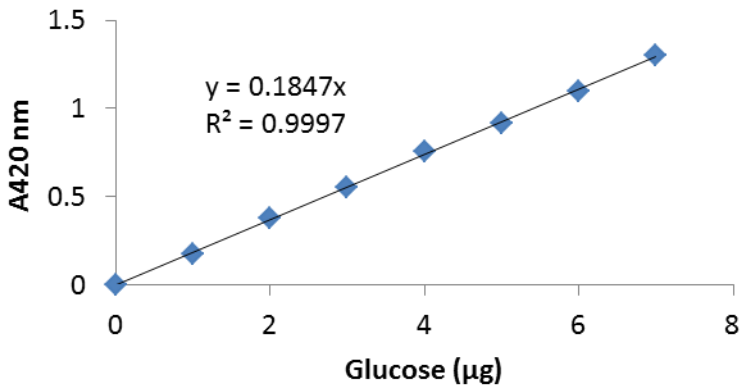


Figure 2.6 A reference calibration curve for God-Perid assay. Absorbance at 420 nm versus various amounts of glucose was plotted (n=3).

One unit of myrosinase activity was defined as the amount of enzyme liberating 1 µmol of glucose per min. Total activity and specific activity of enzyme can be determined from the formulae:

$$\text{Total activity} = \frac{\text{Average of } A_{420nm} \text{ x Dilution factor x F}}{\text{M of glucose x 30 min}} = \text{X µmol glucose/min}$$

$$\text{Specific activity} = \frac{\text{Total activity}}{\text{Total protein}} = \text{X µmol glucose/min/mg}$$

F; Gradient from calibration curve, M; molecular weight of glucose

2.4 CHROMATOGRAPHY METHODS

2.4.1 High Performance Liquid Chromatography (HPLC)

HPLC was used to detect the degradation rate of glucosinolates by myrosinases or reductases. The method needs a prior desulfation for glucosinolates.

2.4.1.1 Preparation of sulfatase

Sulfatase type H-1 from *Helix pomatia* was purchased from Sigma-Aldrich (UK). 300 mg of sulfatase was dissolved in 12 ml of cold MQ water then 12 ml of cold ethanol was added and stirred. Dissolved sulfatase was centrifuged at 4°C, 3000 x *g* for 6 min. The precipitate was

discarded and 1.5 volumes of cold ethanol was added to supernatant and stirred. Then centrifugation was performed as before

and the supernant was discarded. The precipitate was dissolved in 8 ml of MQ water and loaded on A25 Sephadex (GE Healthcare) layered reservoir then flow through was loaded on C25 Sephadex (GE Healthcare) layered reservoir and flow through was used in HPLC sample preparation.

2.4.1.2 Desulfation of glucosinolates

A needle and a frit was placed to a 3 ml capacity reservoir then the reservoir was pre-packed with 1 ml diethylaminoethyl (DEAE) Sephadex suspension and washed with 0.5 ml of MQ water. Each sample (200 µl) was mixed with 4.8 ml of 70% methanol with 50 µl of internal standard (16 mM sinigrin or glucotropaeolin), 3 ml of sample was loaded on pre-packed column and allowed to drip slowly. The column was washed twice with 0.5 ml 20 mM sodium acetate buffer pH 5.0 then twice with MQ water. The HPLC vials were placed under columns and 75 µl of sulfatase was layered on the surface of the columns. The columns were left overnight at room temperature. The next day, elution was performed with 1.25 ml MQ water in 3 steps (0.5 ml, 0.5 ml and 0.25 ml). The vials were sealed, vortexed and analysed on HPLC.

2.4.1.3 HPLC analytical conditions

Desulfonated glucosinolates were analysed by HPLC using a C18 reversed-phase column, Waters Spherisorb 5µM ODS2 4.6 x 250mm Cartridge. The column was fitted with Spherisorb ODS2 Guard Column, 80Å, 5 µm, 4.6mm X 10 mm (Waters, UK) and connected to an Agilent Technology 1100 pump series and an autosampler Agilent 1100 series. The samples (10 µl) were injected and HPLC runs were carried out using MQ water (solvent A), acetonitrile (Solvent B). A gradient program was used on HPLC: 5% (v/v) acetonitrile for 10 min, 35% (v/v) acetonitrile for 5 min, 40% (v/v) acetonitrile for 1 min, 50% (v/v) acetonitrile for 4 min 30 s, 100% (v/v) acetonitrile for 4 min 30 s and finally 5% (v/v) acetonitrile for 7 min. All runs were performed at 30°C with a flow rate of 1 ml/min. The desulfonated glucosinolates were detected at 229 nm using a Diode Array Detector (Agilent). Glucosinolates were identified using the retention times of known glucosinolate standards and quantified based on the peak areas of samples compared to peak areas of the internal standard. The response factors (Area for standard/Area for equimolar amount glucosinolate) [230] were used to determine

glucosinolate amount in the samples. The glucosinolate content of samples was determined for each glucosinolate (Table 2.10) and calculated as given below.

Amount of Glucosinolate (mol) = Area (Glucosinolate/Internal Standard) * RF *Amount

Glucosinolate	Relative Response Factor
Sinigrin	1
Glucoraphanin	1.07
Glucoerucin	1.04
Glucotropaeolin	0.95

Table 2.10 Response factors for desulfo-glucosinolates at 229 nm relative to desulfo-sinigrin.

For further fragmentation of glucosinolates, the same column and analytical conditions of HPLC with 0.1% formic acid in MQ water (solvent A), 0.1% formic acid acetonitrile (Solvent B) were used. The glucosinolates were detected using an Agilent 1100 LC/MSDF model mass detector Autosampler HP 1200 version.

CHAPTER THREE

3 MYROSINASES FROM HUMAN GUT BACTERIA INVOLVED IN GLUCOSINOLATE METABOLISM

3.1 INTRODUCTION

Consumption of cruciferous vegetables has been reported to provide health promoting effects against various cancers and this effect was attributed to glucosinolates and especially to their degradation products, ITCs [50, 69-72]. Myrosinase is the enzyme responsible for degradation of glucosinolates and formation of valued ITCs. Glucosinolates and myrosinases are kept separately from each other in the plants. Due to tissue damage, integrity of cells is disrupted and glucosinolates and myrosinase come into contact with each other. The hydrolysis of glucosinolates occurs and hydrolysis products such as ITCs, thiocyanates or nitriles are formed [8, 24]. In addition to plant myrosinases, many organisms show myrosinase activity including fungi [131], aphids [132] and gut bacteria [56, 112].

During cooking, plant myrosinases can be heat inactivated then biotransformation of glucosinolates to ITCs is dependent on human gut bacteria. Studies reported that the gut bacteria including *Enterococcus casseliflavus* CP1, *Escherichia coli* VL8 and *Lactobacillus agilis* R16 [56, 57] and *Bacteroides thetaiotaomicron* [112] can possess glucosinolate metabolism. While the plant [117, 119-121] and aphid myrosinases [132, 133] are well characterised, bacterial myrosinases are not. The first bacterial myrosinase was purified 43 years ago from *Enterobacter cloacae* but this study could not identify the gene or protein sequence [231]. There were attempts to identify a bacterial myrosinase with its amino acid sequence for years but they had a limited success [137, 232]. Recently, the very first bacterial myrosinase sequence was identified from a soil strain *C.* WYE1 by Albaser et al. (2016)[118]. This myrosinase was a periplasmic β-glucosidase from Glycosyl Hydrolase 3 (GH3) family with a signal peptide. A preliminary characterisation was performed and this myrosinase was found to be activated by ascorbate like plant myrosinases [118]. However, its crystal structure is yet to be achieved.

Luang-In et al. (2014) isolated several strains from the human gut using a sinigrin enrichment method and studied their glucosinolate degrading ability. The strains *E. casseliflavus* CP1, *E. coli* FI10944 and *L. agilis* R16 were found to degrade sinigrin and produce ITCs and nitriles. *E. casseliflavus* CP1 and *L. agilis* R16 were further studied and it was found that they degraded 53 ± 8% and 10 ± 2% of the glucoraphanin in the media after 24 h. When cell-free extracts of *E. casseliflavus* CP1 and *L. agilis* R16 were tested for myrosinase activity, the extracts did not show myrosinase activity. It was suggested that myrosinase activity of *E. casseliflavus* CP1 and *L. agilis* R16 is cell wall-associated [56]. 2D gel electrophoresis was used to identify sinigrin-

induced bacterial proteins of *E. coli* VL8 and *L. agilis* R16 involved in glucosinolate metabolism. It was concluded that induced proteins mostly belong to carbohydrate metabolism, oxidoreduction and sugar transport system [138].

Myrosinase activity in the human gut bacteria is attributed to the β-glucosidases which also have specificity for glucosinolates. A bacterial myrosinase from human gut has yet to be identified in terms of gene identification. In previous studies, candidate genes from *E. casseliflavus* CP1 [137] and *Lactobacillus* strains [232] which were proposed to have myrosinase activity were cloned and expressed in *E. coli* to identify the myrosinase. However, the recombinant proteins of *Lactobacillus* species from GH1 family (http://www.cazy.org/)[233] were found to be insoluble and no myrosinase activity was determined. In addition, the recombinant proteins of *E. casseliflavus* CP1 (both from GH1 and GH3 family) failed to show myrosinase activity. Some of the recombinant proteins of *E. casseliflavus* CP1 were found to show β-glucosidase activity against various glucosides. Luang-In et al. (2013) performed their study lacking the annotated genome of *E. casseliflavus* CP1, the Uniprot database was used to get information on the β-glucosidases of *E. casseliflavus* CP1. As the genome of *E. casseliflavus* CP1 was sequenced, assembled and annotated in this study, this revealed that many putative β-glucosidases (40 β-glucosidases) are found in the *E. casseliflavus* CP1 genome. It was thought that Luang-In et al. (2013) had not cloned the correct β-glucosidase with myrosinase activity [137]. In this context, *E. casseliflavus* CP1 and *E. coli* FI10944 were studied in this chapter to identify their myrosinase. Cloning and chromatography approches were used for identification.

3.2 HYPOTHESIS

Specific β-glucosidases are responsible for degradation of glucosinolates by human gut bacteria and can be identified using molecular biology or chromatography techniques.

3.3 MATERIALS AND METHODS

3.3.1 Bacterial Strains

The bacterial strains that are mentioned in this chapter were given in Table 3.1.

Bacterial Strain	Source of the strain
L. agilis R16 [56]	Human gut (provided to us)
E. coli FI10944	Human gut (provided to us)
E. casseliflavus CP1 [56]	Human gut (provided to us)
C. freundii FC50	Human gut (isolated in this study)
C. WYE1 [118]	Soil (provided to us)

Table 3.1 Bacterial strains used in the study.

3.3.2 Genome Assembly of Glucosinolate Degrading Bacteria

Bacterial cultures (10 ml) in appropriate culture broth were grown aerobically at 37°C until the culture reached an optical density OD_{600}=1.0. Genomic DNA extraction was performed according to QIAGEN Genomic DNA Protocol (QIAGEN, 2001) and genomic DNA of strains was sequenced and assembled as described in Chapter 2-section 2.2.13.

3.3.3 Screening Glucosinolate Metabolism by Human Gut Bacteria

E. casseliflavus CP1, *E. coli* FI10944 and *L. agilis* R16 were reported to show glucosinolate degradation and ITC formation previously [56]. These strains were checked for their glucosinolate degradation ability in this study to confirm they are still able to degrade glucosinolates. The cultures were grown in the same way as described previously by Luang-In et al. (2014)[56]. Human gut bacteria were cultured in 5 ml of Wilkinson Chalgren (*E. casseliflavus* CP1), MRS-glucose media (*L. agilis* R16) and Nutrient Broth (*E. coli* FI10944) containing 1 mM glucoraphanin at 37°C anaerobically for 24 h. Negative controls consisted of only glucoraphanin incubated in appropriate media for each bacteria. After incubation, cells and negative controls were centrifuged for 15 min at 20000 x *g* at 4°C. The supernatants were removed and transferred into clean tubes. HPLC samples were prepared from these supernatants (200 µl used) and analysed as described before in Chapter 2-section 2.4 to determine glucosinolate degradation. The degradation products were not characterised because the main focus of this part of the study was to confirm that they were still showing the ability to degrade glucosinolates.

E. coli FI10944 was tested for glucosinolate metabolism under different conditions. Instead of using a rich medium, it was grown in M9 minimal media (recipe in Chapter 2-section 2.1.1) with glucoraphanin or sinigrin as sole carbon source. The growth of *E. coli* FI10944 was screened over 16 h by Bioscreen C (Labsystems), also testing the effect of different glucosinolate concentrations on growth. The addition of more glucosinolate after 4th hour was tested to understand the consumption of glucosinolates by cells.

It was also examined if the myrosinase activity of *E. coli* FI10944 cell-free extracts (CFEs) was inducible. *E. coli* FI10944 was grown in 10 ml of Nutrient Broth with or without 1 mM glucoraphanin anaerobically overnight at 37°C then the cultures were used to prepare CFEs by sonication. Cell pellets were washed twice with 0.1 M citrate phosphate buffer pH 7.0 to remove any traces of glucoraphanin from culture in sonication step. Protein concentrations of CFEs were determined by Bradford assay. The CFEs were incubated in 1 ml reaction volume with 0.25 mM glucoraphanin in 0.1 M citrate phosphate buffer pH 7.0 anaerobically for 16 h at 37°C. After incubation, samples were centrifuged at 16000 x *g* for 15 min at 4°C and supernatants were transferred into clean tubes. The supernatants (200 µl) were used in sample preparation for HPLC to determine glucoraphanin degradation by CFEs.

L. agilis R16 was examined to see if pre-culturing in MRS media with sinigrin or glucoraphanin would affect its myrosinase activity and whether this affect would be dependent on glucosinolate type. *L. agilis* R16 cells were pre-cultured in 5 ml of MRS-glucose, MRS with 1 mM glucose, MRS with 1 mM sinigrin, MRS with 1 mM glucoraphanin media anaerobically at 37°C overnight. These pre-cultures were used to inoculate 1 ml of MRS with 1 mM sinigrin, MRS with 1 mM glucoraphanin media. The rest of the MRS with 1 mM sinigrin and MRS with 1 mM glucoraphanin pre-cultures were centrifuged 13000 x *g* for 15 min at 4°C. Clear supernatants (200 µl) were used in sample preparation for HPLC analysis to determine glucosinolate concentration left after incubation. The cultures inoculated from the pre-cultures were incubated anaerobically for 24 h at 37°C. Controls without inoculum were also set up. Later, cultures were centrifuged at 13000 x *g* for 15 min at 4°C. Clear supernatants were transferred to clean tubes and 200 µl of supernatants used in sample preparation for HPLC analysis to determine glucosinolate degradation. While 16 mM sinigrin was used as internal standard for HPLC to determine glucoraphanin degradation, it was replaced with 16 mM glucotropaeolin for sinigrin degradation.

In addition to these strains, 98 new strains from human gut were isolated by glucoraphanin enrichment in this study (Chapter 2-section 2.1.3). Fifteen of them degraded 100% of glucoraphanin in the media. They were identified by 16S rDNA sequencing then results were analysed on The Ribosomal Database Project (RDP) database [234] (Method is in Chapter 2-section 2.2.12). One of the isolates was further studied. This strain, *C. freundii* FC50 was further screened for its growth in M9 minimal media including glucoraphanin or sinigrin. To test the myrosinase activity of CFEs of *C. freundii* FC50, it was grown in 5 ml Nutrient Broth (NB) with or without 1 mM sinigrin and CFEs were prepared as described before. CFEs were quantified by Bradford assay (Method is in Chapter 2-section 2.3.2) then myrosinase activity of CFEs was examined by God-Perid assay (Chapter 2-section 2.3.6).

3.3.4 Identification of β-glucosidases to Clone and Express

The genomes of *E. casseliflavus* CP1 and E. coli FI10944 were sequenced, assembled and annotated as described in Chapter 2-section 2.2.13. As bacterial myrosinases are suggested to be β-glucosidases and more closely related to aphid myrosinases than plant myrosinases [133, 137], all β-glucosidases in the genome of *E. casseliflavus* CP1 and *E. coli* FI10944 were interrogated for significant similarity to aphid myrosinases and a recently identified bacterial myrosinase, *C.* WYE1 myrosinase [118]. The candidate genes were selected and checked for conserved domains then best candidates were cloned and expressed in *E. coli*.

3.3.5 Cloning, Expression of β-glucosidases and Purification of Proteins

The selected genes were cloned using the pET expression system (Novagen). Genomic DNA of *E. casseliflavus* CP1 and *E. coli* FI10944 were extracted by using the Genomic DNA extraction kit with Genomic Tip 20/G columns (QIAGEN) as advised by the supplier. The primers introducing new restriction sites were designed to amplify the candidate genes from *E. casseliflavus* CP1 and *E. coli* FI10944 genome are listed in Table 3.2. When pET28b vector was used, primers were designed to enable in-frame translation of C-terminal 6XHis-tag by insertion of new nucleotides if this was needed. The nucleotide sequences of the cloned genes are shown in Appendix 2. PCR was set up using 5 ng or 10 ng genomic DNA as template and PCR products were checked on agarose gel (Chapter 2-section 2.2.1 and 2.2.2). The annealing temperatures are also given in Table 3.3.

Name	Primer Sequence (5'-3')	Restriction Site
ECG44-F	GAAACA**TA**TGGATCATAAACAACTAAA	NdeI
ECG44-R	GTTTCTC**GAG**GCTAGCACTCTTG	XhoI
ECG4-F	GAATA**CATA**TGGAGACCGTAAAACTAAA	NdeI
ECG4-R	GAAAG**AA**TT**C**AAAATCGTGAGCGTACC	EcoRI
ECG39-F	GGAGATCA**T**ATGAAAATTGATGTGAAT	NdeI
ECG39-R	AACCC**T**CG**AG**A**T**AAACAGTTATAGTTTCTTT	XhoI
ECOLG3-F	GAGGAAGAAAAATCCATATGAAATG	NdeI
ECOLG3-R	AAAG**G**A**A**T**T**CAGCAACTCAAACTCG	EcoRI

Table 3.2 Primers used in the PCR experiments to amplify the candidate β-glucosidase genes . The restriction sites are underlined. The altered nucleotides to introduce the restriction sites are in bold. The inserted nucleotides are highlighted.

Name	T_A, time, cycles	T_E, time	Product Size
ECG44-F ECG44-R	51°C, 30 s, 5X 63°C, 30 s, 20X	72°C, 40s	1436 bp
ECG4-F ECG4-R	55°C, 30 s, 5X 65°C, 30 s, 20X	72°C, 1 min	2162 bp
ECG39-F ECG39-R	50°C, 30 s, 5X 67°C, 30 s, 20X	72°C, 1 min	2101 bp
ECOLG3-F ECOLG3-R	54°C, 30 s, 5X 66°C, 30 s, 20X	72°C, 1 min	2321 bp

Table 3.3 Conditions used in the PCR to amplify the candidate β-glucosidase genes. T_A; annealing temperature, cycles; cycles used in annealing step of PCR, T_E; extension temperature and time used in PCR, Product size; the expected PCR product sizes for β-glucosidase genes. The annealing step was performed using T_m of 100% sequence match part of primers first (5X) and then using the Tm of entire primer (20X).

The insert was ligated to double digested pET15b or pET28b expression vector and gene constructs were transformed into *E. coli* Top10 competent cells (Thermo-Fischer Scientific) for *ecg*44 and *E. coli* DH5α chemically competent cells (homemade, See Chapter 2-section 2.2.9)

for *ecg*4, *ecg*39 and *ecolg*3. Transformant colonies were screened by colony PCR using GoTag polymerase and primers T7P2 (5'-TGAGCGGATAACAATTCCC) and T7T (5'-GCTAGTTATTGCTCAGCGG) (See Chapter 2 section 2.2.11 for method). The clones that gave positive results for PCR were selected for plasmid purification. These clones were inoculated in L Broth with ampicillin (100 μg/ml final concentration) or L Broth with kanamycin (30 μg/ml final concentration) and grown overnight at 37°C, with shaking at 250 rpm. The recombinant plasmids were purified using EZNA Plasmid Mini Kit II (Omega Bio-Tek) and then sequenced. T7P2 and T7T primers were used to sequence pET15b-ecg44 construct. Due to the larger size of the other recombinant plasmids, internal primers were designed to sequence the whole gene. The internal primers used in sequencing of pET28b-ecg4, pET28b-ecg39 and pET28b-ecolg3 constructs were ECG4_I (5'-GACCTCCTTTAATACGATCCA), ECG39_I (5'-GCTTTGTCATTAGTGATTTCGT) and ECOLG3_I (5'-GATTATATGCCGCCGTACA) respectively. The plasmids used and created in this study are shown in Table 3.4. The sequences were compared with the expected sequences to confirm that there was no mutation in the cloned genes. After that, the plasmid was transformed into *E. coli* BL21(DE3) cells (One Shot competent cells, Invitrogen). The colonies were picked and checked by colony PCR. The positive colonies were grown overnight in L Broth plus antibiotic then overnight grown cultures were used to prepare bacterial stocks. Bacterial stocks were kept at -80°C and one of the positive clones was further used in protein expression and induction studies.

Plasmid	Relevant Characteristics	Antibiotic Resistance
pET15b	Expression vector	Ampicillin
pET28b	Expression vector	Kanamycin
pET15b-ecg44	*ecg44* gene insert from *E. casseliflavus* CP1 ligated to pET15b	Ampicillin
pET28b-ecg4	*ecg4* gene insert from *E. casseliflavus* CP1 ligated to pET28b	Kanamycin
pET28b-ecg39	*ecg39* gene insert from *E. casseliflavus* CP1 ligated to pET28b	Kanamycin
pET28b-ecolg3	*ecolg3* gene insert from *E. coli* FI10944 ligated to pET28b	Kanamycin

Table 3.4 Plasmids used and generated in this study.

Protein expression in *E. coli* BL21(DE3) encoding the cloned genes was induced with IPTG at a final concentration of 0.5 mM for different time intervals. For *E. coli* BL21(DE3) cells expressing pET15b-ecg44, inductions were performed for 1 h, 2 h, 3 h, or 4 h at 37°C, shaking at 250 rpm. Protein expression in *E. coli* BL21(DE3) encoding pET28b-ecg4, pET28b-ecg39 and pET28b-ecolg3 were induced at 37°C for 3 h or 25°C for 4 h. Control cultures with *E. coli* BL21(DE3) pET15b (or pET28b) grown in presence of 0.5 mM IPTG (empty vector controls) and control cultures of *E. coli* BL21(DE3) expressing the cloned genes grown without 0.5 mM IPTG (uninduced cultures) were always included. After each induction period, the cells were centrifuged at 3220 x *g* for 15 min at 4°C. The supernatant was removed and cell paste was frozen at -20°C. The CFEs were prepared by sonication as described in Chapter 2-section 2.3.1 and run on a SDS-PAGE gel (denaturing, reducing conditions) including negative controls too. The CFEs were checked for the presence of 6XHis-tag by Western Blot (Chapter 2-section 2.3.4) and partially purified by Ni-NTA columns as described in Chapter 2-section 2.3.5.

3.3.6 Dialysis of the Recombinant Proteins

Purified proteins were dialysed overnight to remove excessive amounts of imidazole from Ni-NTA column purification (Method is in Chapter 2-section 2.3.5). Spectra/Por porous membrane tubing with 500-1000 Da cut off (Spectrum Labs) was used for dialysis. The membrane tubing was soaked in MQ water for 30 min in advance and rinsed with MQ water before use. The dialysis was performed in an appropriate amount of buffer (at least 100X bigger than volume of the protein) at 4°C for 18 h. The proteins were removed from the tubing and transferred into new clean tubes. The dialysed proteins were quantified by Bradford assay and tested by God-Perid assay for myrosinase activity. In addition, dialysis was performed in ammonium sulphate precipitation steps to exchange buffer.

3.3.7 Measurement of Myrosinase Activity

God-Perid assay was used to measure the glucose release upon glucosinolate or β-glucoside breakdown by myrosinase activity or β-glucosidase activity respectively [137]. The method was described in Chapter 2-section 2.3.6 in detail. The reaction mixture (300 µl), consisted of enzyme and 2 mM sinigrin in 0.1 M citrate phosphate buffer pH 7.0 was incubated aerobically for 30 min at 37°C. The buffer used in the assay can change depending on the enzyme. The mixture was boiled for 5 min to inactivate the enzyme and then 1 ml of God-Perid reagent added to the mixture and incubated for 15 min at 37°C. The mixture was transferred to a

spectrophotometer cuvette and the absorbance at 420 nm was measured. A calibration curve was prepared using glucose to quantify the glucose released by the enzyme activity.

The effect of ascorbic acid on myrosinase activity of a recombinant protein, ecg44, was tested using God-Perid assay. The reactions were set up in 300 µl of volume consisting of 2 mM sinigrin, 5 µg of ecg44 and 1 mM ascorbic acid in 0.1 M citrate phosphate buffer pH 7.0. To test the effect of different cofactors and the amount of enzyme, the assays were set up using 10, 50, 200 µg of ecg44 with or without 10 mM $MgCl_2$ or $ZnCl_2$.

3.3.8 β-glucosidase Assay Using 4-nitrophenyl β-D-glucopyranoside as Substrate

This assay measures transformation of 4-nitrophenyl β-D-glucopyranoside (p-NPG) into the product p-nitrophenol (p-NP) that absorbs light at 405 nm [235].

p-NPG (COLOURLESS) ⟶ **d-glucose + p-NP (YELLOW)**

The samples were prepared in 100 µl of sodium phosphate buffer pH 8 containing 4 mM 4-Nitrophenyl β-D-glucopyranoside and incubated for 30 min at 37°C then 100 µl of 1 M Na_2CO_3 was added to stop the enzymatic reaction. To determine the enzyme activity, a standard curve was prepared using 5-50 nmol of p-nitrophenol in 100 µl sodium phosphate buffer pH 8 (Figure 3.1). The enzyme assays were set up in 96 well plates. The absorbance was measured using a ThermoMax Plate Reader (GMI, US) at 405 nm. One activity unit was defined as the amount of activity generating 1 µmol p-nitrophenol per minute.

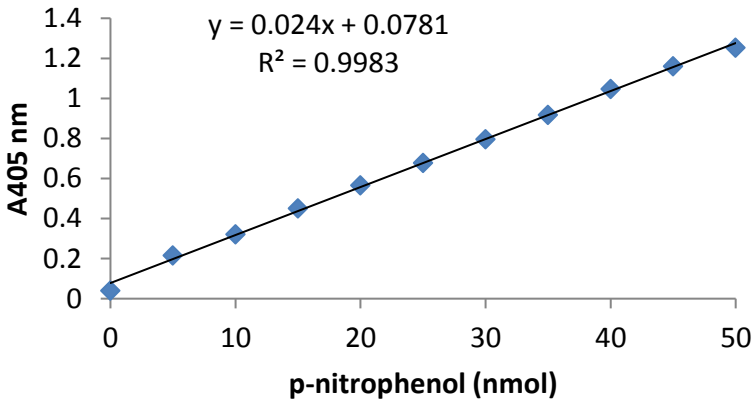


Figure 3.1 A reference calibration curve for β-glucosidase assay. Absorbance at 405nm versus various amounts of p-nitrophenol was plotted (n=3).

3.3.9 Purification of the Bacterial Myrosinase using Fast Protein Liquid Chromatography (FPLC)

The method was adapted from a previous study [231]. Purification attempt 1: *E. coli* FI01944 was grown in 500 ml of Nutrient Broth with 1 mM sinigrin at 37°C overnight (OD_{600}= 1.7) then the protein extract prepared from this overnight grown culture by $(NH_4)_2SO_4$ precipitation. The overnight grown culture was spun down at 18000 x *g* for 20 min at 4°C. The cell pellet was resuspended in 0.2 M sodium phosphate buffer pH 7.0 (9 ml/ g cell paste) with 1X protease inhibitor cocktail (100X concentrate, Melford). The cells were sonicated for 45 s and kept on ice for 1 min to cool down and this step was repeated 7 times. After sonication, the suspension was centrifuged at 13000 x *g* for 25 min at 4°C. The supernatant was removed as soluble CFE. CFE was saturated with $(NH_4)_2SO_4$ at 80% (w/v) saturation at 4°C then centrifuged at 3220 x *g* for 15 min at 4°C. The supernatant was removed and the precipitate was dissolved in 30 ml of 0.2 M sodium phosphate buffer pH 7.0 as a protein extract. The protein extract was dialysed against 10 mM sodium phosphate buffer pH 7.0 for 18 h at 4°C. This protein extract was used in Ion Exchange (IEX) Chromatography to purify the bacterial myrosinase of *E. coli* FI10944.

Anion exchange chromatography was performed using Resource Q 1 ml column (GE Healthcare) attached to an ÄKTA FPLC (GE Healthcare) equipped with a UPC-900 wavelength detector (GE Healthcare) using Unicorn 3.20 software (GE Healthcare). The buffers used in IEX chromatography were 10 mM Na-phosphate buffer pH 7.0 as starting buffer and 10 mM Na-phosphate plus 1 M NaCl buffer pH 7.0 as elution buffer. These buffers were prepared freshly and filtered through 0.22 µm Millex-GS filters (Merck Millipore). The column was equilibrated with starting buffer at 4 ml/min flow rate and then samples were run at 4 ml/min flow rate and the fractions (1.25 ml) were collected. The fractions from IEX chromatography were tested for myrosinase activity by God-Perid Assay using sinigrin as the substrate. The fractions showing myrosinase activity were further purified by gel filtration (GF). GF was performed with a Superdex 75 10/300 GL column (GE Healthcare). The column was attached to ÄKTA FPLC (GE Healthcare) equipped with a UPC-900 wavelength detector (GE Healthcare) using Unicorn 3.20 software (GE Healthcare). The buffer used was 10 mM Na-phosphate 150 mM NaCl pH 7.0 and it was filtered using a Millipore 0.22 µm GSTF membrane filter before use. The column was equilibrated with this buffer. Following column equilibration, the proteins were eluted and the fractions (1.25 ml) were collected. Collected fractions were tested for

myrosinase activity by God-Perid assay and the fractions with myrosinase activity were determined. These fractions from IEX and GF with myrosinase activity were analysed by LCMS/MS to determine the protein sequences.

Purification attempt 2: The buffer conditions for myrosinase activity were optimised using different buffers in CFE preparation. CFEs were prepared from overnight grown 10 ml cultures of Nutrient Broth by sonication as reported before. CFEs of *E. coli* FI10944 in various buffers were prepared including 20 mM Na Citrate Buffer (pH 3.0, 4.0, 5.0 and 6.0), sodium citrate phosphate buffer (pH 5.0, 6.0 and 7.0), Tris-HCl buffer (pH 7.0, 8.0 and 9.0), sodium phosphate buffer (pH 6.0, 7.0 and 8.0). The protein concentrations of CFEs were determined by Bradford Assay and the myrosinase activity was determined by God-Perid assay and optimum buffer was determined for myrosinase activity. *E. coli* FI01944 was grown in 500 ml of Nutrient Broth without glucosinolates at 37°C overnight (OD_{600}= 1.9). The overnight grown culture was centrifuged at 18000 x *g* for 20 min at 4°C. The cell pellet was resuspended in 20 mM Na citrate buffer pH 5.0 (9 ml/ g cell paste) with 1X protease inhibitor cocktail (100X concentrate, Melford). CFE was prepared by sonication as described in the first purification attempt. The CFE was saturated with $(NH_4)_2SO_4$ at 80% (w/v) saturation at 4°C then centrifuged at 3220 x *g* for 15 min at 4°C. The supernatant was removed and the paste was dissolved in 20 mM Na citrate buffer pH 5.0. This protein extract was dialysed against 20 mM Na citrate buffer pH 5.0 overnight at 4°C to remove $(NH_4)_2SO_4$. The dialysed protein extract was run on FPLC to perform GF. GF was performed with a Superdex 75 10/300 GL column (GE Healthcare). Sample was run on gel filtration using 20 mM Na-citrate plus 200 mM NaCl buffer pH 5.0. The buffer was prepared freshly and filtered through Millipore 0.22 µm GSTF membrane filter before use. The sample was run with 0.5 ml/min flowrate and fractions (0.5 ml) were collected. The protein eluates were monitored at 280 nm. The fractions showing myrosinase activity were analysed by LCMS/MS to determine the protein sequences.

3.3.10 Identification of the Protein Sequences Obtained from FPLC by LCMS/MS

The fractions showing myrosinase activity obtained from FPLC were run on SDS-PAGE gels under denaturing, reducing conditions as described in Chapter 2-section 2.3.3. Gel pieces of interest were excised and washed with 1 ml of 200 mM ammonium bicarbonate (ABC) in 50% (v/v) acetonitrile to equilibrate the gel to pH 8 and remove the stain for 15 min twice. This solution was removed and the gel was washed with 1ml of acetonitrile (Fisher Scientific). After acetonitrile wash, solution was removed and left to air dry to remove any remaining

acetonitrile. The samples were incubated with 10 mM dithiothreitol in 50 mM ABC for 30 min at 60°C to reduce any cysteine thiol side chains then the solution was removed. The samples were alkylated with 1 ml of 100 mM iodoacetamide in 50 mM ABC in the dark at room temperature for 30 min. The gel pieces were then washed with 1 ml of 200 mM ABC in 50% (v/v) acetonitrile for 15 min twice followed by leaving them in 1 ml of acetonitrile to dehydrate and shrink the gel pieces. The protein sample was digested by adding 100 ng of trypsin (modified porcine trypsin; Promega), in 10 µl of 10 mM ABC before incubation at 37°C for 3 h. After digestion the samples were acidified with 10 µl of 1% (v/v) formic acid for 10 min. The digest solution was transferred to a clean tube and the gel pieces were then washed with 50% acetonitrile and agitated by tapping the tube and incubated for 10 min. Samples were frozen using dry ice and stored until mass spectrometry analysis [236]. Sample preparation and LCMS/MS run were performed by Dr. Fran Mulholland (IFR).

The approach of LCMS/MS method and Mascot database (http://www.matrixscience.com/) is to take a sample of the protein of interest and digest it with a proteolytic enzyme, trypsin and then analyse the digest mixture by mass spectrometry. An instrument with MS/MS capability provides information about the structure by recording the fragment ion spectrum of the peptide. The experimental mass values are compared with calculated peptide mass or fragment ion mass values. An appropriate scoring algorithm is used to identify the closest match or matches. If the protein of interest is present in the sequence database, then the database pulls out that precise entry. If the sequence database does not contain the protein, then the database finds out those entries which exhibit the closest homology.

3.4 RESULTS

3.4.1 Myrosinase Activity of Human Gut Bacteria

In this chapter we mainly focused on *E. casseliflavus* CP1, *E. coli* FI10944 and *L. agilis* R16 which were reported to have myrosinase activity previously by Dr. Vijitra Luang-In [137]. In addition to these strains, 98 bacterial strains with myrosinase activity were isolated by a glucoraphanin enrichment method. Ten of these isolates were identified at genus level by 16S rDNA sequencing and 5 isolates (Isolate 23, 32, 35, 50 and 57) were found to belong *Enterobacteriaceae* family. The results are shown in Table 3.5 and RDP database results of isolates are given in Appendix 3. Because of time limitation, not all of these strains could be

further studied. One of the isolates, isolate 50, was selected for further investigation of its glucosinolate degrading ability. Its genome was sequenced, assembled and annotated. The results from 16S rDNA and genome sequencing revealed that it was a strain of *C. freundii* and it was designated as *C. freundii* FC50.

Isolates	**Genus**
59, 61, 66, 71	*Enterococcus*
33, 36, 48, 49, 58, 60	*Escherichia/Shigella*

Table 3.5 Identities of 10 human gut isolates at genus level. The isolates were obtained using a glucoraphanin enrichment method.

When *E. casseliflavus* CP1, *E. coli* FI10944 and *L. agilis* R16 (SEM images are shown in Figure 3.2) were cultured from previously frozen stocks, they were still able to degrade glucosinolates in *in vitro* conditions. The samples prepared from cultures were analysed on HPLC and the glucoraphanin degradation by cells was determined compared to negative controls which include only glucoraphanin incubated in appropriate media. According to the results, *E. casseliflavus* CP1, *L. agilis* R16 and *E. coli* FI10944 used up 7.3 ± 2.0%, 12.4 ± 0.8% and 100% of the glucoraphanin supplied to the media respectively (HPLC chromatograms are shown in Figure 3.3).

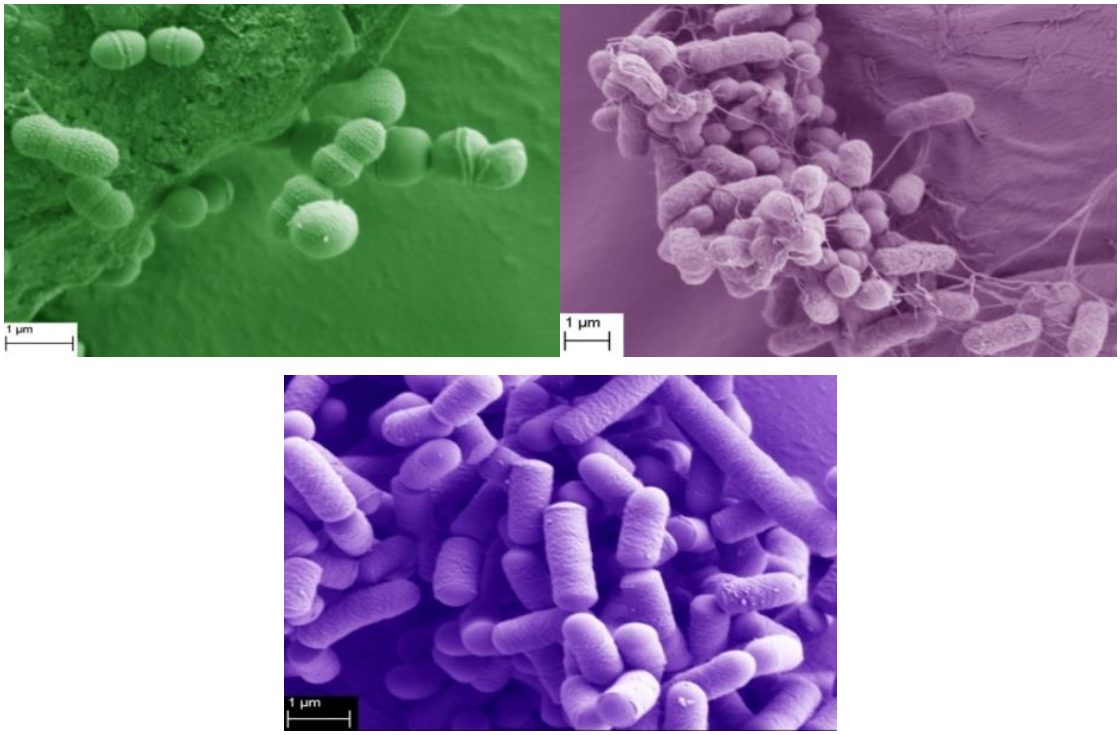


Figure 3.2 SEM images of *E. casseliflavus* CP1 (up left), *E. coli* FI10944 (up right) and *L. agilis* R16 (below). Scanning electron microscopy was performed using a Zeiss Supra 55 VP FEG SEM (Zeiss, Germany) operated at 3 kV by Rachael Stanley at Institute of Food Research.

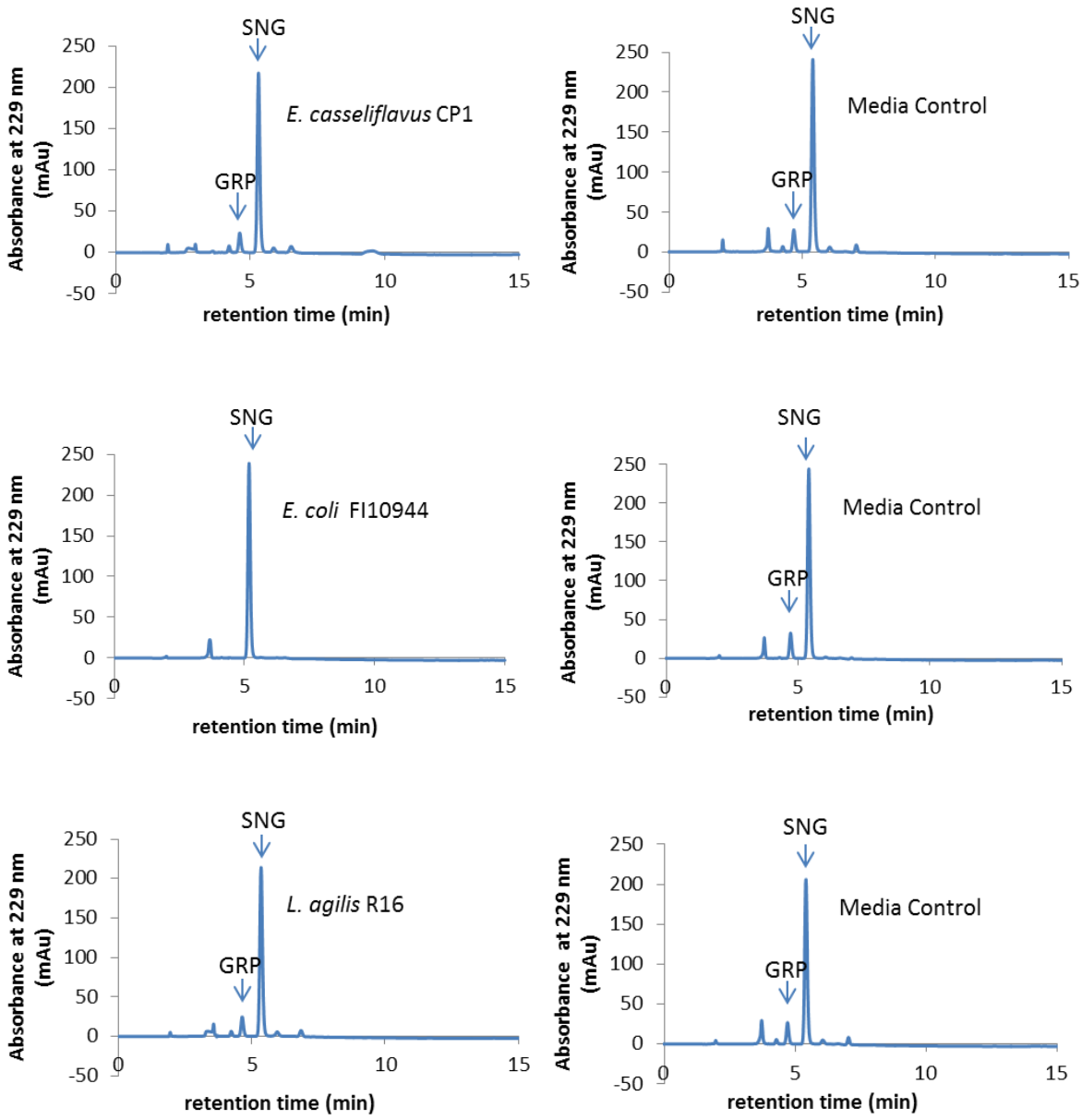


Figure 3.3 HPLC chromatograms of hydrolysis of glucoraphanin by *E. casseliflavus* CP1, *E. coli* FI10944 and *L. agilis* R16. SNG; sinigrin, GRP; glucoraphanin. Sinigrin was used as internal standard. Sinigrin retention time (5.3-5.4 min), glucoraphanin retention time (4.5-4.7 min). Medium controls consisted of glucoraphanin incubated in appropriate media for each bacterium.

The growth of *E. coli* FI10944 was analysed over time. The results are shown in Figure 3.4. This growth pattern shows that sinigrin is being used as carbon source by *E. coli* FI10944 (Figure 3.4 a, b, d). General media have lots of different carbon sources but sinigrin or glucoraphanin was the sole carbon source in these experiments. This explains the low growth rate of *E. coli* FI10944 in M9 minimal media with sinigrin compared to growth rate previously

achieved in general media. These experiments showed that *E. coli* FI10944 reached the highest OD_{600} in ~4 h (Figure 3.4 a, b). With the same amount of glucose in the media instead of glucosinolate, the culture could only reach an OD_{600} of ~0.55 but addition of more glucose at 4th hour achieved a higher growth rate indicating the carbon source is the growth limiting factor (Figure 3.4 c). This result led the study to try the addition of more sinigrin after 4th hour (Figure 3.4 d). This effect almost doubled the growth and showed the availability of free glucosinolate (sinigrin) is the growth limiting factor for *E. coli* FI10944.

a

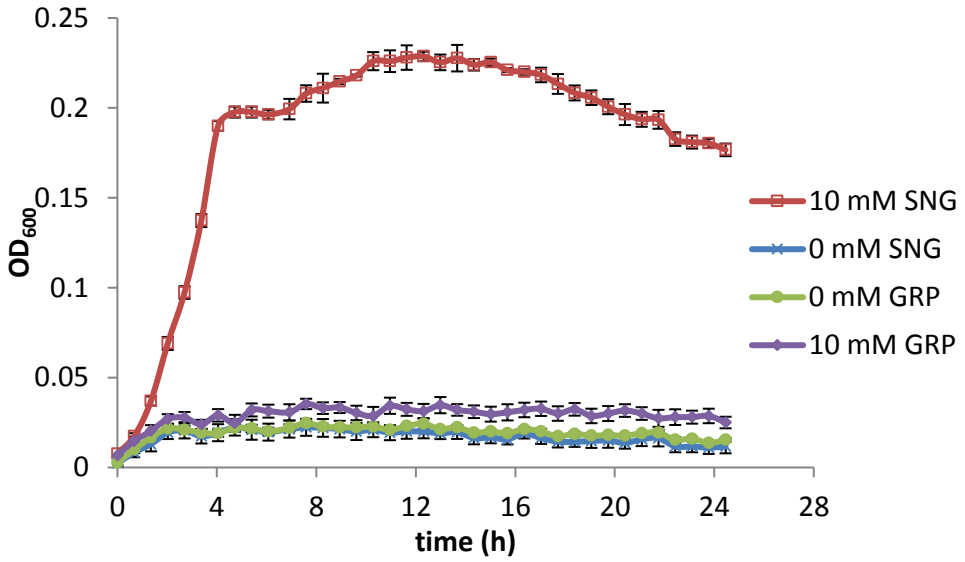
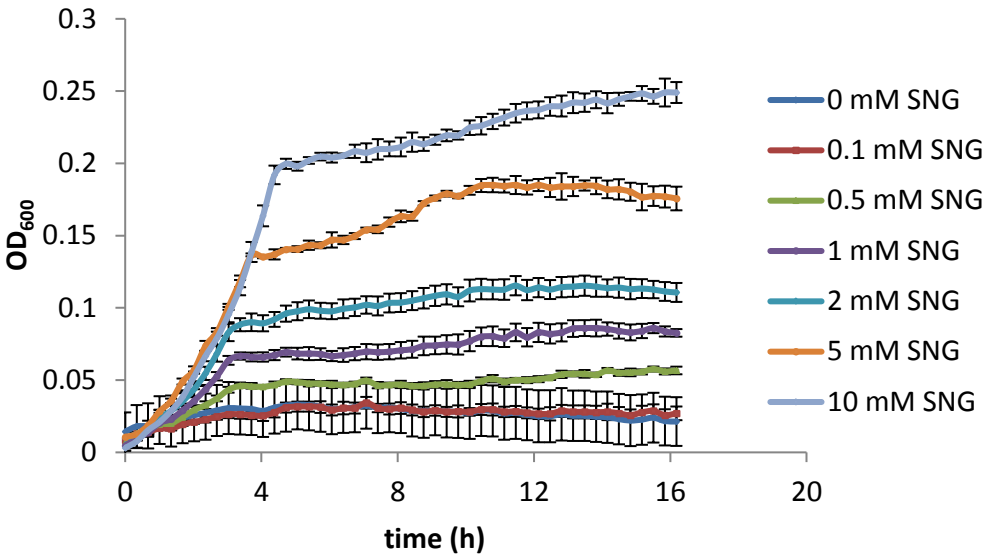


Figure 3.4 Growth curves of *E. coli* FI10944 in M9 minimal media at different growing conditions. a) Growth curve at 37°C for 24 h in M9 minimal media with sinigrin (SNG) or glucoraphanin (GRP) as sole carbon source. b) Growth curve at 37°C for 16 h in M9 minimal media with sinigrin (SNG) at different concentrations (0-10 mM). c) Growth curve at 37°C for 16 h in M9 minimal media with 10 mM glucose to test the effect of extra glucose on growth rate. d) Growth curve at 37°C for 16 h in M9 minimal media with 10 mM sinigrin (SNG) to test the effect of extra sinigrin addition at 4th hour.

b



c

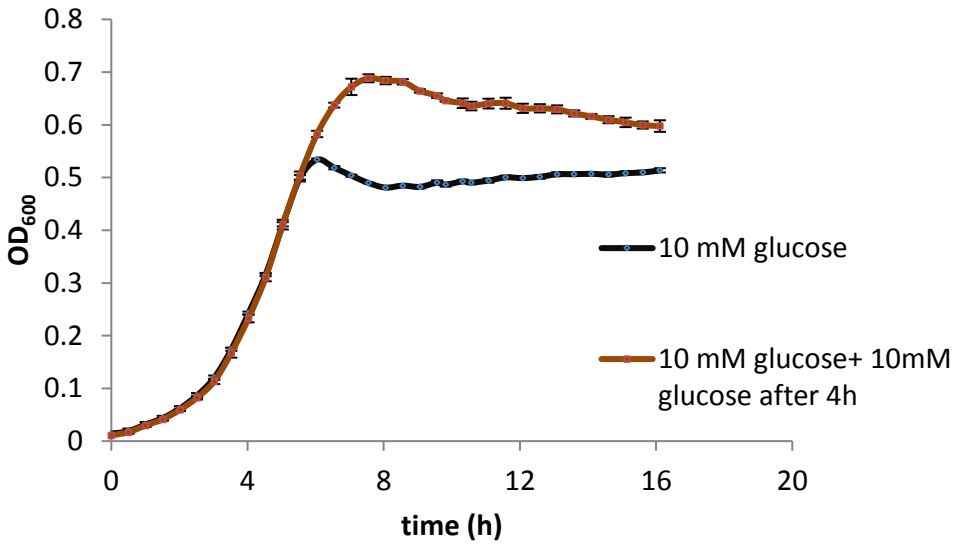


Figure 3.4-Continued

d

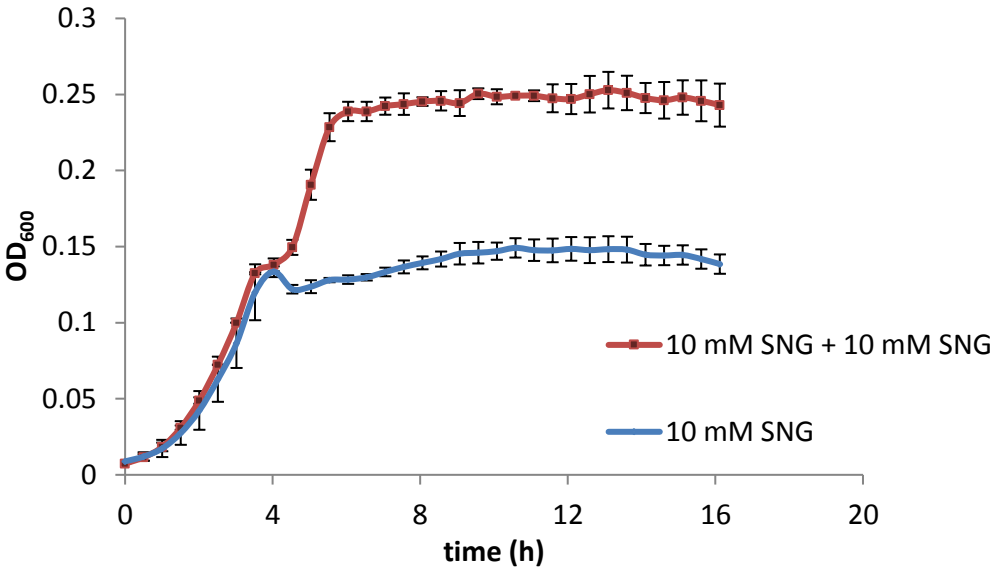
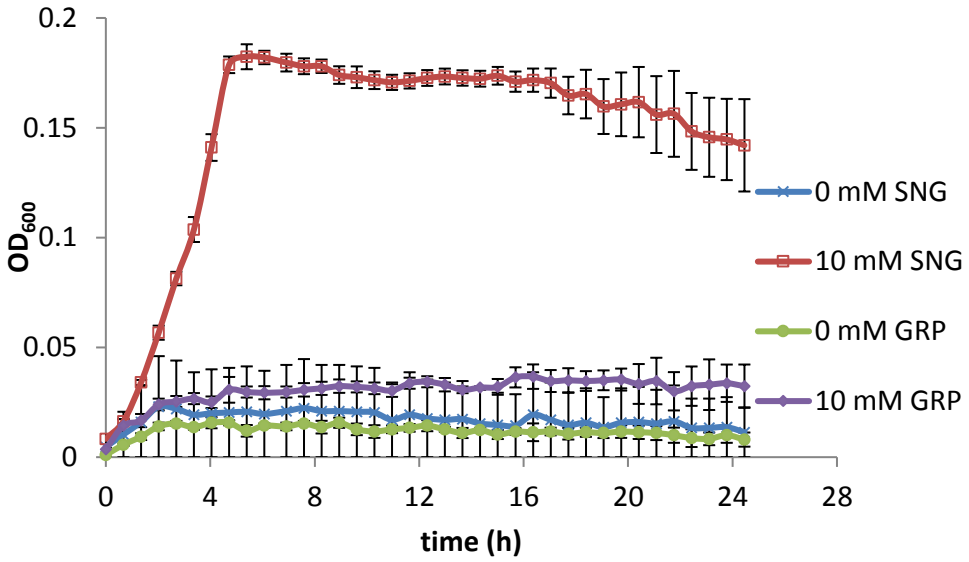


Figure 3.4-Continued

C. freundii FC50 showed the same growth profile with *E. coli* FI10944, it was able to grow in M9 minimal media with sinigrin but not in M9 minimal media with glucoraphanin (Figure 3.5 a). It reached highest OD_{600} at ~ 5^{th} hour, this is later than *E. coli* FI10944. The addition of more sinigrin at 4^{th} hour resulted in a higher growth rate (Figure 3.5 b).

a



b

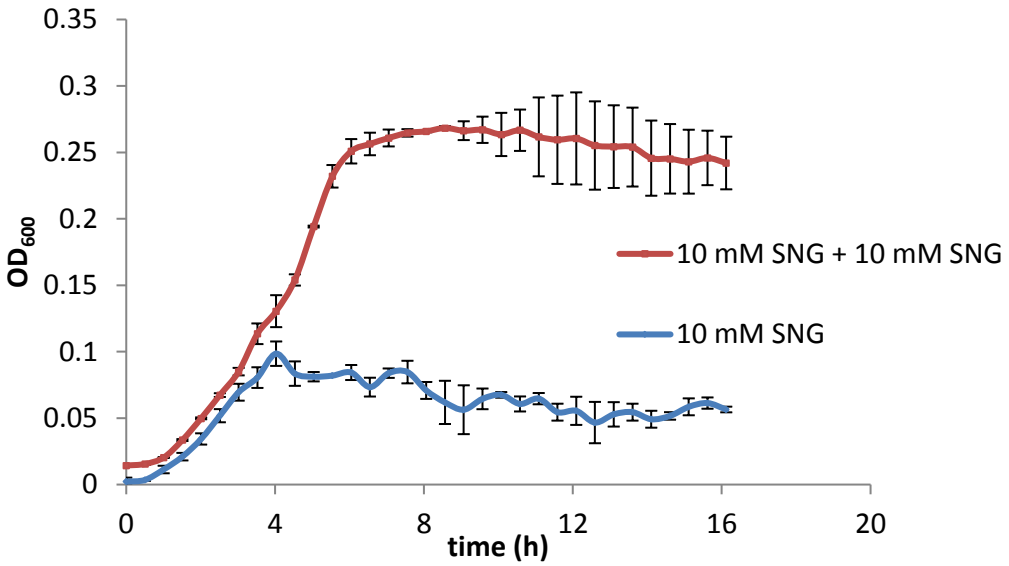


Figure 3.5 Growth curves of *C. freundii* FC50 in M9 minimal media at different growing conditions. a) Growth curve at 37°C for 24 h in M9 minimal media with sinigrin (SNG) or glucoraphanin (GRP) as sole carbon source. b) Growth curve at 37°C for 16 h in M9 minimal media with 10 mM sinigrin (SNG) to test the effect of extra sinigrin addition at 4th hour.

It is not known why *E. coli* FI10944 and *C. freundii* FC50 could not use glucoraphanin in the media as carbon source and did not grow (Figure 3.4 a and 3.5 a). This might be something about the glucoraphanin stock used in the study or a chemical absent from the M9 minimal

medium because *E. coli* FI10944 was able to use up glucoraphanin in fermentation studies before when it was grown in Nutrient Broth. In general, this experiment gave us an overview of the effect of sinigrin and glucoraphanin as a sole carbon source on *E. coli* FI10944 and *C. freundii* FC50 growth.

When CFEs of *C. freundii* FC50 were tested for myrosinase activity, the myrosinase activity of CFEs obtained from cultures grown in Nutrient Broth and Nutrient Broth with 1 mM sinigrin were 0.12 ± 0.1 µg glucose/min/mg enzyme and 0.15 ± 0.1 µg glucose/min/mg enzyme respectively. CFE obtained from culture grown in Nutrient Broth with 1 mM sinigrin had a significantly higher myrosinase activity compared to CFE obtained from the culture grown in Nutrient Broth one ($p=0.018$, two sample two tailed t-test). The genomic DNA of *C. freundii* FC50 was extracted, its genome was sequenced and assembled as described in Chapter 2. However, further characterisation of the glucosinolate metabolism by this strain could not be studied.

During the purification of myrosinase of *E. coli* FI10944, CFEs and protein extracts were prepared to be used in FPLC analysis. God-Perid assay was carried out to determine the myrosinase activity of these protein extracts using sinigrin as substrate. *Sinapis alba* myrosinase (Sigma) was used as a positive control. The results are shown in Table 3.6. The negative controls did not show any myrosinase activity. Negative control 1 consisted of extracts in buffer without substrate. This was set up to check that myrosinase activity was originated from CFE enzyme activity not from any glucose left in the extracts. The negative control 2 consisted of only sinigrin in buffer. This was set up to confirm that degradation was originated from an enzymatic activity. The results showed that CFE of *E. coli* FI10944 had myrosinase activity on sinigrin.

Samples	Specific Activity (µmol glucose/min/mg enzyme)
Cell-free extract (CFE)	0.08 ± 0.00
Precipitated extract	0.08 ± 0.00
Dialysed protein extract	0.15 ± 0.00
S. alba myrosinase	0.44 ± 0.05

Table 3.6 Specific myrosinase activity of extracts of *E. coli* FI10944. The myrosinase activity was assessed by God-Perid assay and sinigrin was used as substrate in the assay. The results are given as means of triplicates + SD.

As *E. coli* FI10944 was able to degrade all of the glucoraphanin supplied to Nutrient Broth previously. We wished to examine if the myrosinase activity in CFEs of *E. coli* FI10944 was inducible. HPLC was used to assess glucoraphanin degradation by CFEs. Glucoraphanin-induced CFEs showed a degradation rate of 1.68 ± 0.89 glucoraphanin/µg protein (12.1 ± 4.7% degradation of original glucoraphanin concentration). Uninduced CFEs were able to degrade at a rate of 1.37 ± 0.53 nmol glucoraphanin/µg protein (10.4 ± 5.5% degradation of original glucoraphanin concentration). There was no significant difference for glucoraphanin degradation rate of induced or uninduced CFEs of *E. coli* FI10944 ($p=0.641$, two sample two tailed t-test). However, it should be noted that samples had high SD values.

L. agilis R16 was examined to see if pre-culturing in MRS media with glucoraphanin, sinigrin, glucose or without glucose would change its myrosinase activity. The glucoraphanin and sinigrin degradation of samples were compared to negative controls which consisted of only glucoraphanin in MRS media and degradation rates were calculated. The glucosinolate contamination from pre-cultures were also taken into account in calculations. Results were summarised in Table 3.7. Pre-culturing in different conditions did not produce a remarkable difference in sinigrin or glucoraphanin degradation of *L. agilis* R16. It degraded 97-99% of the sinigrin in the MRS media and 9-14% of the glucoraphanin after 24 h incubation. Degradation rate of glucoraphanin was low as observed previously. It concluded that *L. agilis* R16 was not able to degrade glucoraphanin at as a high rate as sinigrin even if a pre-culturing in presence of glucosinolates was performed.

Pre-culturing Conditions	Glucoraphanin Degradation (%)	Sinigrin Degradation (%)
Grown with 1 mM sinigrin	9.1±6.6	97.4±2.6
Grown with 1 mM glucoraphanin	14.5±2.9	99.3±1.2
Grown with no glucose	13.0±2.2	99.2±1.2
Grown with 1 mM glucose	12.6 ±2.8	97.2±2.5

Table 3.7 Glucoraphanin and sinigrin degradation rates of *L. agilis* R16 pre-cultured at different conditions. *L. agilis* R16 was pre-cultured in MRS with 1 mM sinigrin, glucoraphanin, glucose or without glucose. The results are given as means of triplicates + SD.

3.4.2 Cloning and Expression of β-glucosidases

3.4.2.1 Selection of β-glucosidases to Clone and Express

The search of myrosinase-like genes in human gut bacteria was based on the aphid myrosinase sequence (Glycosyl Hydrolase Family 1 - GH1) but after identification of a bacterial myrosinase from soil belong to GH3, this sequence was also included in the gene search. In the *E. casseliflavus* CP1 genome, 40 β-glucosidases were found from different Glycosyl Hydrolase (GH) families. These β-glucosidases were aligned with aphid myrosinases (from *Brevicoryne brassicae* and *Acyrthosiphon pisum*) and searched for any significant similarity and the presence of conserved amino acid residues that have a key role in myrosinase activity.

The 3 best candidate genes were selected, they are β-glucosidases (EC 3.2.1.21): named as *ecg*1, having 30% identity with aphid myrosinase (from *Brevicoryne brassicae*); *ecg*43 having 31% sequence identity with aphid myrosinase (from *Brevicoryne brassicae*); *ecg*44 having 33% identity with aphid myrosinase (from *Brevicoryne brassicae*). The multiple sequence alignment of 3 β-glucosidases with aphid myrosinase from *Brevicoryne brassicae* and *Acyrthosiphon pisum* is shown below (Figure 3.7). The alignment was performed using Clustal Omega [237].

The proton donor and nucleophile are the key elements of myrosinase activity. In *B. brassicae* myrosinase, Glu 167 and Glu 374 were reported to be these elements respectively (Figure 3.6) [133]. The 3 β-glucosidases possess these amino acid residues as can be seen on the alignment (highlighted). His 39 was reported as $Zinc^{+2}$ binding, dimer forming amino acid residue but not conserved in the 3 bacterial β-glucosidases. Tyr 309, Trp 416, Trp 424 and Phe 432 in aphid myrosinase were reported to form a hydrophobic pocket and enable the

recognition of the glucose ring. Tyr 309, and Trp 416 were conserved in 3 β-glucosidases. Trp 424 was conserved in ecg43 and ecg44 but not in ecg1. Phe 432 residue was not conserved in any of the β-glucosidases but there was another hydrophobic amino acid Tyr in this position. Trp 123 in aphid myrosinase was reported to be involved in recognition of sulphur by van der Waals contact with sulphur of the thioglucosidic bond and this residue was only conserved in ecg44 but ecg1 and ecg43 have a Phe instead.

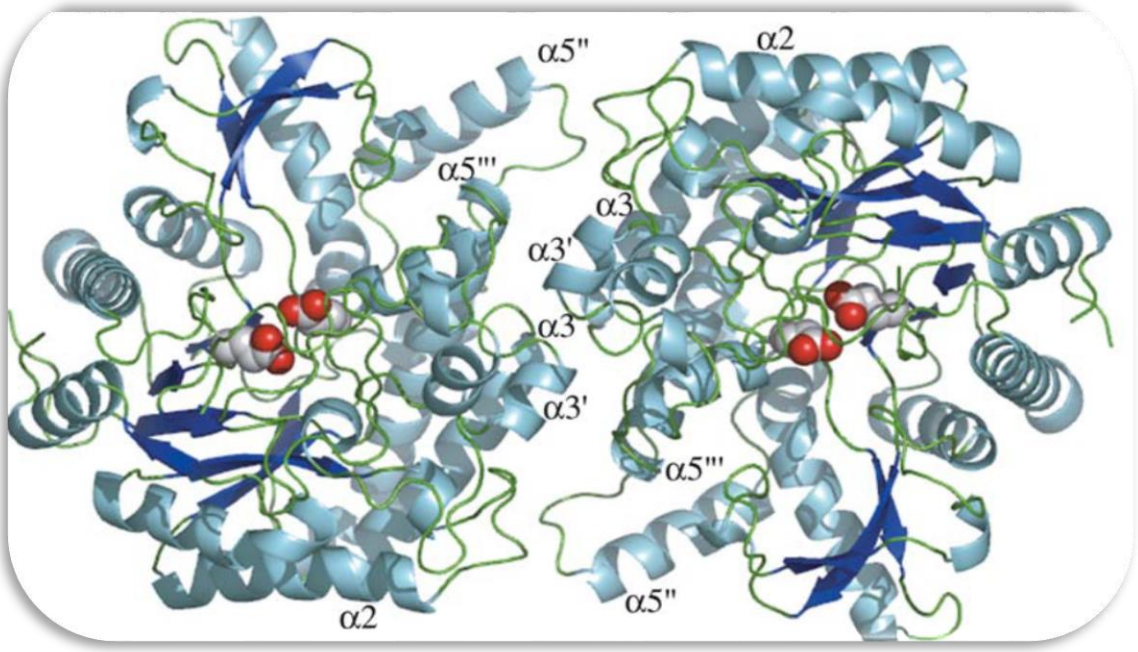


Figure 3.6 The structure of the aphid myrosinase. The β-strands are shown in dark blue, α-helices in light blue and loops in green. The two catalytic glutamic acid residues are shown as in red indicating the position of active sites. The image was obtained from Huseybe et al., 2005 [134].

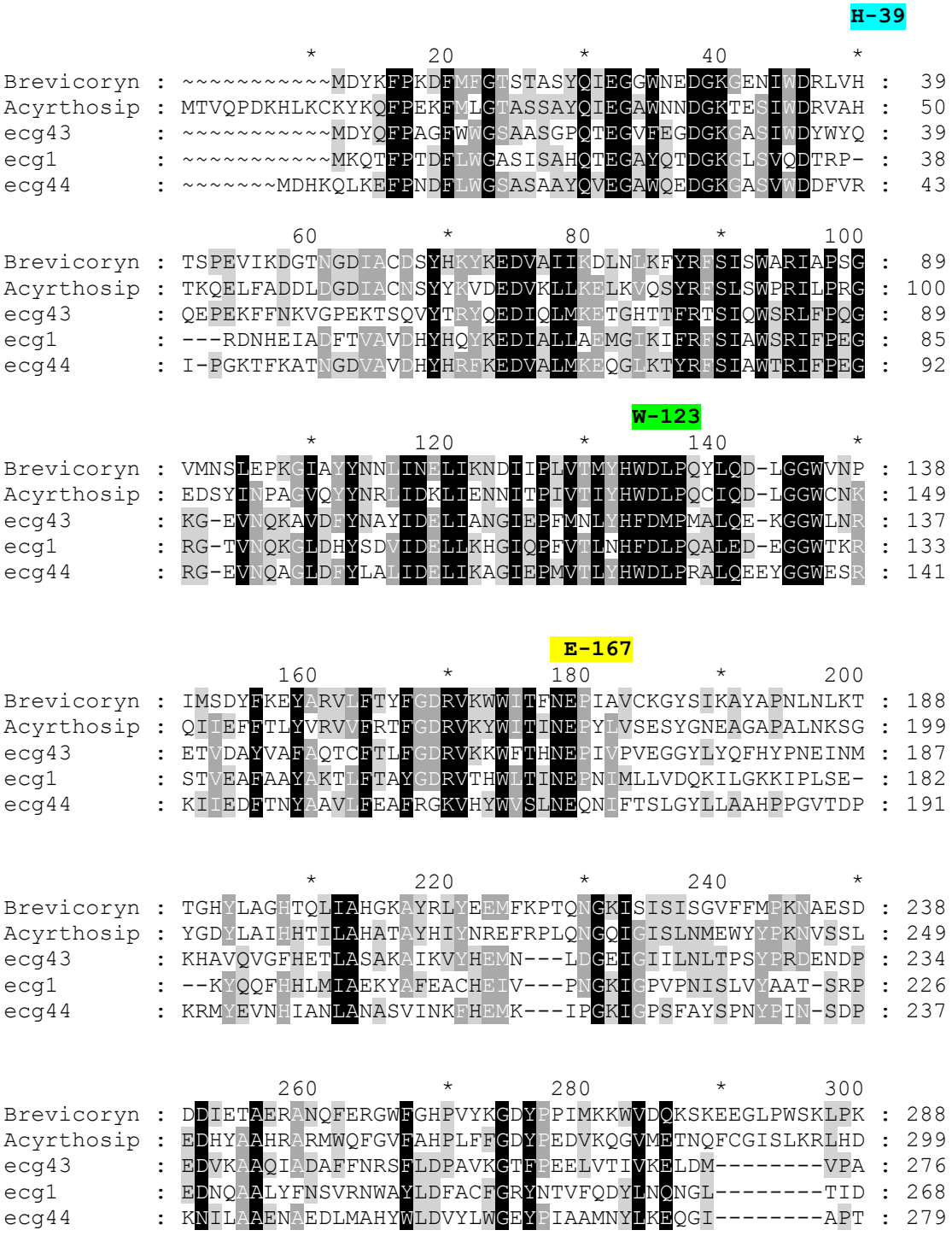


Figure 3.7 Alignments of putative β-glucosidases; *ecg*1, *ecg*43, *ecg*44 with aphid (*B. brassicae*) myrosinase and pea aphid (*A. pisum*) myrosinase. Brevicoryn; *B. brassicae* myrosinase, Acyrthosip; *A. pisum* myrosinase. The acid/base residue Glu 167 (E) and nucleophile of *B. brassicae* myrosinase are highlighted in yellow. Tyr 309 (Y), Trp 416 (W), Trp 424 (W) and Phe 432 (F) of *B. brassicae* myrosinase were highlighted in pink. His 39 (H) and Trp 123 (W) of *B. brassicae* myrosinase are highlighted in blue and green, respectively.

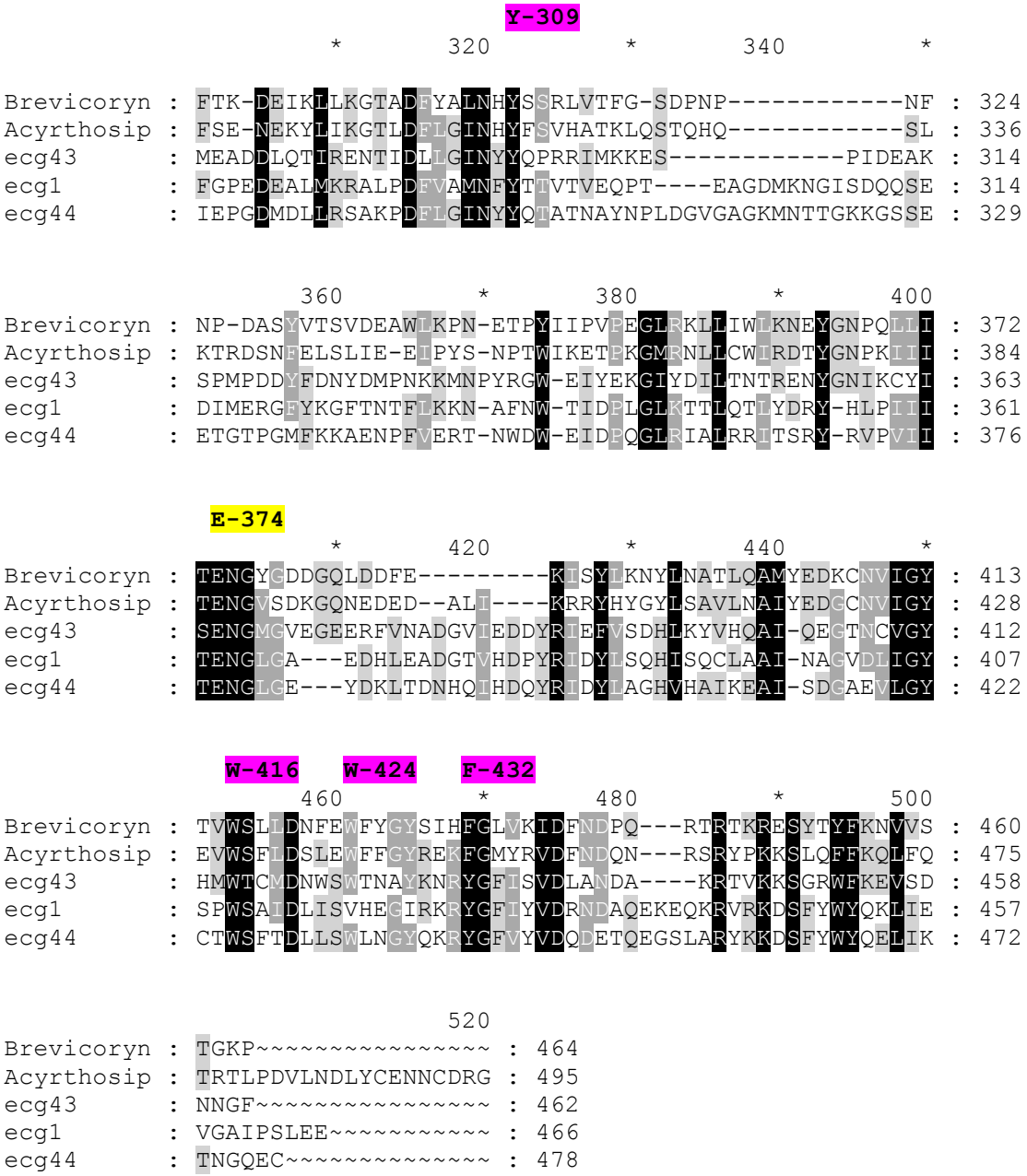


Figure 3.7-Continued

The ecg44 was selected for further experiments as it has most of the conserved amino acid residues and showed the highest sequence identity (33%) with aphid myrosinase (*B. brassicae*).

During the study, the first bacterial myrosinase sequence was identified and it is a periplasmic β-glucosidase from GH3, not from GH1 like aphid and plant myrosinases [118]. Based on this sequence, the genomes were reinterrogated and new candidate genes were selected for cloning and expression. Especially, the importance of a conserved amino acid sequence, SDW,

was reported in enzyme activity of *C.* WYE1 myrosinase [118]. Whilst the candidate genes were being selected, this sequence was taken into account as well. One candidate gene with the SDW sequence (*ecg*4) and one without it (*ecg*39) from *E. casseliflavus* CP1 were selected. In addition, one periplasmic glucosidase from *E. coli* FI10944 was selected (*ecolg*3). The alignment of selected genes with *C.* WYE1 myrosinase is shown in Figure 3.8 and SDW sequence has been highlighted. The selected genes and their sequence identity with *C.* WYE1 myrosinase is shown in Table 3.8.

Bacterium	Gene	Enzyme Family	Sequence Identity (%)
E. casseliflavus CP1	*ecg*44	GH1	*33
E. casseliflavus CP1	*ecg*4	GH3	29
E. casseliflavus CP1	*ecg*39	GH3	24
E. coli FI10944	*ecolg*3	GH3	24

Table 3.8 The genes cloned and expressed in this part of the study. Sequence identity indicates the sequence identity with *C.* WYE1 myrosinase, except *, * indicates the sequence identity with aphid (*B. brassicaea*) myrosinase. As ecg44 is from GH1 family (http://www.cazy.org/)[233], it was compared to aphid myrosinase.

```
                    *        20         *        40         *
CMYR    : MLTAFKMNTVSLAVLVCLPLSVSASIQSAQQPELGYDTVPLLHLSGLSFK :  50
ECG39   : ~~~~~~~~~~~~~~~~~MKIDVN---------------------------- :   6
ECG4    : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
ECOLG3  : ~~~~~~~~~~MKWLCSVGIAVSLAL----QPALAD--------------- :  21

                60         *        80         *       100
CMYR    : DLNRDGKLNPYEDWRLSPQTRAADLVKRMSVAEKAGVMMHGIAPAEG--- :  97
ECG39   : -MRKN---------RLQYTEAAEKILKQMTLKEKVHLMSGNMVLAD---- :  42
ECG4    : ~~~~~~~~~~~~~~~METVKLKELIDQMTVAEKVGQLVQLTPDFFAQGG :  34
ECOLG3  : DLFGNHPL-----TPEARDAFVTELLKKMTVDEKIGQLRLISVGPDNPKE :  66

                    *       120         *       140         *
CMYR    : --------------STFGNGSVYDSEATRKMVVDAHVNSLITRLNGEEPA : 133
ECG39   : -VRR-----DMA------NGQHYNE------------------------- :  55
ECG4    : EITGPMQQWQMSEDQLFTIGSVLGTHTAE--------------------- :  63
ECOLG3  : AIR------EMI--KDGQVGAIFNTVTRQ--------------------- :  87
```

Figure 3.8 Alignment of the putative β-glucosidases with *C.* WYE1 myrosinase. The SDW motif is highlighted in red. CMYR; *C.* WYE1 myrosinase.

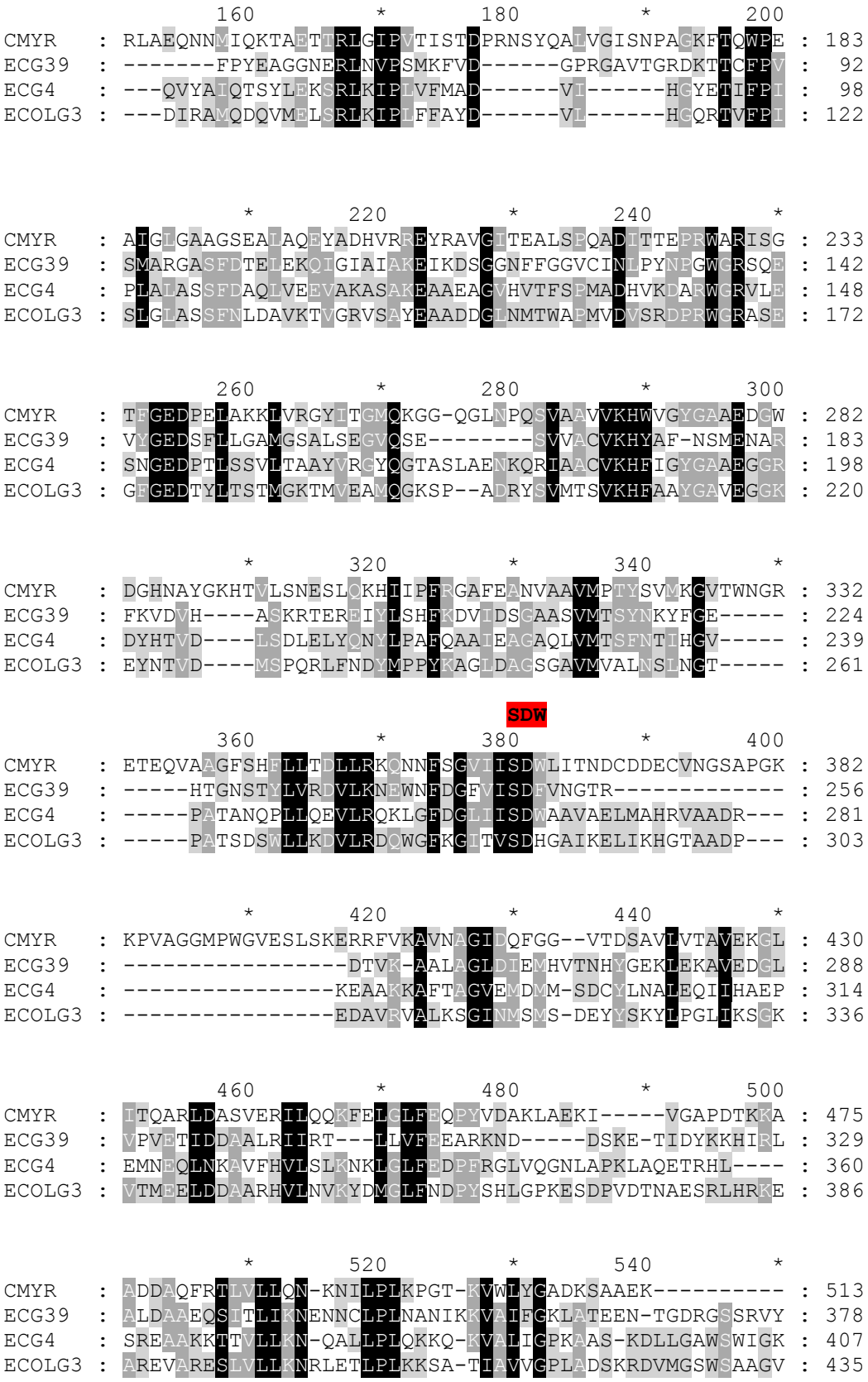


Figure 3.8 Continued

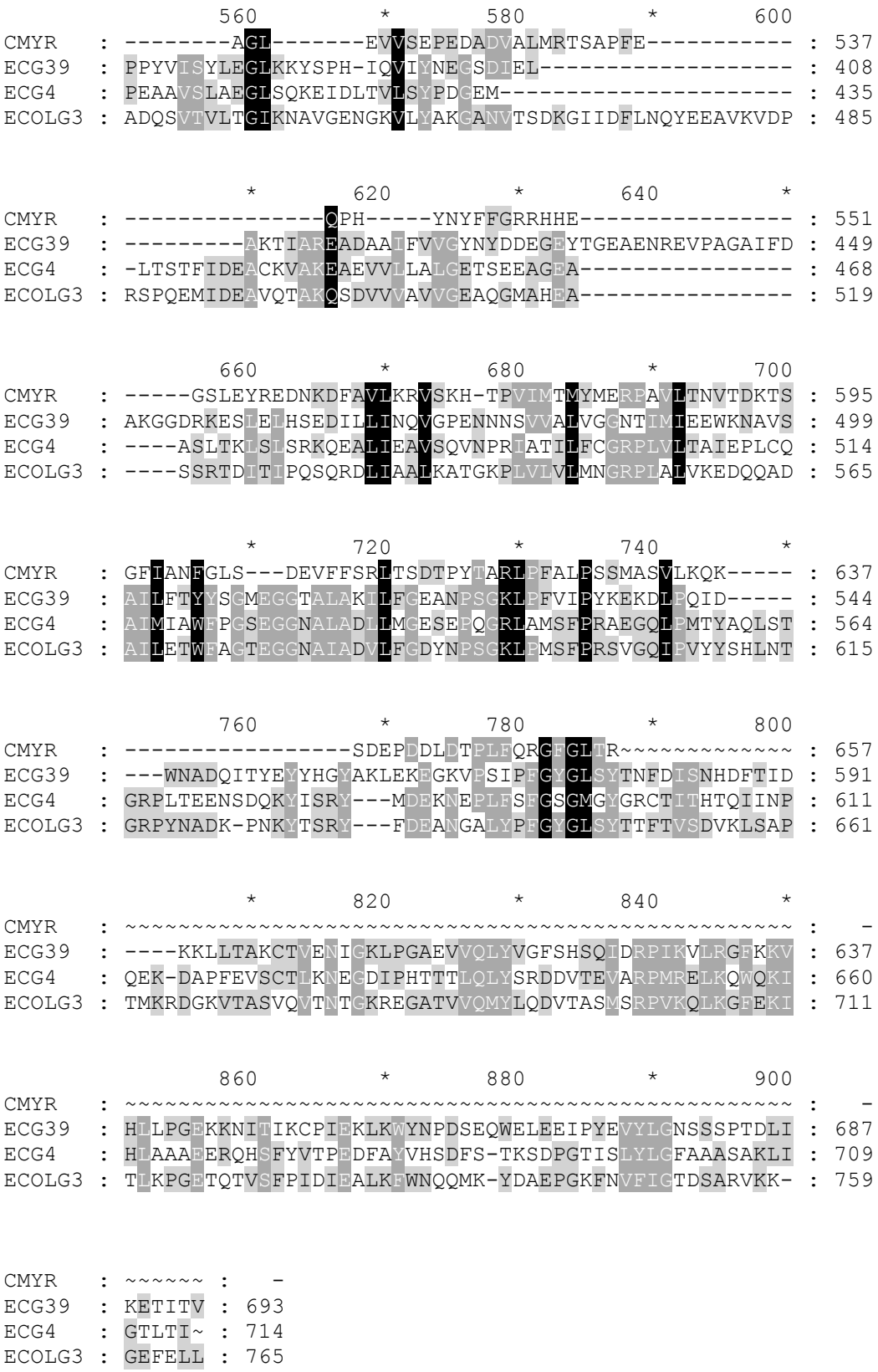


Figure 3.8 Continued

3.4.2.2 Cloning of Putative β-glucosidases of *E. casseliflavus* CP1 and *E. coli* FI10944

Four β-glucosidase genes of *E. casseliflavus* CP1 and *E. coli* FI10944 were cloned. One of them was from GH1 family (*ecg*44) and the rest were from GH3 family (*ecg*4, *ecg*39, *ecolg*3). Genomic DNA (gDNA) of *E. casseliflavus* CP1 and *E. coli* FI10944 were used to amplify the genes. The gel electrophoresis of gDNA is shown in Figure 3.9.

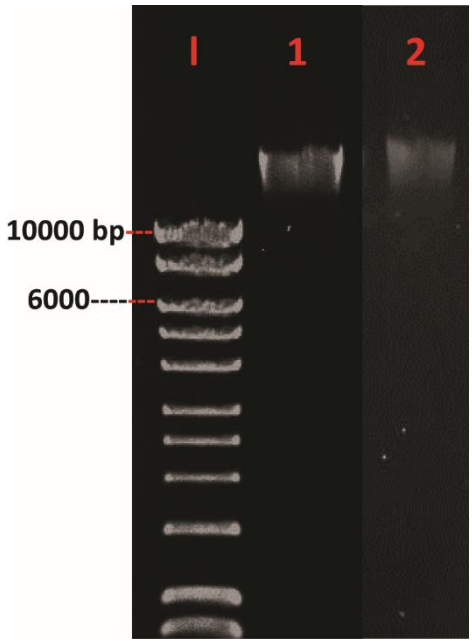


Figure 3.9 Agarose gel electrophoresis of bacterial gDNA. These gDNAs were used as template in PCR to amplify the genes. gDNA samples were run on 0.8 % agarose gel. l; Hyperladder, 1; gDNA of *E. coli* FI10944, 2; gDNA of *E. casseliflavus* CP1.

Specific primers (Table 3.1) were designed to produce PCR products. Some of the products gave a few bands on agarose gel then the PCR was repeated with optimized conditions (Table 3.2) until specific PCR products with expected sizes were obtained. After amplification, the inserts were ligated to double digested pET15b (*ecg*44) or pet28b (*ecg*4, *ecg*39, *ecolg*3) expression vectors (the maps of the vectors are given in Appendix 1) then the gene constructs were transformed into *E. coli* DH5α cells. The cells were plated on L agar with ampicillin (100 μg/ml final concentration) or kanamycin (30 μ/ml final concentration) plates. Several colonies were picked and tested by colony PCR to test if they have the plasmid with the correct gene inserts (Figure 3.10).

For the *ecg*44 gene construct, none of the screened colonies were found to have the correct insert so the ligation PCR was performed to test if the ligation worked. The ligation PCR was performed using the same T7P2, T7T primers (Chapter 2) and the same method as colony PCR. The PCR gave the expected gene product ~1600 bp so it was confirmed that the ligation worked. The transformation of gene constructs into *E. coli* cells was repeated using *E. coli* Top10 competent cells (Thermo-Fischer Scientific). The colonies were checked by colony PCR again and this time one positive transformant with correct insert was found. This positive colony was then grown on L agar with ampicillin (100 µg/ml). After extraction of the recombinant plasmid, it was sequenced and confirmed to have no mutation in the gene insert. The recombinant pET15b-*ecg*44 plasmid was transformed into *E. coli* BL21 (DE3) cells for protein expression studies.

The *ecg*4, *ecg*39, *ecolg*3 were amplified by PCR (Figure 3.9). Then each gene construct was ligated into pET28b vector and transformed into *E. coli* DH5α cells as described before.

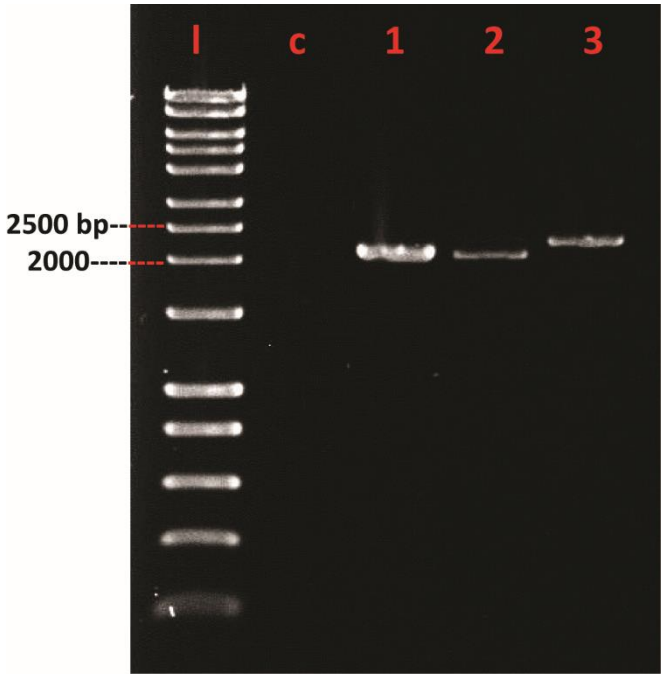


Figure 3.10 Agarose gel electrophoresis of PCR products. PCR products were run on 1% agarose gel. l; Hyperladder, c; PCR negative control, 1; ecg4 of *E. casseliflavus* CP1, 2; ecg39 of *E. casseliflavus* CP1, 3; ecolg3 of *E. coli* FI10944. The products showed the expected sizes; ecg4 (2162 bp), ecg39 (2101 bp), ecolg3 (2321 bp).

After transformation, several colonies were selected randomly and colony PCR was performed for each gene construct. The colony PCR results showed that there were gene

inserts with correct sizes in the recombinant plasmids (Figure 3.11). A few of the colonies with the expected PCR products were selected. They were grown on L agar with kanamycin (30 µg/ml) and the recombinant plasmids (pET28b-*ecg*4, pET28b-*ecg*39 and pET28b-*ecolg*3) were extracted. Sequencing confirmed that there was no mutation in the gene inserts. The recombinant pET28b-*ecg*4, pET28b-*ecg*39 and pET28b-*ecolg*3 plasmids were transformed into *E. coli* BL21 (DE3) cells for protein expression.

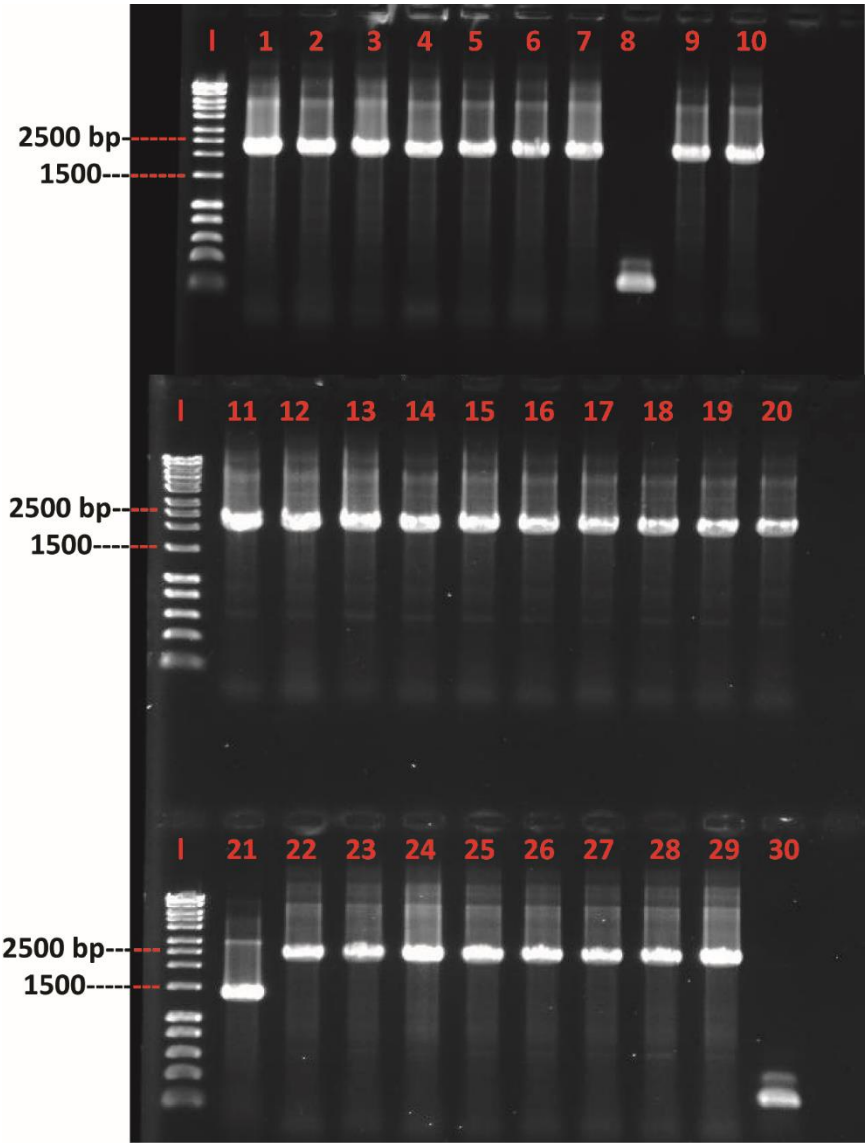


Figure 3.11 Agarose gel electrophoresis for colony PCR experiments. l; hyperladder, 1-10; selected colonies for *E. coli* DH5α pET28b-ecg4, 10-20; selected colonies for *E. coli* DH5α pET28b-ecg39, 20-30; selected colonies for *E. coli* DH5α pET28b-ecolg3. The products were run on 1% agarose gel. The expected PCR product sizes; 291 bp for pET28b vector with no insert, 2397 bp for recombinant pET28b-ecg4 plasmid, 2301 bp for recombinant pET28b-ecg39 plasmid, 2547 bp for recombinant pET28b-ecolg3 plasmid.

3.4.2.3 Protein Expression and Purification of Putative β-glucosidases of *E. casseliflavus* CP1 and *E. coli* FI10944

The protein expression of *E. coli* DH5α expressing pET28b-*ecg*4 was tested using different time intervals (1, 2, 3, 4 h) at 37°C. The 3 h protein expression induction resulted in the highest recombinant ecg44 protein expression, giving the recombinant protein size within the expected size (~56 kDa) with 6XHis-tag. The induction for 2 h and 4 h seemed to produce similar amount of protein when cell-free extracts (CFE) were checked on SDS-PAGE gel. The protein gel image is shown in Figure 3.12a. The CFEs were checked for the presence of 6X-His tag by Western blot as described in Chapter 2 (Figure 3.12b).

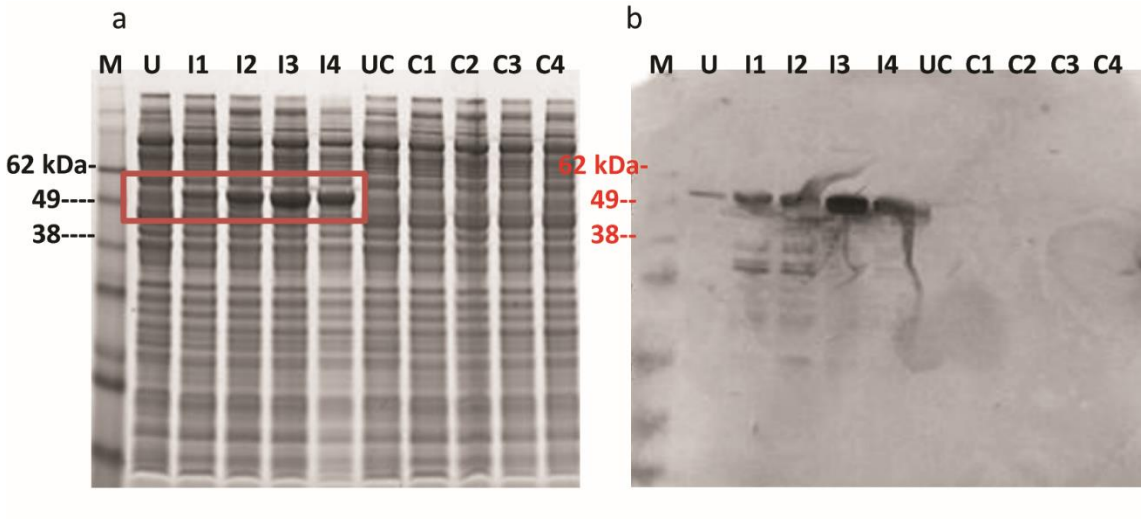


Figure 3.12 SDS-PAGE and Western blot of the CFEs from *E. coli* BL21 (DE3) expressing pET15b-*ecg*44. a) Protein gel where CFEs were run on 10% Bis-Tris gel under denaturing, reducing conditions in MES buffer. b) Western blot (using anti-His tag antibody) of CFEs which was run within the same order of protein gel in a. 10 µg of protein was loaded per lane. M: See Blue Protein Marker, U: CFE obtained from uninduced *E. coli* BL21 (DE3) expressing pET15b-*ecg*44 cells, I1-I4: CFEs obtained from 1-4 h induced *E. coli* BL21 (DE3) expressing pET15b-*ecg*44 cells, UC: CFE obtained from uninduced *E. coli* BL21 (DE3) pET15b cells, C1-C4; CFE obtained from 1-4 h induced *E. coli* BL21 (DE3) pET15b cells (empty vector controls).

The expression of the recombinant ecg44 protein was determined by Western Blot using a His-tag antibody (Novagen). Recombinant proteins with expected protein sizes were recognised by the His-tag antibody so the purification of ecg44 was performed using Ni-NTA column. The *E. coli* BL21 (DE3) cells expressing ecg44 were induced for 3 h at 37°C, with shaking at 250 rpm. The cells were spun down and Ni-NTA purification was performed as described in Chapter 2. All fractions (3.25 µl) from the purification steps were run on a protein gel (Figure 3.13).

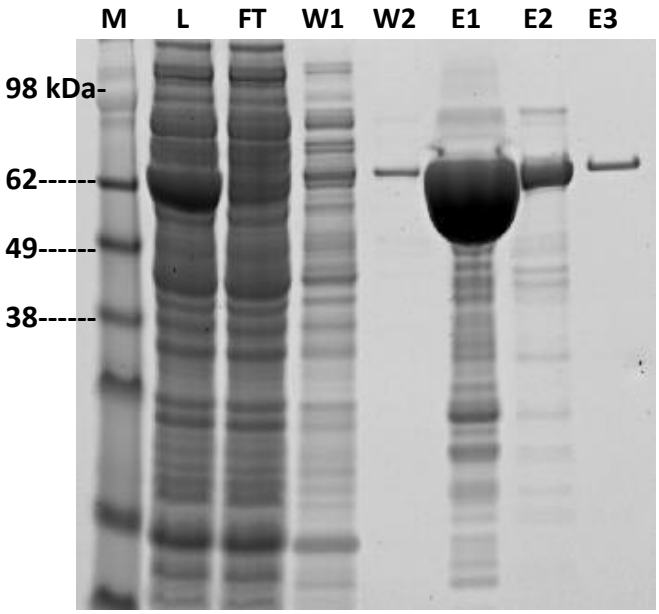


Figure 3.13 SDS-PAGE of the Ni-NTA purified ecg44 proteins from *E. coli* BL21 (DE3) expressing pET15b-*ecg*44. 3.25 µl of fractions were loaded per lane. Proteins were run on 10% Bis-Tris Gel under denaturing, reducing conditions in MES buffer. L: Lysate, FT: Flow through, W: Wash, E: Eluate, M; Protein marker.

The optimum protein expression conditions of *E. coli* BL21(DE3) expressing pET28b-*ecg*4 and pET28b-*ecg*39 were found to be 25°C for 4 h. However, 3 h protein expression induction at 37°C resulted in a better expression of the recombinant proteins for *E. coli* BL21(DE3) expressing pET28b-*ecolg*3. Using these optimum conditions, protein expression was induced and the cells were centrifuged for 15 min at 3220 x *g*, 4°C. The supernatant was removed and the cell paste was frozen at -20°C. The CFEs were prepared by sonication and run on a protein gel (denaturing, reducing conditions). The protein gel image is shown below (Figure 3.14). The recombinant ecg4 (82.5 kDa), ecg39 (80.7 kDa) and ecolg3 (87.8 kDa) are visible on the gel within expected sizes with 6XHis-tags. The CFEs were checked for the presence of 6X-His tag by Western blot as described in Chapter 2 (Figure 3.14b).

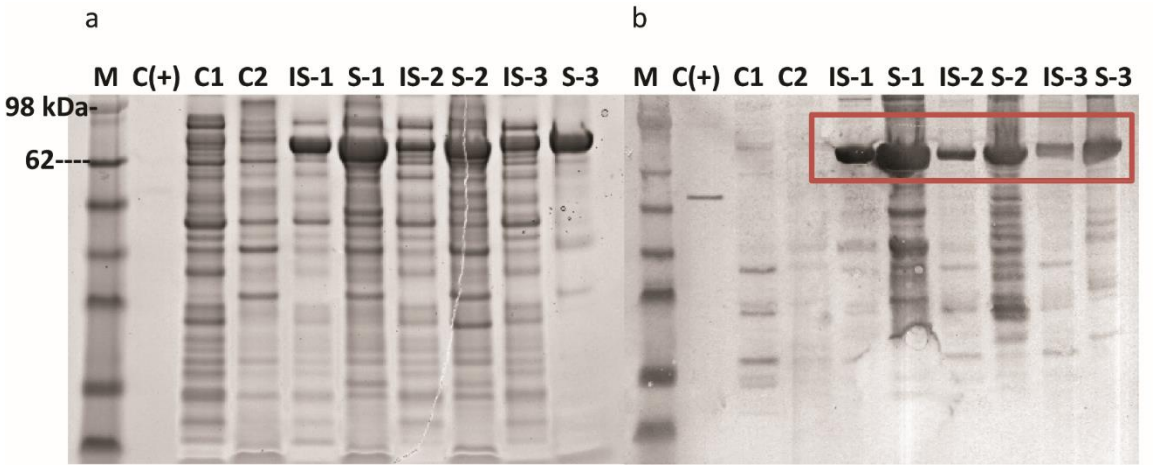


Figure 3.14 SDS-PAGE and Western Blot of the CFEs from *E. coli* BL21 (DE3) expressing pET28b-*ecg*4, pET28b-*ecg*39 and pET28b-*ecolg*3. a) Protein gel where CFEs were run on 10% Bis-Tris gel under denaturing, reducing conditions in MES buffer. b) Western blot (using anti-His tag antibody) of CFEs which was run within the same order of protein gel in a. 10 µg of protein was loaded per lane. M: See Blue Protein Marker, C (+): ecg44, 6XHis-tagged protein, C1 and C2: CFEs obtained from *E. coli* BL21 (DE3) expressing pET15b (empty vector control), IS: Insoluble extract, S: Soluble extract, 1: CFE obtained from *E. coli* BL21 (DE3) expressing pET28b-*ecg*4, 2: CFE obtained from E. coli BL21 (DE3) expressing pET28b-*ecg*39, 3: CFE obtained from *E. coli* BL21 (DE3) expressing pET28b-*ecolg*3.

The ecg4, ecg39 and ecolg3 were purified in the same way as ecg44 by Ni-NTA purification (Figure 3.15).

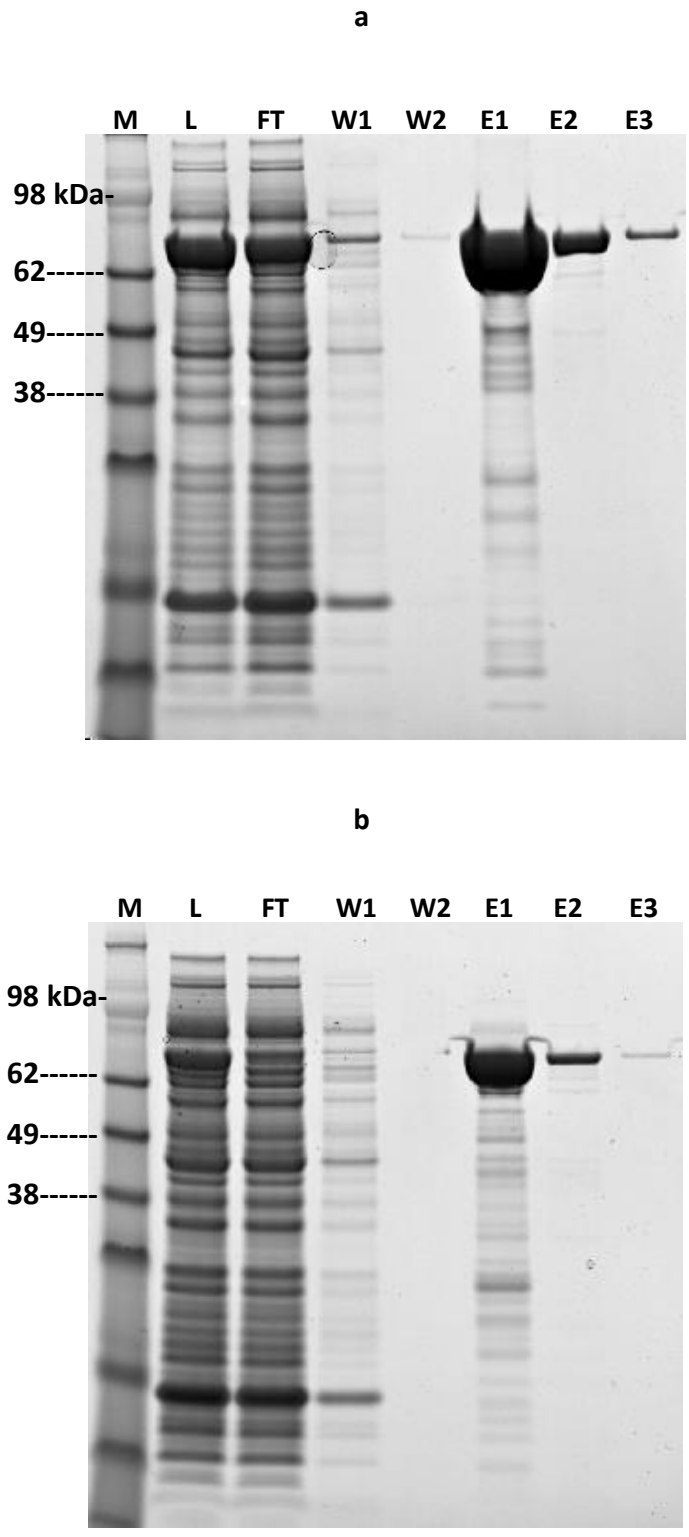


Figure 3.15 SDS-PAGE of the Ni-NTA purified ecg4 (a), ecg39 (b) and ecolg3 (c) proteins from *E. coli* BL21 (DE3) expressing pET28b-*ecg*4, pET28b-*ecg*39 and pET28b-*ecolg*3 respectively. 3.25 µl of fractions were loaded per lane. Proteins were run on 4-12% Bis-Tris Gel under denaturing, reducing conditions in MES buffer. L: Lysate, FT: Flow through, W: Wash, E: Eluate, M; Protein marker.

c

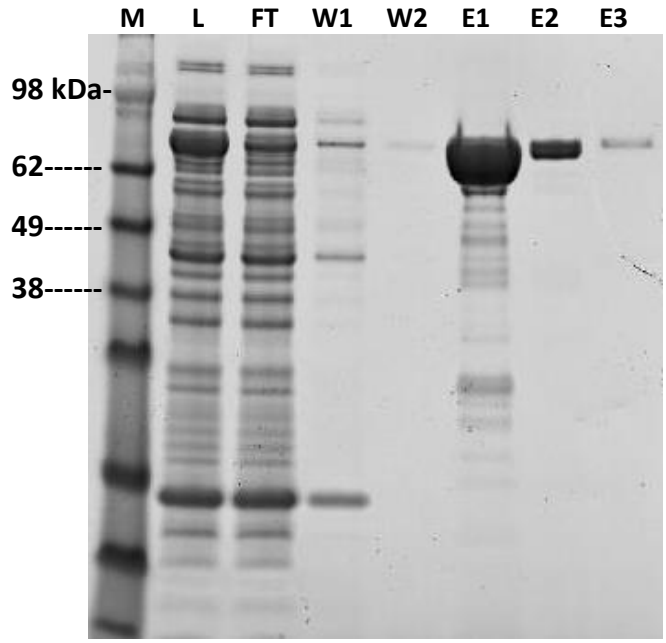


Figure 3.15-Continued

After Ni-NTA purification, ecg44 was dialysed against 25 mM Tris-HCl pH 7.4 buffer. The dialysed ecg44 was quantified by Bradford assay. Myrosinase assay (God-Perid) was carried out using 200 µg of ecg44 (~12 µM) *S. alba* myrosinase was used to as a positive control. The positive control showed a specific enzyme activity of 0.17 µmol/min/mg enzyme for sinigrin. However, ecg44 did not show any myrosinase activity against sinigrin or glucoraphanin.

The effect of ascorbic acid on myrosinase activity of ecg44 was tested using the God-Perid assay. No myrosinase activity was determined when 1 mM ascorbic acid was used. The presence of cofactors ($MgCl_2$ or $ZnCl_2$) or different amount of enzyme (10, 50, 200 µg of ecg44) did not make any difference and no myrosinase activity was observed. To test if any factor from CFE was needed for myrosinase activity of ecg44, CFE of *E. coli* BL21(DE3) expressing ecg44 was added to purified protein. However, this did not make any difference on the enzyme activity. The ecg44 is a β-glucosidase so β-glucosidase activity assay was performed using 4-nitrophenyl β-D-glucopyranoside as substrate but it did not show β-glucosidase activity either.

The Ni-NTA purified ecg4, ecg39 and ecolg3 were dialysed against 20 mM citrate phosphate buffer pH 7.0. The myrosinase activity of dialysed and undialysed ecg4, ecg39 and ecolg3 proteins were tested using the God-Perid assay. The results revealed that these proteins do not have any myrosinase activity. To test if myrosinase activity of these recombinant proteins

is in need of any cofactor from CFE, CFE of *E. coli* FI10944 showing myrosinase activity was added to the recombinant proteins. This did not change the result and recombinant proteins did not show myrosinase activity. They were also tested for β-glucosidase activity using 4-nitrophenyl β-D-glucopyranoside as substrate. β-glucosidase activity of ecg4 and ecg39 were 2.4 ± 0.1 µmol p-nitrophenol/min/mg enzyme and 0.08 ± 0.02 µmol p-nitrophenol/min/mg enzyme respectively when in undialysed proteins. β-glucosidase activity levels in ecg4 and ecg39 were found to be 1.6 ± 0.2 p-nitrophenol/min/mg enzyme and 0.07 ± 0.01 p-nitrophenol/min/mg enzyme respectively when dialysed proteins were used. In summary, ecg4 and ecg39 proteins were active, they showed β-glucosidase activity. However, they did not show myrosinase activity under the conditions tested.

3.4.3 Purification of the Bacterial Myrosinase of *E. coli* FI10944

In the first attempt to purify the putative bacterial myrosinase, overnight grown culture of *E. coli* FI10944 (OD_{600} = 1.7) was used to prepare protein extract and then fractionated by $(NH_4)_2SO_4$ precipitation. Precipitate was dialysed and then run on IEX chromatography and fractions were collected (Method in section 3.3.8). The fractions A8, A9 and A10 showed myrosinase activity within the 13 fractions tested (Figure 3.16). The enzyme activities of fractions were 0.11 ± 0.00, 0.24 ± 0.01 and 0.01 ± 0.00 µmol glucose/min/mg enzyme respectively. The positive control, a commercial plant myrosinase of *S. alba*, had an enzyme activity of 0.17 µmol glucose/min/mg enzyme. The negative control, with only sinigrin in buffer, did not show any enzyme activity. Precipitated and dialysed protein extracts and fractions showing myrosinase activity (A8, A9 and A10) were run on a protein gel (Figure 3.17). Bands shown on Figure 3.17 were excised and analysed by LC-MS\MS. As fraction A9 showed the highest activity, this fraction was further purified using gel filtration to identify the protein responsible for myrosinase activity.

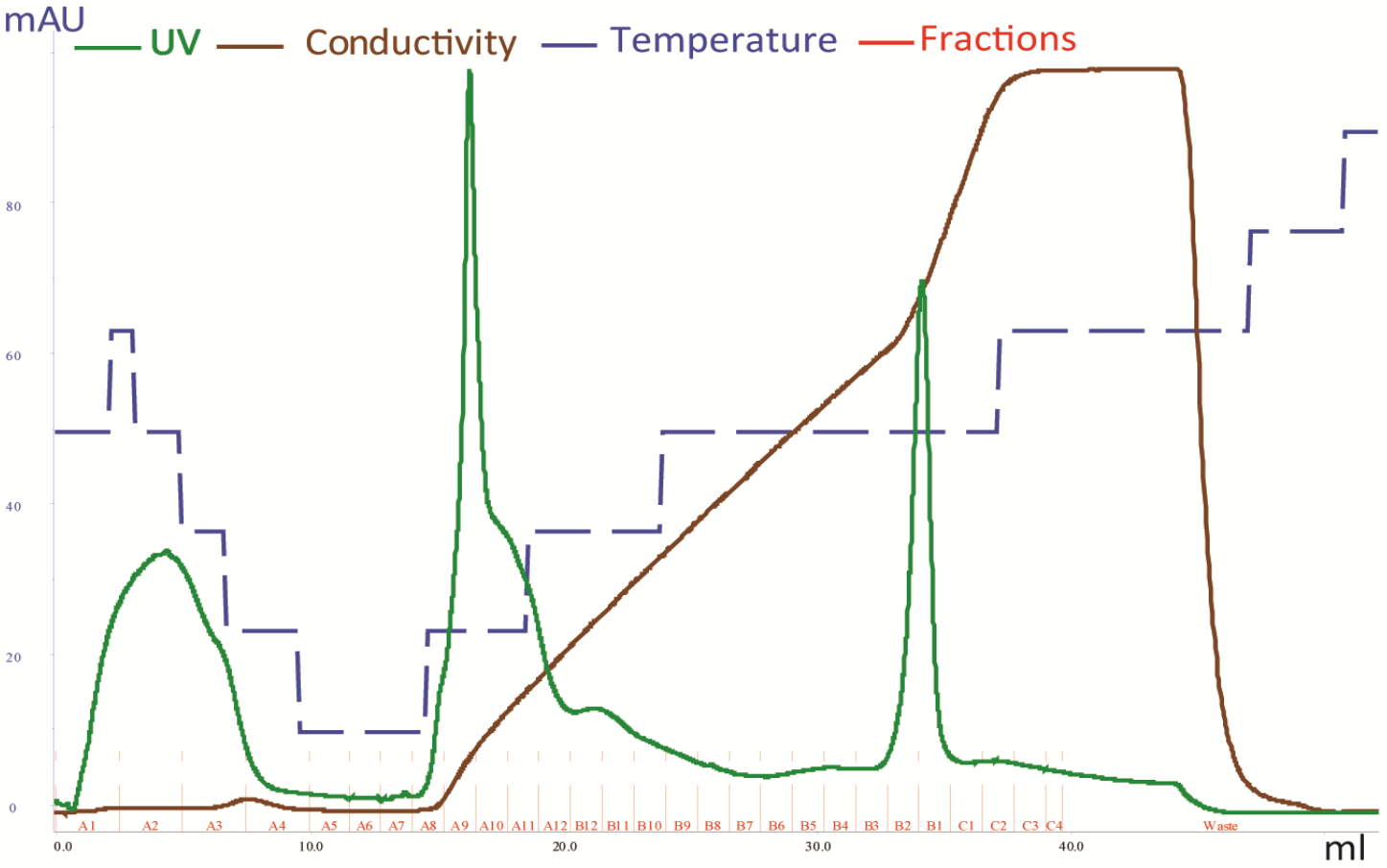


Figure 3.16 Chromatogram showing IEX fractions from ammonium sulphate precipitated proteins of *E. coli* FI10944. The protein eluates were monitored at 280 nm. Fractions were collected.

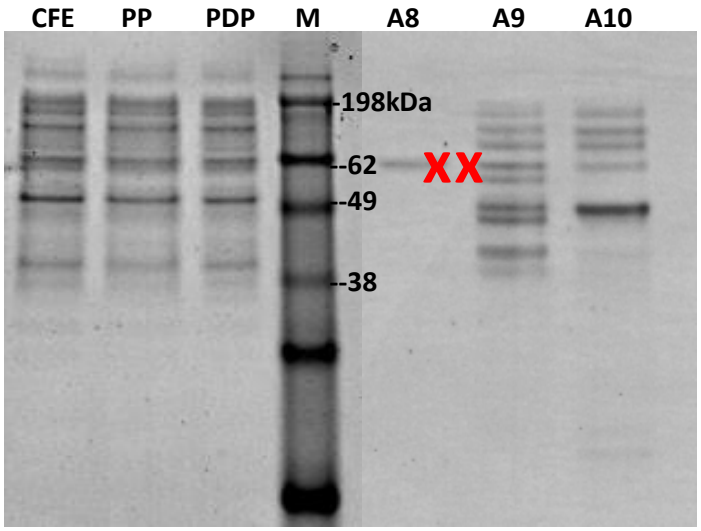


Figure 3.17 SDS-PAGE showing protein extracts of *E. coli* FI10944 and fractions from IEX chromatography. 10 µg of protein was loaded per lane. CFE; cell-free extract, PP; precipitated protein extract, PDP; precipitated, dialysed protein, M; See blue protein marker, A8-F10; fractions from IEX chromatography, X; Protein bands for excision. Proteins were run on 10% Bis-Tris Gel under denaturing conditions.

Mascot analysis showed that the excised A8 protein band and its equivalent in A9 are flagellin, no glucosidase sequence was obtained from this analysis. The fraction A9 was further purified by gel filtration using a Superdex 75 (GE Healthcare) column (Method in Chapter 2) and 33 fractions were obtained. These fractions were also tested for myrosinase activity. The fraction showing the highest myrosinase activity was further analysed by LC-MS/MS. When Mascot results were interrogated with *E. coli* FI10944 genome or compared to *C.* WYE1 myrosinase, no glucosidase sequence or conserved domains related to myrosinase activity were found. The results revealed that phosphoglycerate kinases and glycerophosphoryl diester phosphodiesterases dominated the fraction.

In the second attempt of purification of bacterial myrosinase by chromatography, the conditions for myrosinase activity were optimised with different buffers. The buffers 20 mM Na Citrate pH 5.0 or 6.0 were found to be better for protein yield under the conditions tested (Figure 3.18).

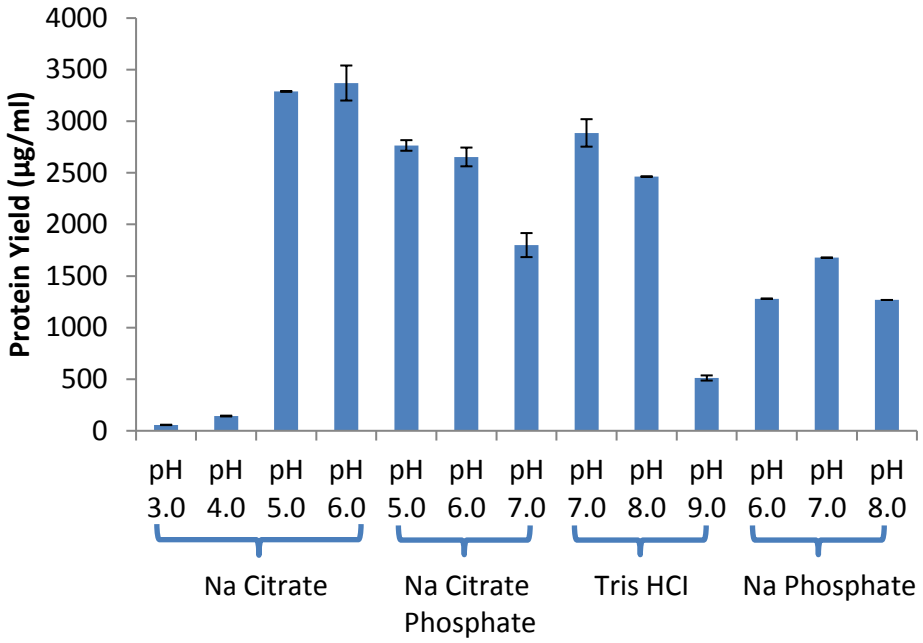


Figure 3.18 Protein yields of CFEs of *E. coli* FI10944 in various buffers. All of the buffers were at a concentration of 20 mM. The protein amount was determined by Bradford Assay in triplicates and the results are given as mean of triplicates ± SD.

20 mM Na Citrate buffer pH 5.0 was concluded to be better for myrosinase activity under the conditions tested (Figure 3.19). CFEs in 20 mM Na Citrate buffer pH 5.0 was tested for myrosinase activity by God-Perid assay and they showed a myrosinase activity of 0.13 ± 0.00 µmol glucose/min/mg enzyme. Optimum pH for myrosinase activity was found to be pH 5.0.

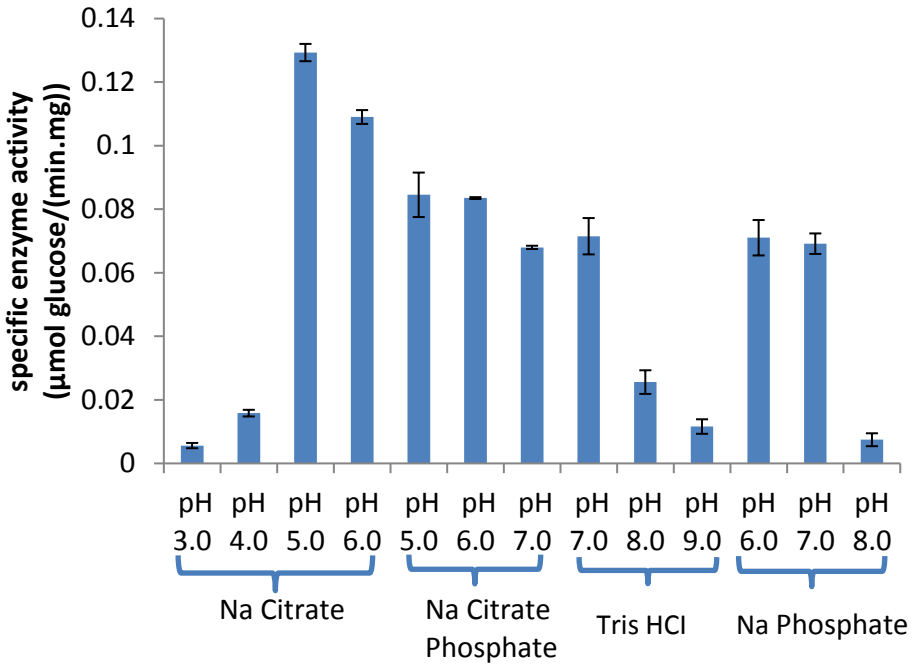


Figure 3.19 Specific enzyme activity of CFEs of *E. coli* FI10944 in various buffers. All of the buffers were used at a concentration of 20 mM. The myrosinase assay was performed to determine activity and the results are given as mean of triplicates ± SD.

E. coli FI01944 was grown in 500 ml of Nutrient Broth without glucosinolates at 37°C overnight, OD_{600} of overnight grown culture was 1.9. The protein extract was prepared as described before and run on FPLC to perform size exclusion chromatography (Method in Chapter 2). Four of the fractions showed myrosinase activity and one of the fractions was run on 10% Bis-Tris gel under denaturing, reducing conditions for 10 min then a protein band was excised and analysed by LC-MS/MS.

When this experiment was run previously, it was found that the fraction showing myrosinase activity was dominated by phosphoglycerate kinases. This time, enzyme activity optimisation was performed in advance to get a better chance to find out the responsible enzyme for myrosinase activity in the fraction. The protein sequences detected in the samples were analysed by Mascot (Results are given in Appendix 4). The results concluded that the enzymes found in the fraction are strongly related with the glycolysis pathway, including phosphoglycerate kinases but glucosidases were also found in the fraction. These glucosidases were a predicted cytoplasmic α-amylase (~57 kDa) and a periplasmic trehalase (~64 kDa). Notably, cytoplasmic α-amylase showed significant matches. No β-glucosidases were identified in the fraction. The conditions used in chromatography in two attempts were summarized in Table 3.9.

	1st attempt	2nd attempt
Culturing Media	NB with 1 mM sinigrin	NB
Purification Method	IEX Chromatography followed by GF	GF
Fractions with Myrosinase activity	A8, A9 and A10 from IEX A12 and B12 from GF	B6, B5, B4, B3
Fractions tested by LCMS/MS	A8 from IEX B12 from GF	B4
LCMS/MS Results	A8; flagellins B12; phosphoglycerate kinases, diester phosphodiesterases, phosphopento mutases, NAD(P)H dehydrogenases, enolases, glucose-6-phosphate isomerases, etc.	B4; phosphoglycerate kinases, transaldolases, triosephosphate isomerases, glucose-1-phosphatases, cytoplasmic α-amylases, glucose-6-phosphate isomerases, superoxide dismutases, periplasmic trehalase, etc.

Table 3.9 The details and results of chromatography methods used in purification of bacterial myrosinase. The fractions obtained and LCMS/MS results of the fractions. NB; nutrient broth, IEX; ion exchange, GF; gel filtration. LCMS/MS results indicate the Mascot results for each fraction. The experimental mass values are compared with calculated peptide mass or fragment ion mass values. An appropriate scoring algorithm was used to identify the closest match or matches.

3.5 DISCUSSION

In this chapter, human gut bacteria were tested for myrosinase activity then chromatography and cloning methods were used to attempt to identify the myrosinases. *E. coli* FI10944 was not screened for its glucoraphanin degrading ability before this study although it was shown to degrade sinigrin [137]. It degraded all of the glucoraphanin supplied to the media after 24 h. In addition, the glucosinolate degrading ability was observed in CFEs of *E. coli* FI10944. It was suggested that the glucosinolate degrading ability of bacteria may be in the membrane/debris fraction because no myrosinase activity was seen in CFEs of *L. agilis* R16 [56, 210], *E. coli* VL8 [56] and *E. casseliflavus* CP1 [56] but it was present with intact cells. However, our study showed that cell wall-associated activity is not always the case. CFEs of *E. coli* FI10944 and *C. freundii* FC50 showed myrosinase activity. This is in accordance with a study reporting myrosinase activity in CFEs of *Bifidobacterium* and *Lactobacillus* strains [232]. When CFEs of *E. coli* FI10944 were prepared from induced cells (grown in Nutrient Broth) and uninduced cells (grown in Nutrient Broth with 1 mM glucoraphanin), no significant difference was seen between degradation rates of induced or uninduced CFEs of *E. coli* FI10944. These results suggest that myrosinase activity of *E. coli* FI10944 may be constitutive. However, there was myrosinase activity in CFEs prepared from uninduced *E. coli* FI10944 cells. This result suggests that myrosinase activity is not always associated with the presence of glucosinolates as the inducer. A slight induction effect of sinigrin was observed in CFEs of *C. freundii* FC50 for myrosinase activity. When *C. freundii* was grown in Nutrient Broth with 1 mM sinigrin instead of Nutrient Broth, CFE obtained had slightly higher myrosinase activity. This suggests that myrosinase activity in CFE of *C. freundii* FC50 can be further induced.

E. casseliflavus CP1 used up 7.3% of the glucoraphanin supplied to the media; this was lower than the degradation rate (53%) reported previously by Luang-In et al. (2014) [56]. The same media, incubation time (24 h) were used under anaerobic conditions. The bacteria were isolated and shortly after used in fermentation studies [56] but we tested the activity after a certain storage time away from gut environment. This might be the reason for the loss of glucoraphanin degrading ability of *E. casseliflavus* CP1. During the storage, bacteria might have lost a large part of its myrosinase activity. The same effect was seen in our research group for antimicrobial activity of gut strains. The gut strains possessing bactericidal effects lost their activity during storage [238].

L. agilis R16 used up 12.3% of the glucoraphanin in the media that is same range to 10.0% reported previously [56]. *L. agilis* R16 kept its glucoraphanin degrading ability at the same rate after long storage time. When, *L. agilis* R16 was pre-cultured in MRS with 1 mM sinigrin, glucoraphanin, glucose or without glucose, it still showed a high sinigrin degradation rate (~97-99%) and low glucoraphanin degradation rate (~9-14%). Pre-culturing in media with glucosinolates did not produce a remarkable difference for glucoraphanin and sinigrin degradation compared to pre-culturing in media with or without glucose. Instead of using one step pre-culturing, several pre-culturing steps or using M9 minimal media with glucosinolates as the sole carbon source could be investigated. In fact, the glucosinolate side chain (the size or structure) might be the limiting factor. As glucoraphanin is a methylsulfinylalkyl-glucosinolate possessing a sulphoxide group with a high polarity, the sulphoxide group in the side chain might be the reason for the lower degradation rate. It was reported that *L. agilis* was able to degrade all of the glucoerucin (reduced form of glucoraphanin) but only 10% of the glucoiberin which has sulphoxide group in the structure after 24 h when it was cultured in MRS anaerobically [56]. As *L. agilis* R16 could degrade glucoerucin but not glucoraphanin, it might lack an efficient reductase activity to reduce sulphoxide group.

When the growth of *E. coli* FI10944 and *C. freundii* FC50 were examined over time in M9 minimal media with glucoraphanin and sinigrin, it was found that both strains could grow in M9 minimal media with sinigrin. However, they did not grow on M9 minimal media with glucoraphanin. When *E. coli* FI10944 was grown in Nutrient Broth it was able to consume glucoraphanin. The reason for glucoraphanin not being used as sole carbon source by *E. coli* FI10944 and *C. freundii* FC50 might be the lack of a chemical absent in the M9 minimal media.

In general, our results showed that human gut bacteria have a diverse glucosinolate metabolism capacity and action mechanism to degrade different glucosinolates. A reduction step which is the reduction of sulphoxide group in glucoraphanin structure is likely to be involved in the degradation of glucoraphanin by *E. coli* FI10944. After reduction of glucosinolate, myrosinase in the bacteria may act on the reduced glucosinolate. The same reduction activity was proposed to be involved in the degradation of glucoiberin and glucoraphanin by *Escherichia coli* VL8 [56], *Escherichia coli* Nisssle 1917 and *Enterobacter cloacae* ATCC13047 [55]. In addition, a reductase, methionine sulphoxide reductase A, was suggested to be responsible for this reduction [55]. These enzymes and reduction mechanism are studied in Chapter 4.

To identify the myrosinase of *E. coli* FI10944, buffer conditions for myrosinase activity were optimised and optimum pH for myrosinase activity against sinigrin was found to be pH 5.0. This is different from the pH optima of *C.* WYE1 myrosinase (pH 6.0) [118] and *Brevicoryne brassicae* aphid myrosinase (pH 5.5) [132] for sinigrin degradation. Two attempts were performed to identify myrosinase of *E. coli* FI10944 by FPLC. When crude protein extract of *E. coli* FI10944 was prepared and fractionated by chromatography, no β-glucosidases were identified in the fractions. In the second attempt, predicted cytoplasmic α-amylase and periplasmic trehalase were identified in the purified fraction. Cytoplasmic α-amylase and periplasmic trehalase are enzymes to break down α-bonds in polysaccharides. However, glucosinolates are β-thioglycosides so it is unlikely that cytoplasmic α-amylase and periplasmic trehalase might be involved in glucosinolate metabolism. So far, it was not possible to identify the myrosinase by chromatography. Our results suggested that myrosinase activity in CFE of *E. coli* FI10944 is not inducible. While myrosinase activity in CFE of *C. freundii* was slightly induced by sinigrin and CFEs of *C. freundii* showed myrosinase activity. It will be worth trying to grow *C. freundii* in M9 minimal media with sinigrin as sole carbon source then purify by chromatography as performed for *C.* WYE1 by Albaser et al. (2016)[118].

Heterologous expression approach was also used to identify the bacterial myrosinase of the human gut bacteria. It was challenging to select the glucosidases to clone. The bacteria studied have many glucosidases and our selection was based on similarity to identified myrosinases. As all identified and characterised myrosinases were from GH1 family so far the selection of genes to clone was based on glucosidases from GH1 family in first place. However, the identification of first bacterial myrosinase, *C.* WYE1 myrosinase [118], changed the perspective in this study. This myrosinase belongs to GH3 family so made us to look for glucosidases from other families such as GH3. Four genes from *E. casseliflavus* CP1 and *E. coli* FI10944 were successfully cloned and expressed and 2 of the proteins (ecg4 and ecg39) expressed were determined to have β-glucosidase activity against p-NPG but they showed no myrosinase activity. These proteins might need some cofactors and conditions for myrosinase activity that were not investigated in this study. Two other proteins (ecg44 and ecolg3) did not show any enzyme activity. This might be because of misfolding of proteins or the proteins might need some cofactors. The optimum temperature for protein expression induction in *E. coli* BL21(DE3) cells encoding pET28b-ecolg3 was found to be 37°C. However, lowering

induction temperature could be tried to overcome misfolding risk. As ecolg3 from *E. coli* FI10944 is a periplasmic β-glucosidase, it might also need its natural environment for proper folding [239]. Moreover, there are many β-glucosidases in *E. casseliflavus* CP1 (3 out of 40 putative β-glucosidases were cloned) and *E. coli* FI10944 (1 out of 5 putative β-glucosidases were cloned) which have not been studied. Three 6-phospho-β-glucosidases encoded by the genes *bglA* and *ascB* (GH1) and *chbF* (GH4) in *E. coli* O157:H7 were examined for their effect on myrosinase activity of *E. coli* O157:H7 by Cordeiro et al. (2015). It was suggested that 6-phospho-β-glucosidase encoded by the ascB gene is associated with sinigrin degradation in *E. coli* O157:H7 [240]. Despite its low sequence identity with *C.* WYE1 myrosinase, 6-phospho-β-glucosidase encoded by ascB gene in *E. coli* FI10944 could be cloned and expressed.

The glucosinolate metabolism in the human gut is not just covered by myrosinases. The role of methionine sulphoxide reductases will be discussed in the Chapter 4 in detail. In addition, sulfatases that are responsible for transforming glucosinolates into desulfo-glucosinolates are also important. Sulfatase of *E. coli* VL8 was cloned and expressed in *E. coli* by Luang-In et al. (2013) and the enzyme was found to be active against glucosinolates [137]. Our study did not focus on sulfatases because they were already characterised. A recent study reported that *L. agilis* R16, *E. casseliflavus* CP1 and *E. coli* VL8 could degrade desulfo-glucosinolates and gave only nitriles not ITCs as products [57]. As two of the recombinant proteins of *E. casseliflavus* CP1 (ecg4, ecg39) were active and showed β-glucosidase activity against p-NPG, these proteins might be checked for their activity against desulfo-glucosinolates. The GC-MS method was set up for detection of nitriles but this was yet to be checked. If this degradation is possible then it will reveal another mechanism for degradation of glucosinolates by gut bacteria. This mechanism may include sulfatase activity followed by β-glucosidase activity to produce nitriles as products of glucosinolate degradation. Many other possible mechanisms including the importance of phosphorylation of 6-OH group of glucosinolates are suggested in the literature [137, 138]. As glucosinolates are β-thioglucosides, it was suggested that phosphorylation of glucosinolates by β-glucoside kinase before the action of myrosinase might be important for recognition of glucosinolate structure [137]. However, this hypothesis has not been tested so far.

In summary, our attempt to purify bacterial myrosinase from human gut by cloning and chromatography approaches could not succeed in identification of the myrosinase. However, the data accomplished in this study can be inspiring for future studies. Two β-glucosidases of

E. casseliflavus CP1 were successfully cloned, expressed and found to have β-glucosidase activity. These two β-glucosidases might have activity against desulfo-glucosinolates. *E. coli* FI10944 and *C. freundii* FC50 were identified with myrosinase activity and their myrosinase activity was not cell wall-associated.

CHAPTER FOUR

4 BACTERIAL REDUCTASES FROM HUMAN GUT BACTERIA INVOLVED IN GLUCOSINOLATE METABOLISM

4.1 INTRODUCTION

Methionine residues in proteins can be oxidised to methionine sulphoxides (MetSO) and this process can damage the biological function of the proteins (Figure 4.1). As methionine is very susceptible to oxidation by reactive oxygen species and nitrogen species, organisms are evolved to deal with this oxidative damage [152]. Oxidation of methionine to MetSO is reversible and MetSO can be reduced back to methionine by separate enzyme systems [155].

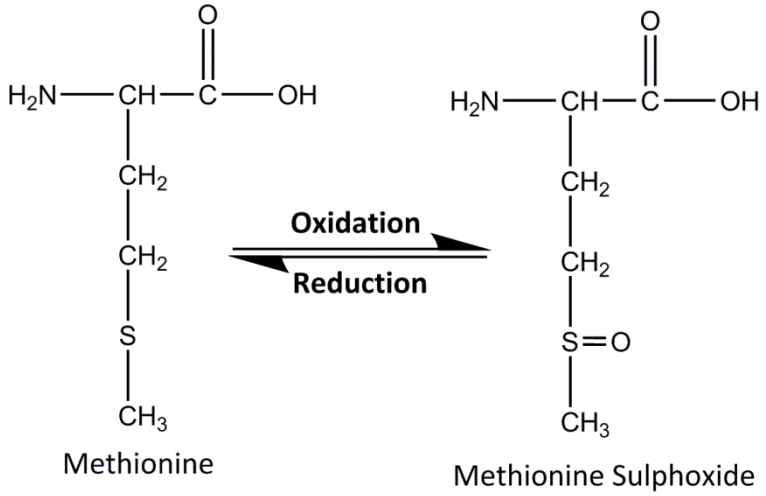


Figure 4.1 Structures of methionine and its oxidation product methionine sulphoxide Adapted from [152].

Methionine sulphoxide reductases (Msr) are the enzymes that catalyse the reduction of these sulphoxides and repair the enzyme function in many organisms [241]. These enzymes were also reported to show specificity for a range of *S*-stereomers of compounds like N-acetyl-methionine-*S*-sulphoxide and *S*-sulindac [242]. Main Msr enzymes include methionine sulphoxide reductase A (MsrA) and methionine sulphoxide reductase B (MsrB), that are responsible for reduction of *S*-epimers and *R*-epimers of MetSO respectively in free or protein bound MetSO. Another class of Msr can only reduce free MetSO not protein bound MetSO, and are called as free methionine sulphoxide reductases (fMsr). If they act on *R*-epimers of free MetSO then they are called as fRMsr. These various Msr enzymes can reduce different stereoisomers of MetSO. In addition, they were suggested to be responsible for the reduction of methylsulfinylalkyl glucosinolates like glucoraphanin and glucoiberin to methylthioalkyl glucosinolates like glucoerucin and glucoiberverin.

The reduction of methylsulfinylalkyl glucosinolates (such as glucoraphanin) is important in glucosinolate metabolism. This removes sulphoxide group of methylsulfinylalkyl glucosinolates and enables the myrosinase to act on the resulting product. A study reported the conversion of glucoraphanin to its reduced counterpart (glucoerucin) by human gut microbiota in a fermentation batch model using faecal inoculum and this conversion did not occur in heat-sterilised samples. Therefore, it was suggested that reductase activity was enzymatic [165]. This reductase activity was reported to be inducible in cell-free extracts (CFE) of *E. coli* VL8 by glucoraphanin or gluconasturtiin. It was suggested that the induction was triggered by the glucosinolate core rather than methylsulfinyl structure. The reductase activity in CFEs *E. coli* VL8 was found to be Mg^{+2} and NADPH dependent and oxygen independent [56]. In the literature, MsrA was suggested to be responsible for reduction of methylsulfinyl glucosinolates by *Escherichia coli* Nisssle 1917 and *Enterobacter cloacae* ATCC13047. However, the effect MsrA on methylsulfinyl glucosinolates is yet to be identified. In addition, the effect of different Msr enzymes on methylsulfinyl glucosinolates needs further investigation.

This chapter focused on identifying the enzymes that are responsible for reduction of methylsulfinyl glucosinolates. To achieve this, 8 putative *msr* genes from *L. agilis* R16, *E. coli* FC44 and *E. coli* VL8 were cloned with a 6X His-tag using a pET15b cloning vector and expressed in *E. coli* BL21 (DE3). The cloned genes were then induced with IPTG for protein expression and their activity was assessed.

4.2 HYPOTHESIS

Methionine sulphoxide reductases are responsible for reduction of methylsulfinyl glucosinolates (such as glucoraphanin) to their reduced counterparts (such as glucoerucin). The role of these enzymes in the reduction of glucosinolates can be identified using molecular biology techniques.

4.3 METHODS

4.3.1 Testing Reductase Activity of Human Gut Bacteria

E. coli VL8 was obtained using a sinigrin enrichment method from Luang-In et al. (2014) [56] and *E. coli* FC44 was isolated using a glucoraphanin enrichment (Chapter 2-section 2.1.3) method in this study. *E. coli* VL8 was tested for its ability to reduce glucoraphanin to

glucoerucin. *E. coli* VL8 was grown in Nutrient Broth with 1 mM glucoraphanin anaerobically for 16 h at 37°C. The cultures (n=3) were centrifuged at 16000 x g for 15 min at 4°C. Supernatants were removed and filtered through 0.2 µm PVDF HPLC filters (Avonchem). HPLC samples were prepared from these supernatants (200 µl used) and analysed as described in Chapter 2-section 2.4.1 to determine the reduction of glucoraphanin to glucoerucin.

CFEs of *E. coli* VL8 and *E. coli* FC44 were examined to see if reduction of glucosinolates is inducible by glucosinolates. The cells were cultured in 1 ml of Nutrient Broth with 1 mM glucoraphanin (for glucoraphanin induced samples) or without (for uninduced samples). These cultures were incubated at 37°C anaerobically for 24 h then the cultures were centrifuged at 4000 x g for 15 min at 4°C. Supernatants were removed, pellets were washed with 0.1 M citrate phosphate buffer pH 7.0 to remove any traces of glucosinolate, isothiocyanate or nitrile products. The washed pellets were resuspended in 0.1 M citrate phosphate buffer pH 7.0 with 100X protease inhibitor cocktails (Melford). Sonication was used to prepare CFEs of *E. coli* VL8 and *E. coli* FC44 (Method is in Chapter 2-section 2.3.1). These CFEs were used in assays to test reduction of glucoraphanin. The activity was tested using similar conditions that were reported previously to be optimum for reductase activity of *E. coli* VL8 [56]. The reaction mixture (1 ml) contained 300 µl of CFE, 0.25 mM glucoraphanin in 0.1 M citrate phosphate buffer pH 7.0 (n=3). Glucoraphanin controls which contain only 0.25 mM glucoraphanin in 0.1 M citrate phosphate buffer pH 7.0 were also included. These reactions were anaerobically incubated for 16 h at 37°C. After incubation, samples were centrifuged at 16000 x g for 15 min at 4°C. Supernatants were removed and filtered through 0.2 µm PVDF HPLC filters. Samples were prepared using 200 µl of the supernatants and subjected to HPLC as described in Chapter 2-section 2.4.1.

4.3.2 Identification of Reductase Genes to Clone and Express

The genomes of *L. agilis* R16, *E. coli* FC44 and *E. coli* VL8 were sequenced, assembled and annotated as described in Chapter 2-section 2.2.13. The genomes were interrogated for methionine sulphoxide reductase genes. All of the Msr genes in *L. agilis* R16, *E. coli* FC44 and *E. coli* VL8 were selected to identify the responsible enzyme group of Msr for glucosinolate reduction.

4.3.3 Cloning and Expression of Reductases

The 3 *msr*A genes of *L. agilis* R16 were cloned in first place. The primers introducing new restriction sites were designed to amplify the candidate genes and are listed in Table 4.1. The nucleotide sequences of the cloned genes are shown in the Appendix 2. PCR was performed using High Fidelity Phusion Polymerase (Finnnzymes) as described in Chapter 2-section 2.2.1. Gel electrophoresis (Method is in Chapter 2-section 2.2.2) was used to analyse the PCR gene products (Table 4.2). PCR conditions and expected product sizes are given in Table 4.2.

Name	Primer Sequence (5'-3')	Restriction Site
MSRA1-LBAG-F	ATT**C**A**T**ATGACAGAGACTGCAATATT	NdeI
MSRA1-LBAG-R	AATC**G**GA**T**CCATTTTCAACGCTTAA	Bam HI
MSRA2-LBAG_F	ATTC**AT**ATGGAAACAGCAATTTTTG	NdeI
MSRA2-LBAG-R	CAACTC**GA**GCTTAGTCCACTTTAGGTG	XhoI
MSRA3-LBAG-F	GAC**CAT**ATGAAAACAAAACCTGAAG	NdeI
MSRA3-LBAG-R	AAGTC**TCG**AGTTGTGACTGAACCGC	XhoI
MSRB-LBAG-F	GATA**CAT**ATGGATAAACAGCAAGG	NdeI
MSRB-LBAG-R	ATTCCTC**GA**GTTTATTAACTCAATTATTTA	XhoI
fRMsr-LBAG-F	TTT**CAT**ATGGCTACCACGGAA	NdeI
fRMsr-LBAG-R	TGTT**CT**CGAGCTTACCCACTC	XhoI
fRMsr-FC44-F	CAT**CAT**ATGCTTATTTTAATAGGCTG	NdeI
fRMsr-FC44-R	TAA**G**GA**TC**CTATCAGTAAATGCTACGTT	Bam HI
MsrA-VL8-F	CGAC**CAT**ATGAGTTTATTTGATAAAAAG	NdeI
MsrA-VL8-R	TTTG**G**A**T**CCGTAACGCTATGCT	Bam HI
MsrB-VL8-F	TGAGCA**T**ATGGCTAATAAACCTTC	NdeI
MsrB-VL8-R	CATGCTC**GA**GATAATGTCATCAAGATTC	XhoI

Table 4.1 Primers used in the PCR experiments to amplify the candidate *msr* genes. The restriction sites are underlined. The altered nucleotides to introduce the restriction sites are in bold.

Name	T_A, time, cycles	T_E, time	Product size
MSRA1-LBAG-F MSRA1-LBAG-R	54°C, 30 s, 5X 64°C, 30 s, 20X	72°C, 15 s	559 bp
MSRA2-LBAG-F MSRA2-LBAG-R	54°C, 30 s, 5X 67°C, 30 s, 20X	72°C, 15 s	547 bp
MSRA3-LBAG-F MSRA3-LBAG-R	55°C, 30 s, 5X 67°C, 30 s, 20X	72°C, 30 s	839 bp
MSRB-LBAG-F MSRB-LBAG-R	49°C, 30 s, 5X 64°C, 30 s, 20X	72°C, 15 s	462 bp
fRMsr-LBAG-F fRMsr-LBAG-R	52°C, 30 s, 5X 67°C, 30 s, 20X	72°C, 15 s	538 bp
fRMsr-FC44-F fRMsr-FC44-R	52°C, 30 s, 5X 64°C, 30 s, 20X	72°C, 20 s	595 bp
MsrA-VL8-F MsrA-VL8-R	58°C, 30 s, 5X 66°C, 30 s, 20X	72°C, 20 s	661 bp
MsrB-VL8-F MsrB-VL8-R	49°C, 30 s, 5X 65°C, 30 s, 20X	72°C, 20 s	492 bp

Table 4.2 Conditions used in the PCR experiments to amplify the candidate *msr* genes. T_A; annealing temperature, cycles; number of cycles used in annealing step of PCR. T_E; extension temperature and time used in PCR, Product size; the expected PCR product sizes for *msr* genes. The annealing step was performed using T_m of 100% sequence match part of primers first (5X) and then using the T_m of entire primer (20X).

Restricted PCR gene products were ligated to double digested pET15b vectors. The recombinant plasmids (Table 4.3) were transformed into *E. coli* Top10 competent cells (Thermo-Fischer Scientific) for *msr*A1*, msr*A2*, msr*A3 and into *E. coli* DH5α chemically competent cells (homemade, See Chapter 2-section 2.2.9) for the rest. Then cells were plated on L agar with ampicillin plates (100 µg /ml final concentration). Colonies were screened by colony PCR using primers T7P2 (5'-TGAGCGGATAACAATTCCC-3') and T7T (5'-GCTAGTTATTGCTCAGCGG-3') as described in Chapter 2-section 2.2.11. The clones that gave positive result for PCR were selected for plasmid purification. The recombinant plasmids were purified using EZNA Plasmid Mini Kit II (Omega Bio-Tek) and sequenced using T7P2 and T7T primers. Depending on the sequencing results, one of the positive transformants without any mutation was selected and transformed into *E. coli* BL21(DE3) for protein expression studies. Protein expression in *E. coli* BL21(DE3) encoding *msr* genes was induced with IPTG at a final concentration of 0.5 mM. For *E. coli* BL21(DE3) expressing pET15b-LBAG_*msr*A1, pET15b-LBAG_*msr*A2 pET15b-LBAG_*msr*A3, inductions were carried out for 4 h at 37°C or 24 h at 25°C, shaking at 250 rpm. For the rest, protein expressions were held at 25°C for 4 h and 37°C for 3 h or 4 h to determine optimum conditions. Control cultures with *E. coli* BL21(DE3) pET15b grown in presence of 0.5 mM IPTG (empty vector controls) and control cultures of *E. coli* BL21(DE3) expressing the cloned genes grown without 0.5 mM IPTG (uninduced cultures) were also included.

Plasmid	Relevant Characteristics
pET15b-LBAG_*msr*A1	*msr*A1 gene insert from *L. agilis* R16 ligated to pET15b
pET15b-LBAG_*msr*A2	*msr*A2 gene insert from *L. agilis* R16 ligated to pET15b
pET15b-LBAG_*msr*A3	*msr*A3 gene insert from *L. agilis* R16 ligated to pET15b
pET15b-LBAG_*msr*B	*msr*B gene insert from *L. agilis* R16 ligated to pET15b
pET15b-LBAG_*frmsr*	*frmsr* gene insert from *L. agilis* R16 ligated to pET15b
pET15b-FC44_*frmsr*	*frmsr* gene insert from *E. coli* FC44 ligated to pET15b
pET15b-VL8_*msr*A	*msr*A gene insert from *E. coli VL8* ligated to pET15b
pET15b-VL8_*msr*B	*msr*B gene insert from *E. coli VL8* ligated to pET15b

Table 4.3 Plasmids used and generated in this study.

4.3.4 Protein Extraction by Bugbuster

Bugbuster protein extraction reagent (Novagen) was used to extract proteins according to supplier's advice. To extract the soluble fraction, *E. coli* cells were harvested from culture by centrifugation using a preweighed tube at 3220 x g for 15 min at 4°C. The supernatant was removed and pellet was allowed to drain, the wet pellet was weighed and frozen at -20°C overnight. Next day, the frozen cell pellet was resuspended at room temperature by pipetting or gentle vortexing (5 ml of Bugbuster per gram of wet cell paste). The suspension was incubated on a shaking platform slowly for 10 min at room temperature. Then the sample was centrifuged at 13,000 x g for 20 min at 4°C. The supernatant was transferred to a fresh tube as soluble extract. To purify inclusion bodies, the pellet was resuspended in the same volume of Bugbuster HT with a final concentration of 200 µg/ml lysozyme. The suspension was vortexed and incubated at room temperature for 5 min then 6 volumes of 1/10 diluted Bugbuster HT in deionized water was added to the suspension and vortexed for 1 min. The suspension was centrifuged at 13000 x g for 15 min at 4°C and the supernatant was removed. The inclusion bodies were suspended in ½ of the original culture volume of 1/10 diluted Bugbuster HT, mixed by vortexing and centrifuged at 3220 x g for 5 min at 4°C. This step was repeated for 2 more times. The final pellet was dissolved in 20 mM Tris-HCl 50 mM NaCl buffer pH 7.5.

4.3.5 Purification of Recombinant Proteins

After protein expression induction, the cells were centrifuged at 3220 x *g* for 15 min at 4°C. The cell paste was used to prepare CFE. The CFEs were prepared by sonication as described in Chapter 2-section 2.3.1 and run on a SDS-PAGE gel (denaturing, reducing conditions) including negative controls too. The CFEs were checked for the presence of 6XHis-tag by Western Blot (Chapter 2-section 2.3.4) and partially purified by Ni-NTA column as described in Chapter 2-section 2.3.5.

After Ni-NTA purification, proteins were dialysed overnight to remove excessive amount of imidazole. The proteins were dialysed in Spectra/Por porous membrane tubing with 500-1000 Da cut off (Spectrum Labs). The membrane tubing was soaked in MQ water for 30 min in advance and rinsed with MQ water before use. The dialysis was performed in an appropriate amount of buffer (at least 100X bigger in volume of protein volume) for 18 h at 4°C. The buffer used in dialysis was 25 mM Tris-HCl buffer pH 7.5 for LBAG_MsrA1, LBAG_MsrA2 and LBAG_MsrA3. The dialysis of LBAG-MsrB and VL8_MsrB was carried out using 0.1 M citrate

phosphate buffer pH 7.0. The proteins were removed from the tubing and transferred into new clean tubes. Quantification of protein amounts was performed by Bradford assay as described in Chapter 2-section 2.3.2.

4.3.6 Testing Reductase Activity of Proteins

To test glucoraphanin reduction by CFEs or Ni-NTA column purified proteins, reactions were set up under different conditions. The optimum conditions reported for reductase activity of *E. coli* VL8 [56] and the studies about Msr [150, 243] were taken into account and the methods were adapted from these studies.

Ni-NTA purified and dialysed (in 25 mM Tris-HCl buffer pH 7.5) eluates of LBAG_MsrA1 and LBAG_MsrA2 were used to determine reductase activity in the presence of NADPH or dithiothreitol (DTT). For DTT, the reactions contained 25 mM Tris-HCl pH 7.5, 20 mM DTT, 1 mM glucoraphanin and 10 µg of eluates with or without 10 mM $MgCl_2$ and incubated for 30 min at 37°C. In addition, the effect of increasing the protein amount to 50 µg and incubation time to 1, 2 and 3 h were also tested using Ni-NTA purified MsrA1. For NADPH, the reactions were set up in 200 µl consisting 1 mM $MgCl_2$, 1 mM NADPH, 0.25 mM glucoraphanin in 0.1 M citrate phosphate buffer pH 7.0 and 10 µg of protein was used in the assays. A glucoraphanin control was set up as samples excluding protein. The reactions were incubated for 16 h at 37°C aerobically.

CFEs of *E. coli* BL21 (DE3) expressing pET15b vector (empty vector control), pET15b-LBAG_*msr*B, pET15b-LBAG_*frmsr*, pET15b-FC44_*frmsr, pET15b-VL8_msr*A and pET15b-VL8_*msr*B were prepared using 50 mM Na phosphate buffer pH 7.5 by sonication (Chapter 2-section 2.3.1), quantified by Bradford assay (Chapter 2-section 2.3.2) and tested for glucoraphanin reducing ability. Briefly, a 200 µl reaction mixture contained a buffer of 50 mM Na phosphate buffer pH 7.5, 50 mM NaCl, 20 mM DTT, 0.5 mM glucoraphanin and 100 µg of protein. Two controls, control 1 excluding protein and control 2 excluding protein and DTT, were set up as samples. Samples and controls were incubated for 16 h at 37°C aerobically.

To test the reductase activity of purified, dialysed (in 0.1 M citrate phosphate buffer pH 7.0) eluates of LBAG_MsrB and VL8_MsrB, the reactions were set up under two different conditions. First, the reactions were set up in 0.1 M citrate phosphate buffer pH 7.0 with 10 mM $MgCl_2$, 1 mM NADPH and 0.5 mM glucoraphanin. Second, the activity was tested in 50 mM Na phosphate buffer pH 7.5 with 50 mM NaCl, 20 mM DTT, 0.5 mM glucoraphanin. The

reactions were both set up in a total volume of 200 µl and 10 µg of protein eluates were used. Glucoraphanin controls were set up as samples excluding proteins. The samples and controls were incubated at 37°C for 16 h aerobically.

After each incubation period, 400 µl of acetonitrile was added to stop the reaction. Samples and controls were filtered through 0.2 µm PVDF HPLC filters. Finally, all these reactions were used in sample preparation for HPLC and samples were run on HPLC as described in Chapter 2-section 2.4.1 to determine reduction of glucoraphanin to glucoerucin.

4.4 RESULTS

4.4.1 Reductase Activity of Human Gut Bacteria

Previously, *E. coli* VL8 was found to reduce glucoraphanin to glucoerucin by Luang-In et al. (2013) [56]. In this study, we examined the reductase activity of *E. coli* VL8 and *E. coli* FC44 to transform glucoraphanin to its reduced counterpart glucoerucin. *E. coli* VL8 was able to reduce glucoraphanin in Nutrient media after 16 h incubation and it consumed 40.9 ± 1.5% of the glucoraphanin. However, the glucoerucin formation did not match to the glucoraphanin consumption by cells. The glucoerucin amount was only 1.1 ± 0.2% of the original glucoraphanin amount added to the media. As reduction would produce one glucoerucin molecule for one glucoraphanin molecule, the depletion of glucoerucin might be explained by further metabolism of glucoerucin by myrosinase of *E. coli* VL8 during 16 h incubation period. Identification of these degradation products was not performed in this study.

E. coli FC44 is a gut bacterium isolated during this study using a glucoraphanin enrichment method as described in Chapter 2-section 2.1.3. During this screening, *E. coli* FC44 was found to produce glucoerucin so it was also included in reductase activity assays (Figure 4.2).

CFEs of *E. coli* VL8 and *E. coli* FC44 were prepared by sonication. Cultures with glucoraphanin (induced) and without glucoraphanin (uninduced) were used for CFE preparation to test if there is an inducible reductase activity in these strains. The reduction rates were based on glucoerucin formation and the results are given in Table 4.4. Induced and uninduced CFEs of *E. coli VL8* showed glucoraphanin reduction at a rate of 1.15 ± 0.20 nmol glucoerucin/µg protein and 0.89 ± 0.05 nmol glucoerucin/µg protein respectively. There was no significant difference for reduction rate of induced or uninduced CFEs of *E. coli* VL8 ($p=0.10$, two sample two tailed t-test). Glucoraphanin reduction rates were 1.75 ± 0.32 nmol glucoerucin/µg

protein for induced and 2.01 ± 0.09 nmol glucoerucin/µg protein for uninduced CFEs of *E. coli* FC44 respectively. CFEs of *E. coli* FC44 were found to be more efficient in reducing glucoraphanin to glucoerucin. No significant difference was found between induced or uninduced CFEs of *E. coli* FC44 (p=0.26, two sample two tailed t-test), suggesting that reductase activity is constitutive.

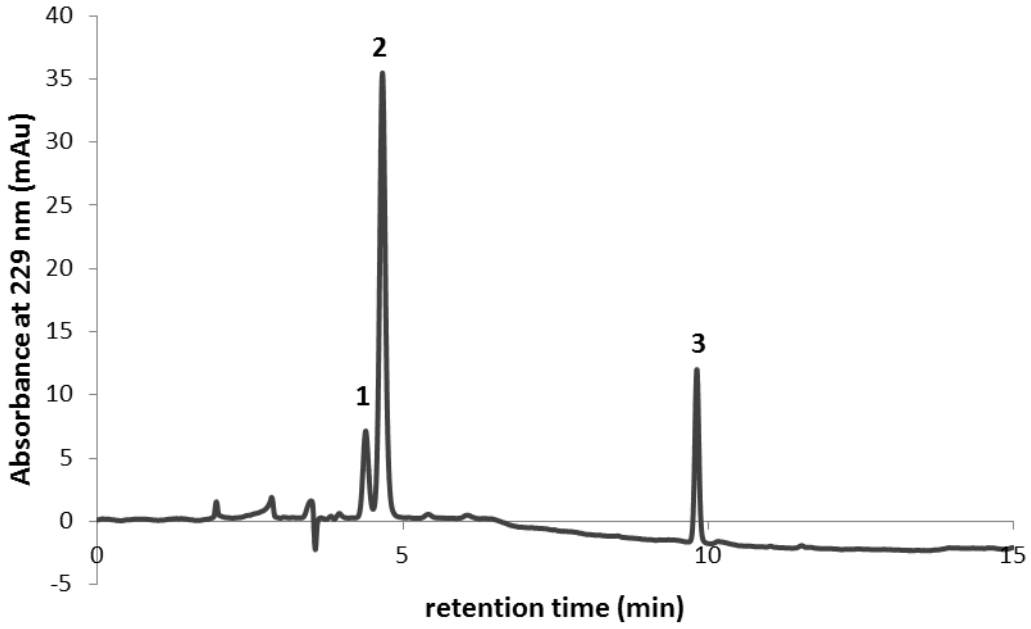


Figure 4.2 HPLC chromatogram showing reduction of glucoraphanin to glucoerucin by *E. coli* FC44. 1; glucoraphanin, 2; sinigrin (internal standard), 3; glucoerucin.

CFEs	Reduction rate (nmol glucoerucin formed/µg protein)
Induced CFE of *E. coli* VL8	1.15 ± 0.20
Uninduced CFE of *E. coli* VL8	0.89 ± 0.05
Induced CFE of *E. coli* FC44	1.75 ± 0.32
Uninduced CFE of *E. coli* FC44	2.01 ± 0.09

Table 4.4 The reduction of glucoraphanin to glucoerucin by glucoraphanin-induced and uninduced CFEs of *E. coli* VL8 and *E. coli* FC44. The results are given as mean of triplicates ± SD.

4.4.2 Identification of Putative Sulphoxide Reductases

The study started with cloning and expression of putative *msr*A genes of *L. agilis* R16. Three putative methionine sulphoxide reductase A (*msr*A), one methionine sulphoxide reductase B

(*msr*B) and one free methionine-*R*-sulphoxide reductase (*frmsr*) were identified in the genome of *L. agilis* R16. However, *E. coli* FC44 and *E. coli* VL8 only have single copies of each *msr* genes. *E. coli* FC44 and *E. coli* VL8 share the same *msr*A nucleotide sequence. There are a few nucleotide differences between *msr*B nucleotide sequences of *E. coli* FC44 and *E. coli* VL8 but they share the same amino acid sequence for MsrB. *E. coli* FC44 and *E. coli* VL8 fRMsr also share same amino acid sequence except two amino acid residues, so only fRMsr in *E. coli* FC44 was selected for cloning. Amino acid sequences of MsrA, MsrB and fRMsr from the Uniprot database were aligned with MsrA (Figure 4.3), MsrB (Figure 4.4) and fRMsr (Figure 4.5) of *L. agilis* R16, *E. coli* FC44 and *E. coli* VL8 using Clustal Omega [237].

Moskovitz *et*. *al*. (2000) reported the presence of a shared amino acid sequence among MsrAs (GCFWG). Site-directed substitution of each amino acid residue in this conserved sequence was found to destroy enzyme activity. However, mutation of Gly residues was reported to result in a partial enzyme activity loss [243]. *E. coli* VL8 and *E. coli* FC44 have this conserved sequence (GCFWG) in their MsrA but *L. agilis* R16 only shares the 'GCFW' part (Shown in Figure 4.3).

Hydrogen bond interactions between Glu 94, Tyr 82 and Tyr 134 in MsrA of *E. coli* to stabilise the oxygen of the sulphoxide was reported by Antonie et al. (2006) [149] and Glu 94 was proposed to be an acid/base catalyst [244]. These residues were conserved in MsrA of *E. coli* VL8 and *E. coli* FC44 at positions 95, 83 and 135 respectively (Figure 4.3). In addition the role of Cys 51 as the nucleophile was reported [244] and this residue is conserved at position 52 (GCFWG sequence) in *E. coli* FC44 and *E. coli* VL8 (Figure 4.3). Some bacteria, like *Neisseria gonorrhoeae*, have a fused form of MsrA with MsrB, named as MsrAB [151]. MsrAB was not found in the bacteria we studied but MsrAB of *Neisseria gonorrhoeae* was also included in the multiple alignments to see the difference among different types of MsrA. In summary, MsrA of *E. coli* VL8 and 3 MsrAs of *L. agilis* R16 were selected for cloning.

```
                      *        20         *        40         *
MSRA_MYCTU : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA3_LBAG : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA1_LBAG : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA2_LBAG : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRAB_NEIG : MKHRTFFSLCAKFGCLLALGACSPKIVDAGTATVPHTLSTLKTADNRPAS :  50
MSRA2_ARAT : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA_ECO12 : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA_FC44  : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA_VL8   : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA_MOUSE : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA_HUMAN : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA_BOVIN : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -

                     60         *        80         *       100
MSRA_MYCTU : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA3_LBAG : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA1_LBAG : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA2_LBAG : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRAB_NEIG : VYLKKDKPTLIKFWASWCPLCLSELGQAEKWAQDAKFSSANLITVASPGF : 100
MSRA2_ARAT : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA_ECO12 : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA_FC44  : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA_VL8   : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA_MOUSE : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA_HUMAN : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA_BOVIN : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -

                      *       120         *       140         *
MSRA_MYCTU : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA3_LBAG : MKTKPEALMRDLYNLIINP------A-TRDWERHLLVQAKNNRQQLSPTG :  43
MSRA1_LBAG : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRA2_LBAG : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ :   -
MSRAB_NEIG : LHEKKDGEFQKWYAGLNYPKLPVVTDNGGTIAQNLNISVYPSWALIGKDG : 150
MSRA2_ARAT : ~~~~~~~~~~~~~~~~~~~~~~MDS----------SL---------- :   5
MSRA_ECO12 : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MS----------L :   3
MSRA_FC44  : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MS----------L :   3
MSRA_VL8   : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MS----------L :   3
MSRA_MOUSE : ~~~~~~~~~~~~~~~~~~~~~~MLSASRRAL--QLLSSANPVR----RMG :  22
MSRA_HUMAN : ~~~~~~~~~~~~~~~~~~~~~~MLSATRRACQLLLLHSLFPVP----RMG :  24
MSRA_BOVIN : ~~~~~~~~~~~~~~~~~~~~~~MLSATRRAL--QLFHSLFPIP----RMG :  22
```

Figure 4.3 Multiple alignments of Recombinant MsrA proteins with other MsrAs from different organisms. The abbreviations used for MsrA from: MYCTU (P9WJM5); *Mycobacterium tuberculosis*, LBAG; *L. agilis* R16, NEIG (P14930); *Neisseria gonorrhoeae*, ARAT (Q9LY15); *Arabidopsis thaliana*, ECO12 (P0A744); *E. coli* strain K12, FC44; *E. coli* FC44, VL8; *E. coli* VL8, MOUSE (Q9D6Y7); *Mus musculus*, HUMAN (Q9UJ68); *Homo sapiens*, BOVIN (P54149); *Bos taurus*. The default settings were used for Clustal Omega. Amino acid sequences were obtained from Uniprot database and the entry numbers of sequences are given in brackets. The 'GCFWG' sequence is shown and highlighted in red. Glu 95 (E) and Cys 52 (C) residues in *E. coli* VL8 and *E. coli* FC44 are highlighted in yellow. Tyr 83 (Y) and Tyr 135 (Y) residues in *E. coli* VL8 and *E. coli* FC44 are highlighted in pink.

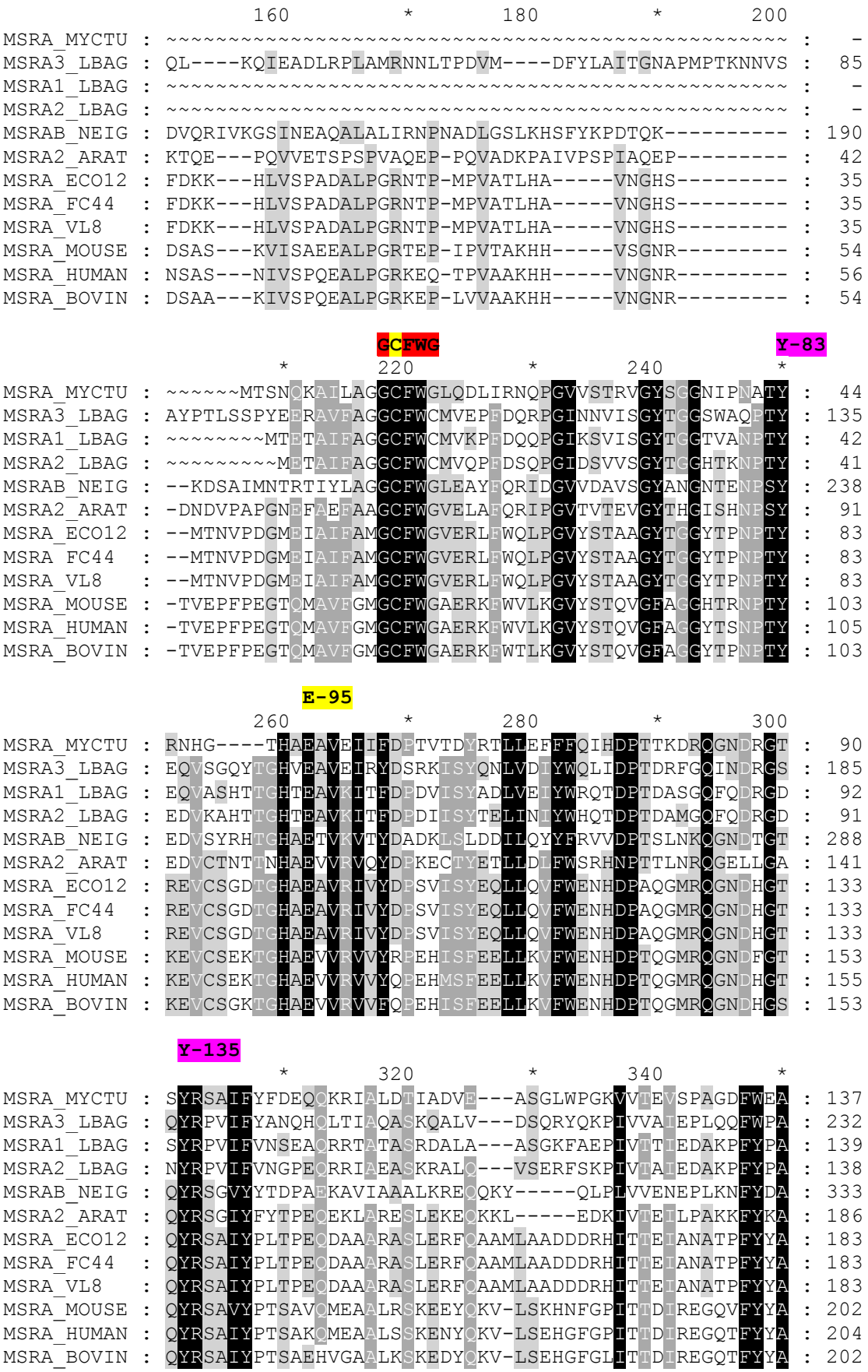


Figure 4.3-Continued

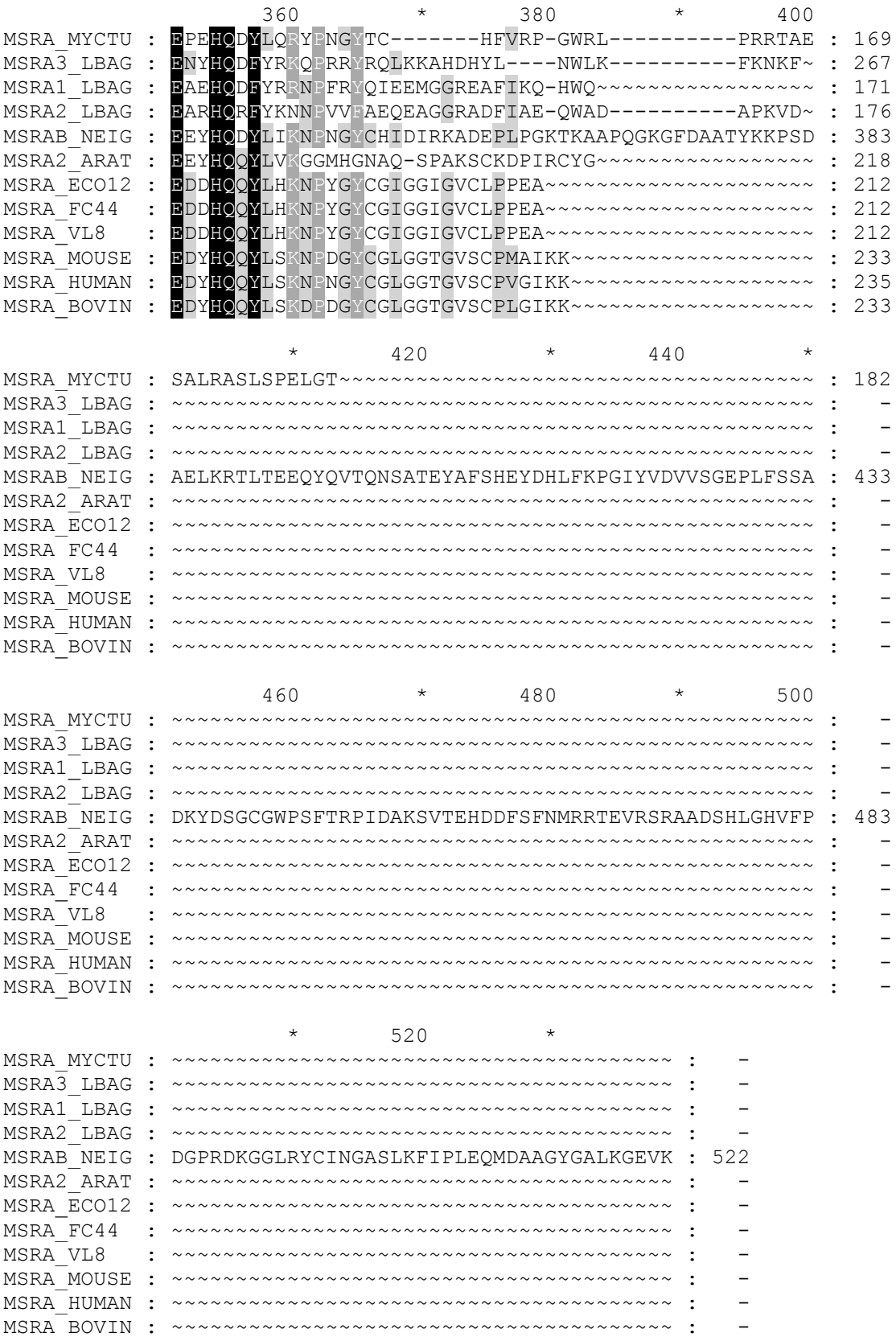


Figure 4.3-Continued

Many MsrB from various organisms were aligned with recombinant MsrB proteins (Figure 4.4). His 103, His 100, Asn 119 and Thr 26 amino acid residues in MsrB of *E. coli* were reported to form hydrogen bonds to stabilise the oxygen of the sulphoxide. These residues were conserved in MsrB of *E. coli* FC44 and *E. coli* VL8 at positions 104, 101, 120 and 27 respectively (Figure 4.4). In addition, His 103 amino acid residue was suggested to be the acid/base catalyst [244] and conserved at His 104 in MsrB of *E. coli* VL8 and *E. coli* FC44. MsrB of *E. coli* VL8 and *L. agilis* R16 were selected for cloning.

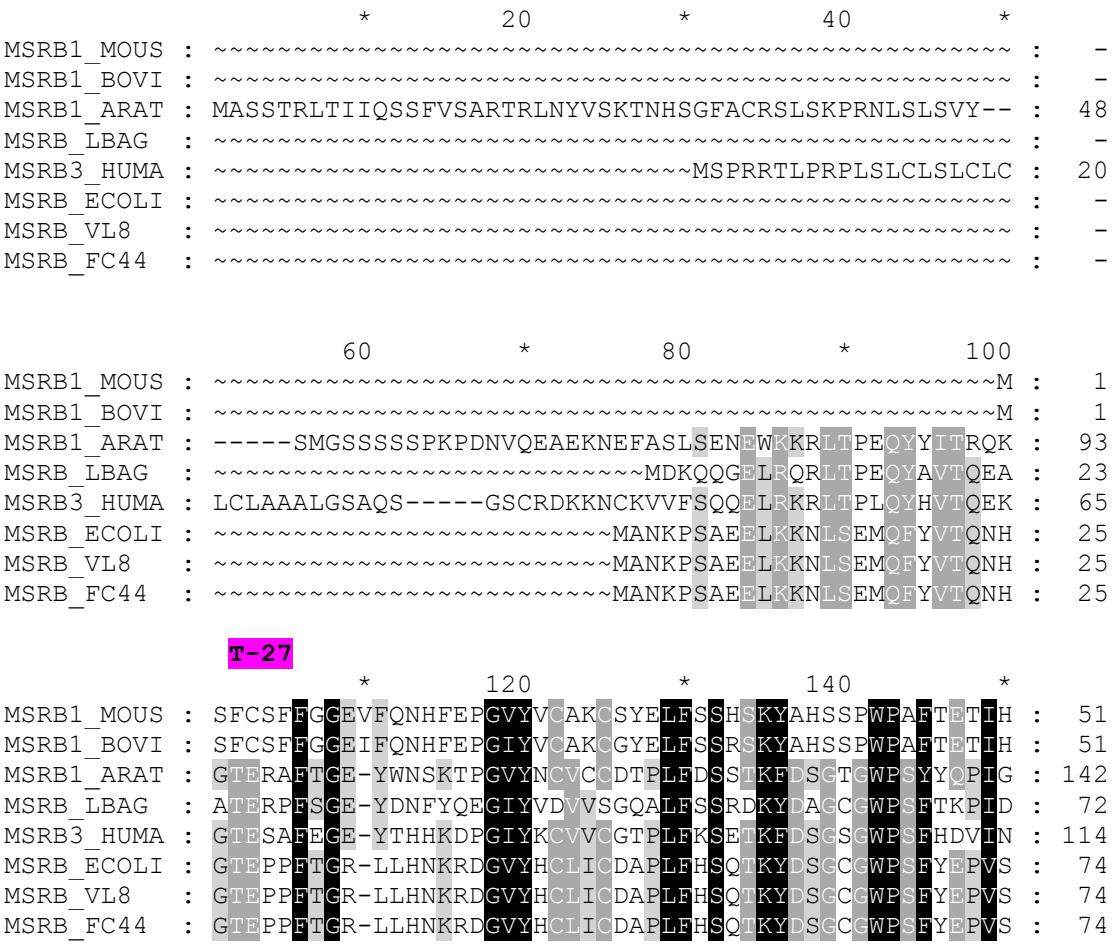


Figure 4.4 Multiple alignments of MsrB proteins from different organisms. The abbreviations used for MsrB from: MOUS (Q9JLC3); *Mus musculus*, BOVI (Q3MHL9); *Bos taurus*, ARAT (Q9C8M2); *Arabidopsis thaliana*, LBAG; *L. agilis* R16, HUMAN (Q81XL7); *Homo sapiens*, ECOLI (P0A746); *E. coli* strain K12, VL8; *E. coli* VL8, FC44; *E. coli* FC44. The default settings were used for Clustal Omega. Amino acid sequences were obtained from Uniprot database and the entry numbers of sequences are given in brackets. His 104 (H) residue in *E. coli* VL8 and *E. coli* FC44 is highlighted in yellow. Thr 27 (T), His 101 (H) and Asn 120 (N) residues in *E. coli* VL8 and *E. coli* FC44 are highlighted in pink.

Figure 4.4-Continued

Many fRMsr from various organisms were aligned with fRMsr proteins in *L. agilis* R16, *E. coli* FC44 and *E. coli* VL8 (Figure 4.5). Cys 94 residue was proposed to be a nucleophile in fRMsr of *E. coli* [245] which was conserved as Cys 76 in *E. coli* FC44 and *E. coli* VL8. This residue was not conserved in fRMsr of *L. agilis* R16. In addition, the acid/base catalyst in fRMsr of *E. coli* was proposed to be Asp 143 by Boschi-Muller et al. (2014) and was conserved at Asp 124 in *L. agilis* R16 and at Asp 125 in *E. coli* FC44 and *E. coli* VL8. fRMsr of *E. coli* FC44 and *L. agilis* R16 were selected for cloning. The initiation of the protein synthesis of fRMsr in *E. coli* strain in the studies [244, 245] starts with a different Met residue and this results in an 18 amino acid shift.

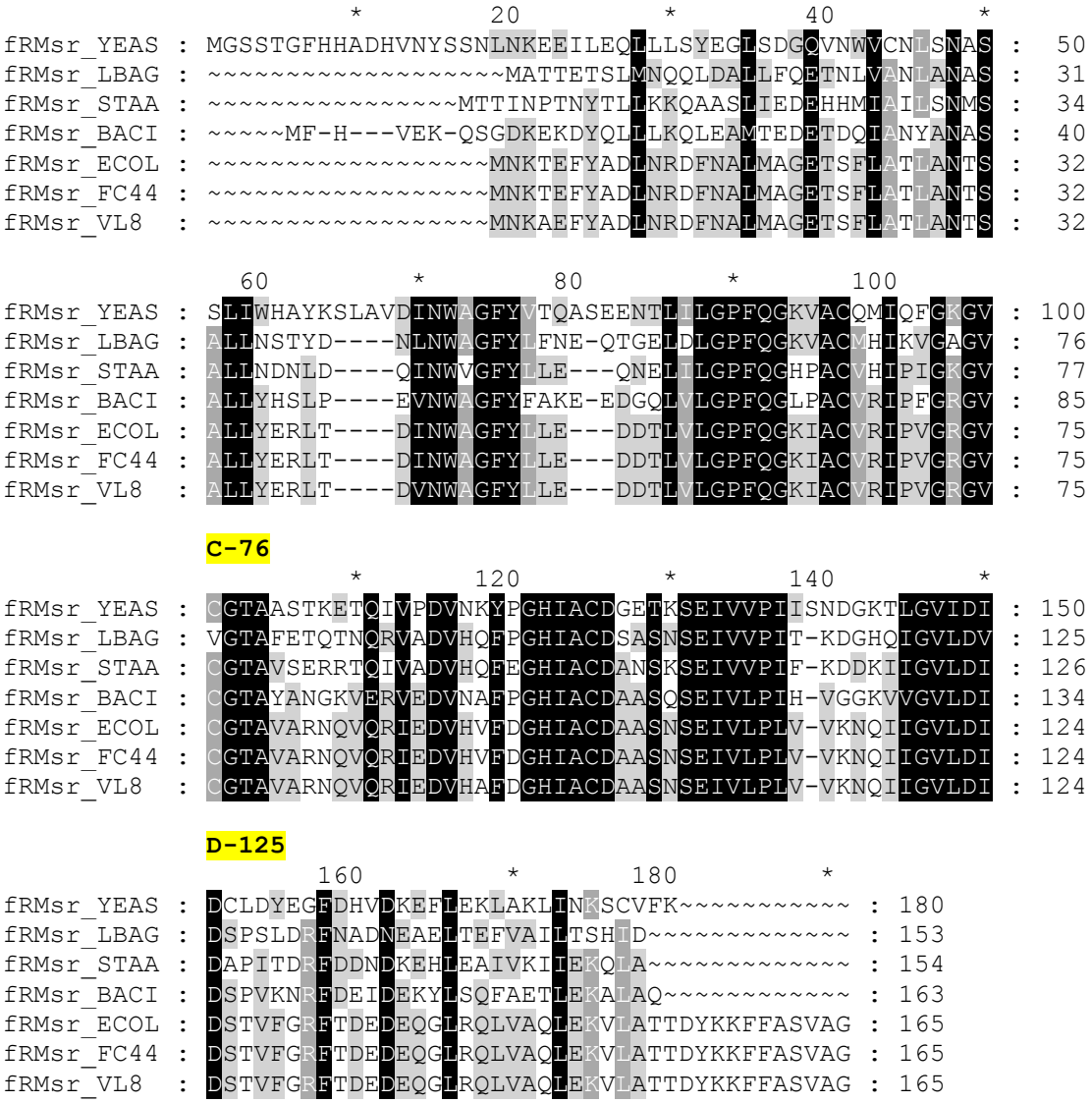


Figure 4.5 Multiple alignments of recombinant fRMsr proteins with fRMsrs from different organisms. The abbreviations used for fRMsr from: YEAST (P36088); *Saccharomyces cerevisiae*, LBAG; *L. agilis* R16, STAA (A0A0D6HHM2); *Staphylococcus aureus*, BACI (A0A0K6L851); *Bacillus subtilis*, ECOL (P76270); *E. coli* strain K12, FC44; *E. coli* FC44, VL8; *E. coli* VL8. The default settings were used for Clustal Omega. Amino acid sequences were obtained from Uniprot database and the entry numbers of sequences are given in brackets. Cys 76 (C) and Asp 125 (D) amino acid residues in *E. coli* VL8 and *E. coli* FC44 are highlighted in yellow.

4.4.3 Cloning of Methionine Sulphoxide Reductases from *L. agilis* R16, *E. coli* VL8 and *E. coli* FC44

The cloned *msr* genes can be listed as: 3 *msr*A, 1 *msr*B and 1 *frmsr* of *L. agilis* R16; 1 *msr*A and 1 *msr*B of *E. coli* VL8; 1 *frmsr* of *E. coli* FC44 (Table 4.5).

Bacteria	Cloned *msr* genes	Abbreviation
L. agilis R16	3 *msr*A	LBAG_ *msr*A
	1 *msr*B	LBAG_ *msr*B
	1 *frmsr*	LBAG_ *frmsr*
E. coli VL8	1 *msr*A	VL8_ *msr*A
	1 *msr*B	VL8_ *msr*B
E. coli FC44	1 *frmsr*	FC44_*frmsr*

Table 4.5 Cloned *msr* genes of *L. agilis* R16, *E. coli* V8 and *E. coli* FC44.

After transformation of recombinant vectors into *E. coli* DH5α cells, many transformants were checked by colony PCR to test if they have correct gene inserts. The expected sizes of PCR products with correct gene inserts from all transformants were observed (Figure 4.6). One of the positive transformants was used for recombinant plasmid preparation and sequenced. Positive recombinant plasmids were transformed into *E. coli* BL21(DE3) for protein expression and the protein expression induction was performed.

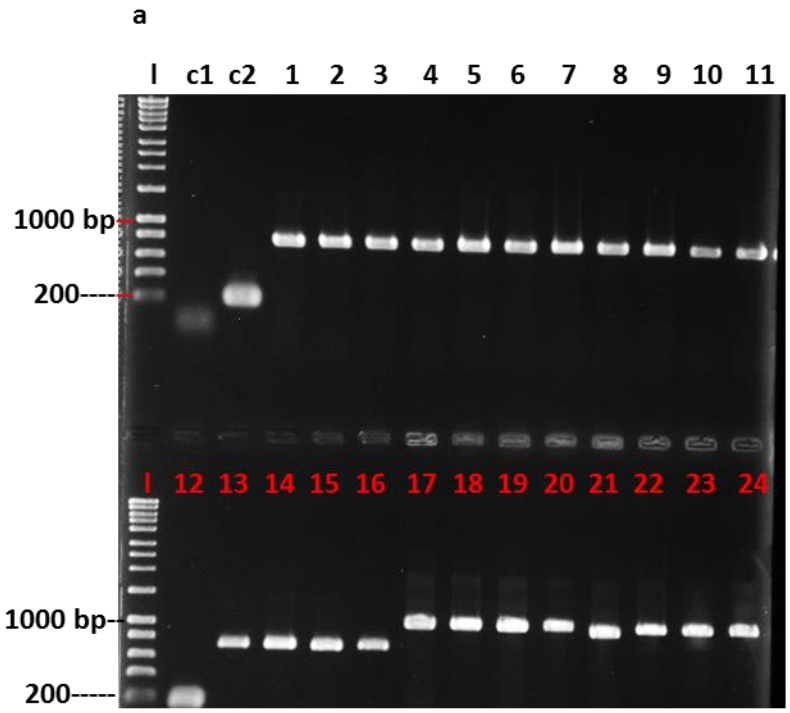


Figure 4.6 Agarose gel electrophoresis from colony PCR experiments. a) 1-8; selected colonies for *E. coli* DH5α pET15b-LBAG_*msr*A1, 8-16; selected colonies for *E. coli* DH5α pET15b-LBAG_*msr*A2, 17-24; selected colonies for *E. coli* DH5α pET15b-LBAG_*msr*A3. b)1-5; selected colonies for *E. coli* DH5α pET15b-LBAG_*msr*B, 6-10; selected colonies for *E. coli* DH5α pET15b-LBAG_*frmsr*, 11-15; selected colonies for *E. coli* DH5α pET15b-FC44_*frmsr*, 16-20; selected colonies for *E. coli* DH5α pET15b-VL8_*msr*A, 21-24; selected colonies for *E. coli* DH5α pET15b-VL8_*msr*B. c)1-7; selected colonies for *E. coli* DH5α pET15b-LBAG_*frmsr*, 8-14; selected colonies for *E. coli* DH5α pET15b-FC44_*frmsr*, 15-24; selected colonies for *E. coli* DH5α pET15b-VL8_*msr*B. l; hyperladder, c1; negative control for PCR (PCR was set up without any genomic DNA), c2; positive control for PCR (PCR was set up with pET15b vector). The expected PCR product sizes; 198 bp for pET15b vector with no insert, 709 bp for pET15b-LBAG_*msr*A1 plasmid, 727 bp for pET15b-LBAG_*msr*A2 plasmid, 1018 bp for pET15b-LBAG_*msr*A3 plasmid, 643 bp for pET15b-LBAG_*msr*B, 717 bp for pET15b-LBAG_*frmsr* plasmid, 770 bp for pET15b-FC44_*frmsr* plasmid, 835 bp for recombinant pET15b-VL8_*msr*A plasmid, 670 bp for pET15b-VL8_*msr*B plasmid.

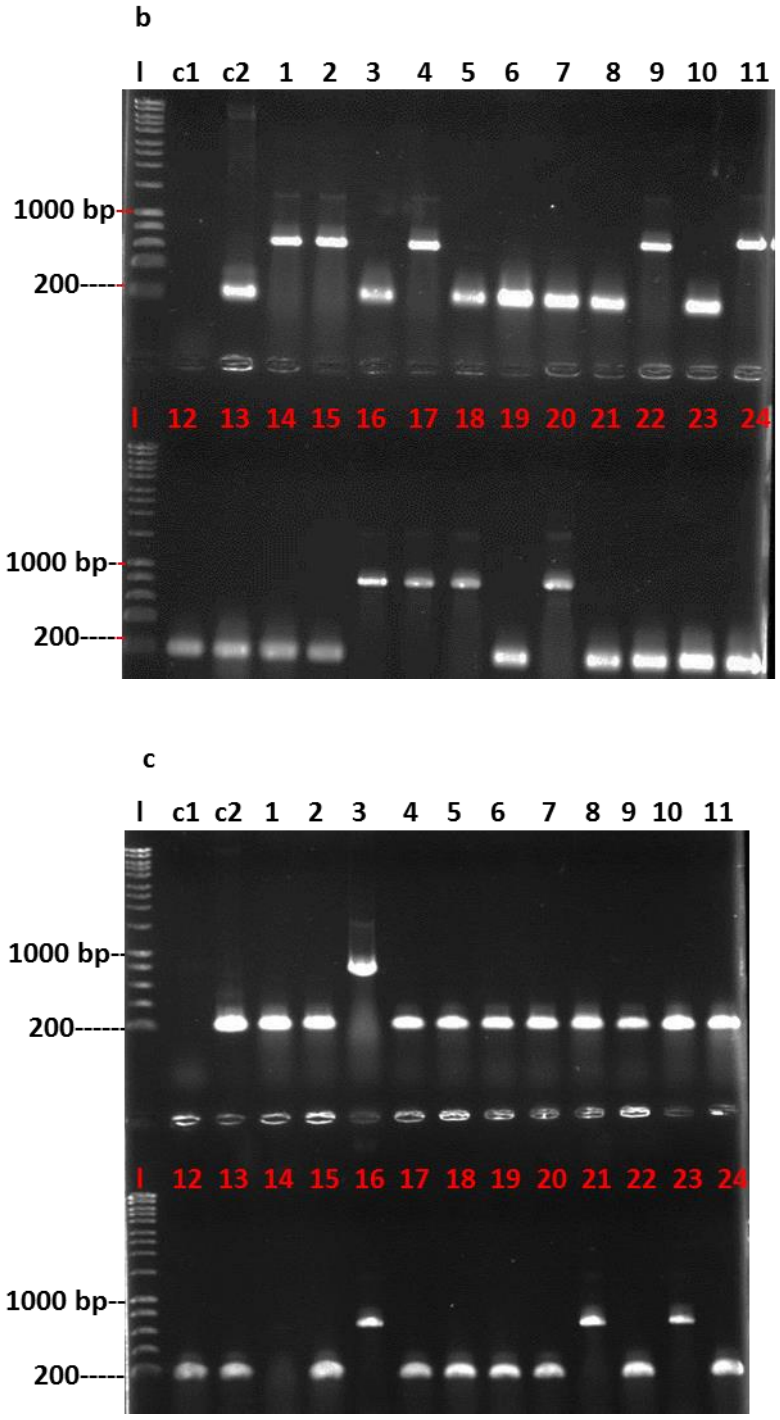


Figure 4.6-Continued

4.4.4 Protein Expression and Purification of Methione Sulphoxide Reductases from *L. agilis* R16, *E. coli* VL8 and *E. coli* FC44

The protein expression inductions in *E. coli* BL21(DE3) expressing pET15b-LBAG_*msr*A1, pET15b-LBAG_*msr*A2 and pET15b-LBAG_*msr*A3 were carried out for 4 h at 37°C or 24 h at 25°C, shaking at 250 rpm. Protein expression induction of MsrA proteins of *L. agilis* R16 at 37°C did not produce sufficient recombinant proteins in the soluble fraction. Recombinant proteins were seen in insoluble fractions (Lane 11, 13, 14, Figure 4.7) so different induction and extraction conditions were tested (Figure 4.8) for MsrA1 protein expression.

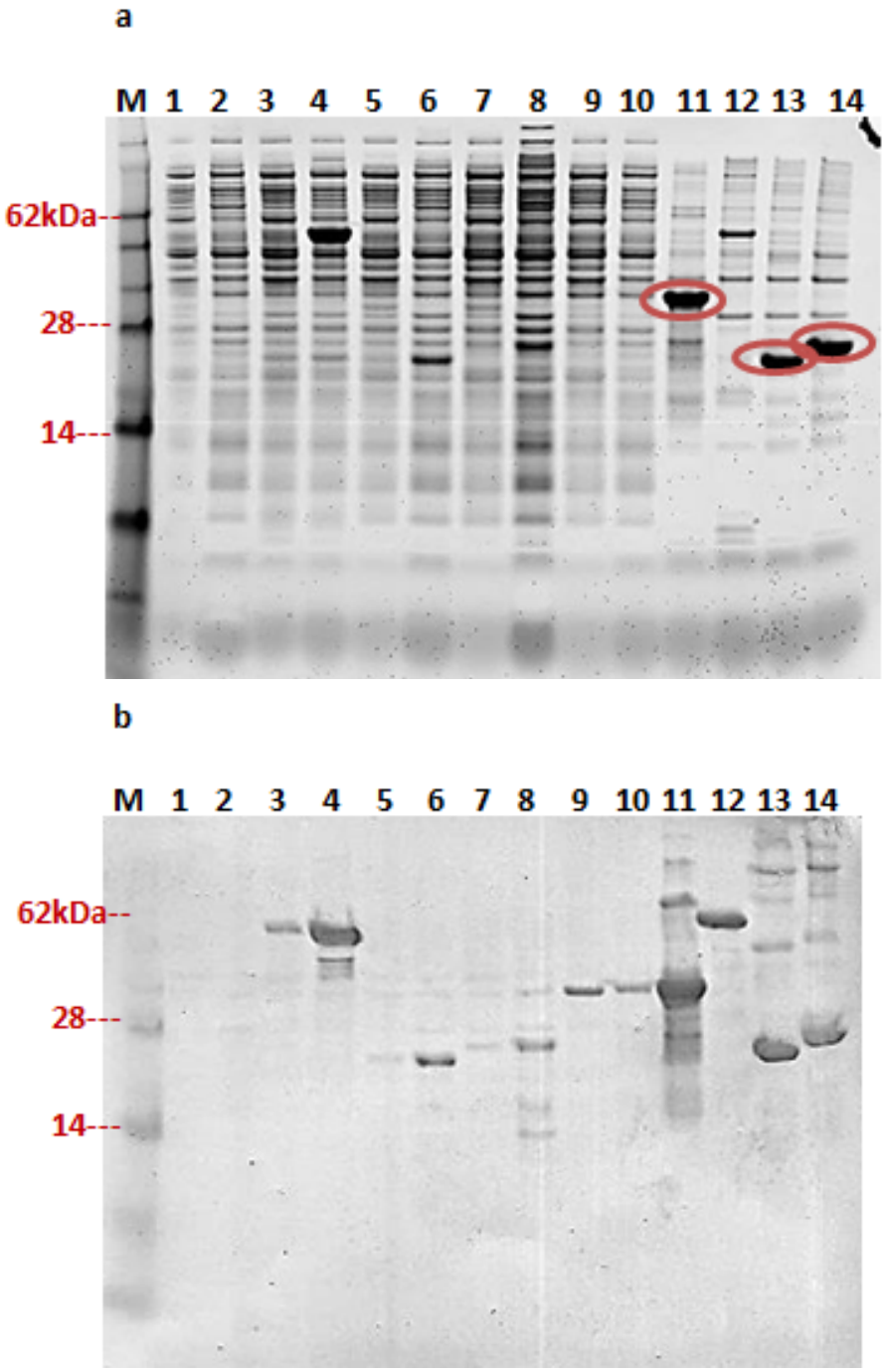


Figure 4.7 SDS-PAGE and Western Blot of the CFEs obtained from *E. coli* BL21 (DE3) cells expressing pET15b-LBAG_*msr*A1, pET15b-LBAG_*msr*A2 and pET15b-LBAG_*msr*A3. a) Protein gel where CFEs were run on 12% Bis-Tris gel under denaturing, reducing conditions in MES buffer. b) Western blot (using anti-His tag antibody) within the same order of protein gel in a. 10 µg of protein was loaded per lane. M; See Blue protein marker. Soluble CFEs of uninduced and 4 h induced *E. coli* BL21 (DE3) cells expressing: 1-2; pET15b (empty vector control), 3-4; pET15b-*ecg*44 (positive control for western-6XHis-tagged protein), 5-6; pET15b-LBAG_*msr*A1, 7-8 pET15b-LBAG_*msr*A2, 9-10; pET15b-LBAG_*msr*A3, Insoluble CFEs of 4 h induced *E. coli* BL21 (DE3) cells expressing: 11; pET15b-LBAG_*msr*A3, 12; pET15b-*ecg*44, 13; pET15b-LBAG_*msr*A1, 14; pET15b-LBAG_*msr*A2. Protein expressions were induced at 37°C with 0.5 mM IPTG.

Protein expression induction at 25°C for 24 h resulted in higher level of expression of MsrA1 (Lane 5, 6 Figure 4.8). The same profile was seen for MsrA2. After optimisation of induction conditions, MsrA1 and MsrA2 proteins were successfully extracted and confirmed to have predicted sizes (Table 4.6). These proteins were checked by Western blot for presence of 6XHis-tag (Figure 4.7) and purified by Ni-NTA column. The purification of MsrA3 was more challenging, the targeted protein was not expressed well by the *E. coli* cells and the Ni-NTA purification of MsrA3 failed (Figure 4.9). Many different protein expression conditions were tested to optimise the conditions. MsrA3 is mostly found in the insoluble fraction. Reducing the IPTG concentration to 0.1 mM and extraction by Bugbuster protein extraction reagent (Novagen) improved the extraction of MsrA3 in the soluble fraction (Figure 4.9 BB Lanes, circled) but MsrA3 was not used in further enzyme assays.

Bacterium	Protein	Predicted Size with 6XHis-tag (kDa)	pI
L. agilis R16	MsrA1	21.4	6.2
L. agilis R16	MsrA2	22	6.2
L. agilis R16	MsrA3	33.4	10
L. agilis R16	MsrB	18	6.3
L. agilis R16	fRMsr	18.3	5.0
E. coli FC44	fRMsr	22	5.8
E. coli VL8	MsrA	25.1	6.0
E. coli VL8	MsrB	17.2	6.5

Table 4.6 Properties of putative methionine sulphoxide reductases. pI; isolelectric point. Predicted sizes and pI of proteins were obtained using the amino acid sequences of the proteins by Editseq-sequence editing software (DNAstar).

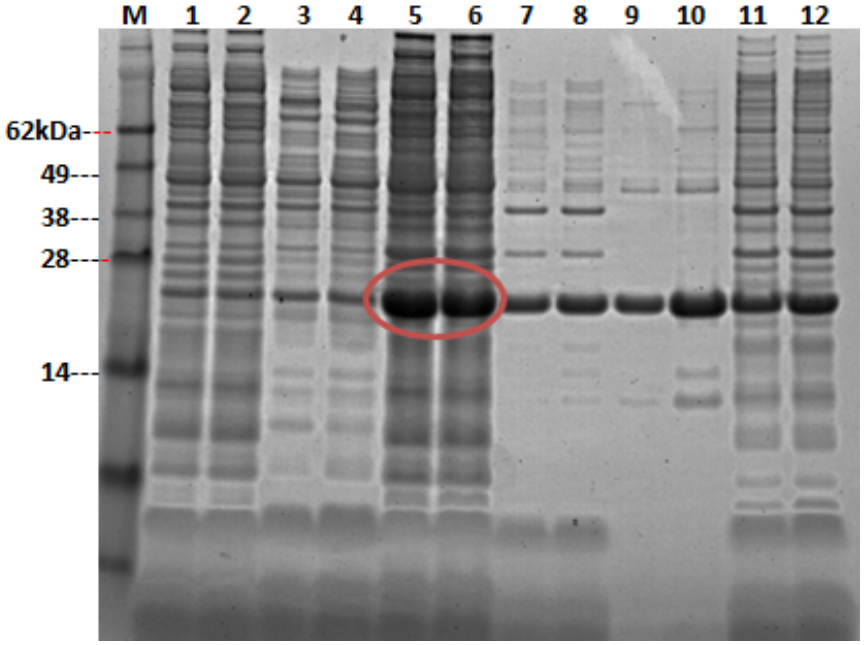


Figure 4.8 SDS-PAGE of CFEs of *E. coli* BL21 (DE3) expressing pET15b-LBAG_*msr*A1. 10 µg of protein was loaded per lane. M: See Blue Protein Marker. Soluble CFEs obtained from induced cells for: 1-2; 3 h at 37°C, 3-4; 4 h at 37°C extracted by Bugbuster, 5-6: 24 h at 25°C. Insoluble CFEs obtained from induced cells for: 7-8; 3 h at 37°C, 9-10: 4 h at 37°C extracted by Bugbuster, 11-12: 24 h at 25°C. Proteins were run on 12% Bis-Tris Gel under denaturing, reducing conditions in MES buffer.

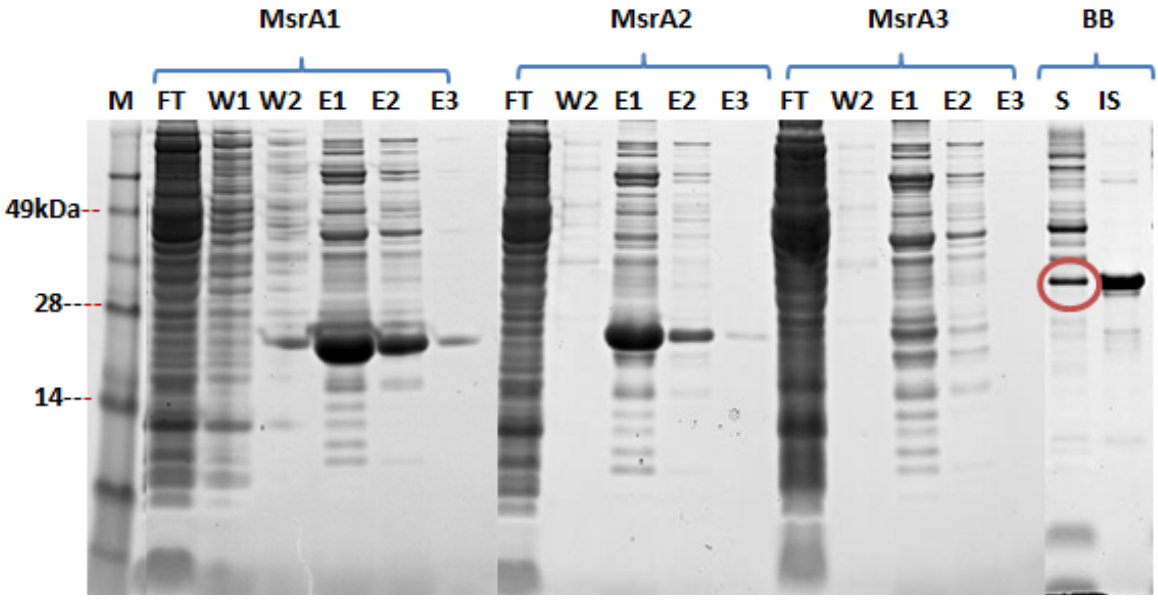


Figure 4.9 SDS-PAGE of the Ni-NTA purified MsrA proteins from *E. coli* BL21 (DE3) expressing pET15b-LBAG_*msr*A1, pET15b-LBAG_*msr*A2 and pET15b-LBAG_*msr*A3. 3.25 µl of fractions were loaded per lane. FT: Flow through, W: Wash, E: Eluate, M; Protein marker, S: Soluble, IS: Insoluble. BB: MsrA3 purified by Bugbuster (Novagen). Proteins were run on 10% Bis-Tris Gels under denaturing, reducing conditions in MES buffer.

To optimise the protein expression for the rest of the Msr proteins, various conditions were examined. The protein expressions were carried out at 25°C for 4 h and 37°C for 3 h or 4 h. The proteins were induced and CFEs were loaded on a SDS-PAGE gel and checked for their expected sizes (See Table 4.6). The optimum protein expression conditions for all of the proteins are shown in Table 4.7. Induction at 37°C for 4 h was found to be optimum under the conditions tested for LBAG_MsrB and VL8_MsrB. For VL8_MsrA expression, 3 h induction at 37°C was found to provide higher levels of protein. These proteins showed a different expression profile compared to MrsA1 and MsrA2 proteins of *L. agilis* R16. Protein expression induction at 25°C resulted in better expression of MrsA1 and MsrA2 proteins of *L. agilis* R16. The protein expression of LBAG_fRMsr and FC44_fRMsr under conditions tested did not produce sufficient proteins in the soluble fraction and proteins were mostly seen in the insoluble fraction.

Bacterium	**Protein**	**Abbreviation**	**Optimum Protein Expression Condition**
L. agilis R16	MsrA1	LBAG_MsrA1	25°C for 24 h
L. agilis R16	MsrA2	LBAG_MsrA2	25°C for 24 h
L. agilis R16	MsrA3	LBAG_MsrA3	37°C for 3 h*
L. agilis R16	MsrB	LBAG_MsrB	37°C for 4 h
L. agilis R16	fRMsr	LBAG_fRMsr	ND
E. coli FC44	fRMsr	FC44_fRMsr	ND
E. coli VL8	MsrA	VL8_MsrA	37°C for 3 h
E. coli VL8	MsrB	VL8_MsrB	37°C for 4 h

Table 4.7 Optimum conditions of *E. coli* BL21 (DE3) expressing *msr* genes. The abbreviations of the proteins are shown and these will be used subsequently for naming the proteins. * Protein extraction was performed by Bugbuster Reagent (Novagen). ND; Not determined, expression of LBAG_fRMsr and FC44_fRMsr under conditions tested resulted in expression of the proteins mostly in insoluble fractions.

CFEs were prepared using optimum protein expression conditions and tested for the presence of 6XHis-tag by Western Blot (See Figure 4.10). Protein expression induction was performed at 37°C for 4 h for LBAG_fRMsr and FC44_fRMsr.

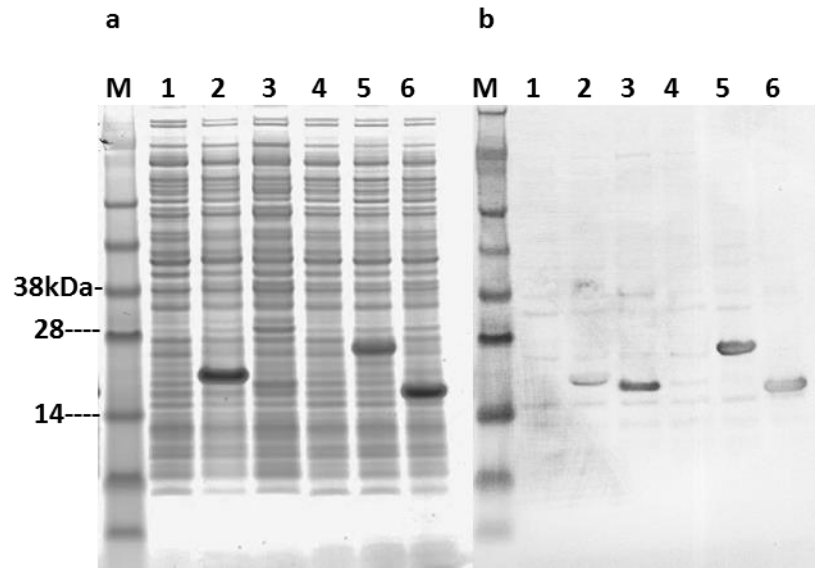


Figure 4.10 SDS-PAGE and Western Blot of the soluble CFEs from *E. coli* BL21 (DE3) expressing recombinant pET15b plasmids expressing *msr* genes. a) Protein gel where CFEs were run on 12% Bis-Tris gel under denaturing, reducing conditions in MES buffer. b) Western blot (using anti-His tag antibody) within the same order of protein gel in a. 10 µg of protein was loaded per lane. M: See Blue Protein Marker. CFEs of *E. coli* BL21 (DE3) cells expressing: 1; pET15b (empty vector control) induced at 37°C for 4 h, 2; pET15b-LBAG_*msr*B induced at 37°C for 4 h, 3; pET15b-LBAG_*frmsr* induced at 37°C for 4 h, 4; pET15b-FC44_*frmsr* induced at 37°C for 4 h, 5; pET15b-VL8_*msr*A induced at 37°C for 3 h, 6; pET15b-VL8_*msr*B induced at 37°C for 4 h.

Proteins were seen to be bound to the anti-His tag except the FC44_fRMsr (Lane 4, Figure 4.10) with the expected sizes (Table 4.6). No binding to the anti-His tag was observed for FC44_fRMsr. LBAG_MsrB and VL8_MsrB were purified by Ni-NTA columns. Ni-NTA purification was performed successfully; eluates were obtained with expected protein sizes (Figure 4.11).

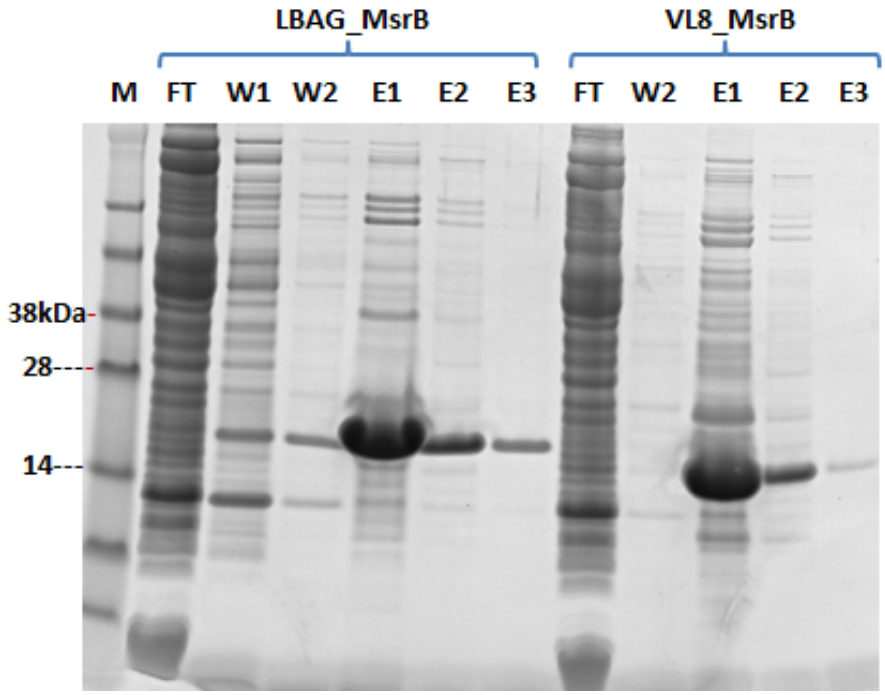


Figure 4.11 SDS-PAGE of Ni-NTA purified MsrB proteins from *E. coli* BL21 (DE3) expressing pET15b-LBAG_*msr*B and pET15b-VL8_*msr*B. 3.25 µl of fractions were loaded per lane. M: See Blue Protein Marker, FT: Flow through, W: Wash, E: Eluate, M; Protein marker, Proteins were run on 10% Bis-Tris Gels under denaturing, reducing conditions in MES buffer. The predicted sizes of LBAG_MsrB and VL8_MsrB are 18 kDa and 17.2 kDa, respectively.

4.4.5 Reductase Activity of Recombinant Methionine Sulphoxide Reductases

Ni-NTA purified, dialysed eluates of MsrA1 and MsrA2 were used to determine glucoraphanin reduction in the presence of dithiothreitol (DTT). Samples were analysed by HPLC to check glucoraphanin degradation and formation of glucoerucin after 30 min incubation. MsrA1 was more effective for glucoerucin formation than MsrA2 (Figure 4.12). MsrA1 and MsrA2 showed glucoerucin formation at a rate of 0.57 ± 0.04 nmol glucoerucin/µg protein and 0.13 ± 0.00 nmol glucoerucin/µg protein respectively. MsrA1 was also tested for reductase activity with longer incubation times (1, 2 and 3 h) and higher protein amount (50 µg). None of these changes in the enzyme assays increased the rate of glucoerucin formation for MsrA1 (data not shown). This reduction did not occur in the control samples without eluates. This suggests that the reduction is an enzymatic reaction. When 1 mM of NADPH was used as reducing agent instead of 20 mM DTT with 1 mM $MgCl_2$ in the assays, no glucoerucin formation was obtained. This study was continued to focus on other bacteria such as *E. coli* VL8 and *E. coli* FC44. In addition, the role of other types of Msr enzymes was investigated.

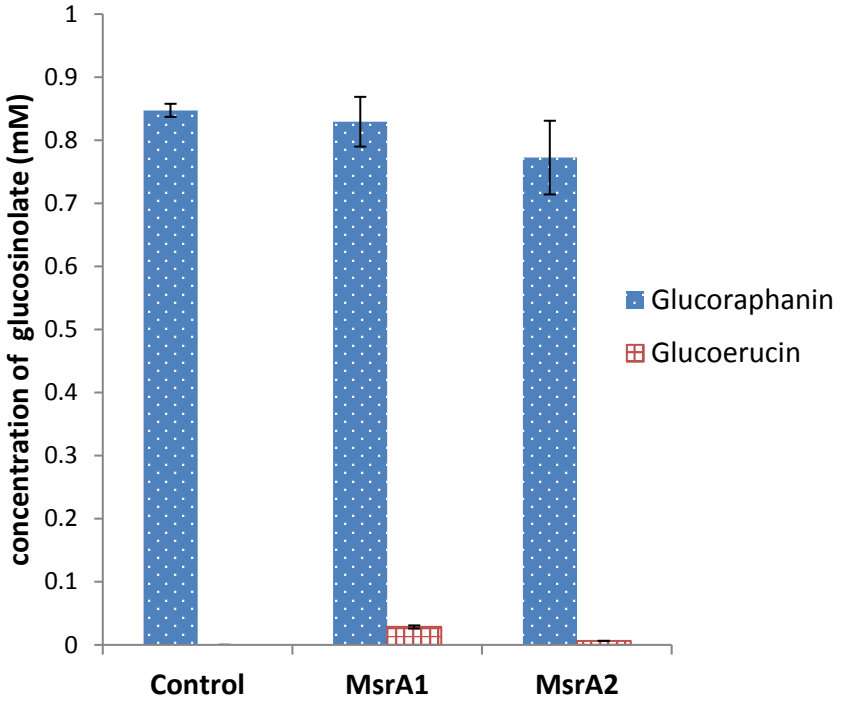


Figure 4.12 Reduction of glucoraphanin to glucoerucin by MsrA1 and MsrA2. Glucoraphanin amount left and glucoerucin amount formed after 30 min incubation are shown. Samples and controls contained 1 mM glucoraphanin originally. Control indicates the glucoraphanin control without any protein that was incubated for 30 min. Assays were set up using same amount of protein (10 µg). Error bars = standard deviation (n=3).

CFEs were tested for reductase activity against 0.5 mM glucoraphanin to examine if they can reduce glucoraphanin to glucoerucin. CFEs of *E. coli* BL21 (DE3) expressing pET15b-LBAG_*msr*B, pET15b-LBAG_*frmsr*, pET15b-FC44_*frmsr, pET15b*-VL8_*msrA* and pET15b-VL8_*msr*B were tested for reductase activity using 20 mM DTT as reducing agent. The results are shown in Figure 4.13. In the presence of DTT without CFE (control 1), glucoraphanin was converted to glucoerucin with a rate of 0.08 ± 0.03 nmol glucoerucin/µg protein. In absence of DTT and any CFE (control 2), there was no glucoerucin formation but glucoraphanin degradation occurred. In addition, control including CFE of *E. coli* BL21(DE3) pET15b (empty vector control) showed glucoraphanin degradation but there was no glucoerucin formation in these controls. LBAG_fRMsr, IS44_ fRMsr and VL8_MsrA samples showed glucoerucin formation at a similar rate (~0.06 ± 0.01 nmol glucoerucin/µg protein) which was also similar to control 1. LBAG_MsrB and VL8_MsrB reduced glucoraphanin at a rate of 0.31 ± 0.02 nmol glucoerucin/µg protein and 0.68 ± 0.02 nmol glucoerucin/µg protein, respectively.

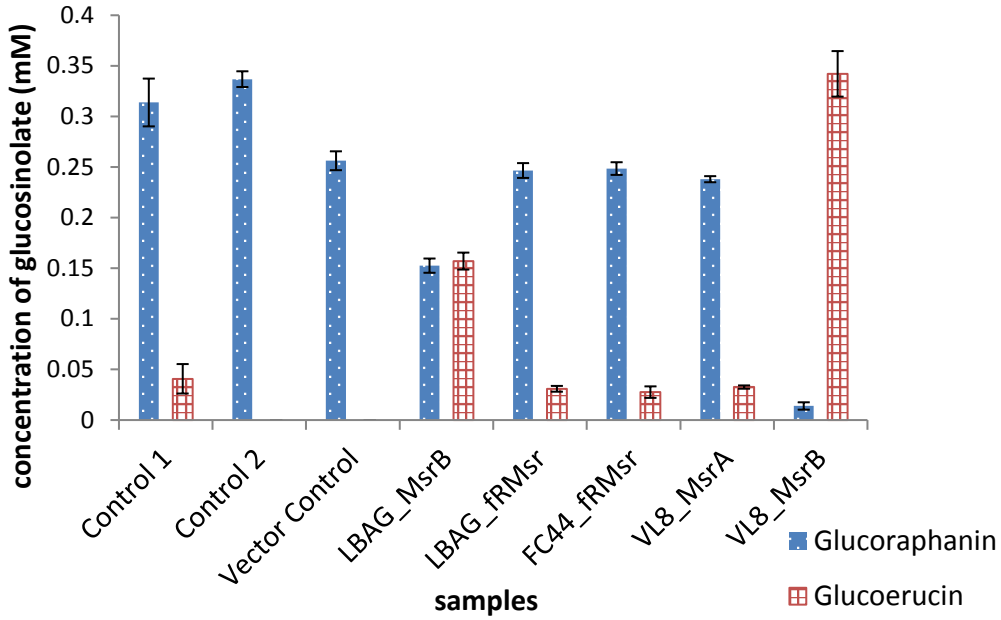
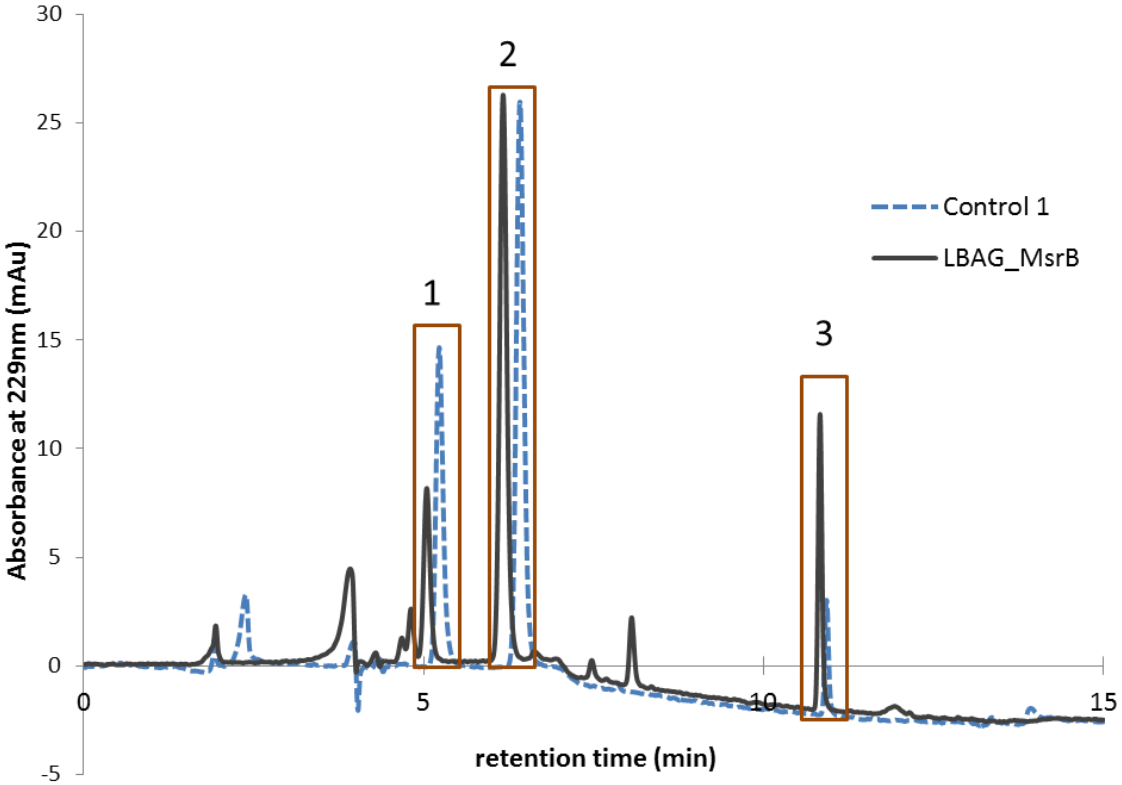


Figure 4.13 Reduction of glucoraphanin to glucoerucin by CFEs of *E. coli* BL21(DE3) expressing *msr* genes. Control 1 indicates the glucoraphanin control without any CFE, Control 2 indicates the control without CFE and DTT addition. Vector Control is the sample that includes CFE of *E. coli* BL21(DE3) with empty pET15b vector. Samples and controls contained 0.5 mM glucoraphanin originally. Assays were set up using same amount of protein (100 µg). Error bars = standard deviation (n=3).

MsrB proteins are likely to be responsible for the reductase activity of *E. coli* VL8 and *L. agilis* R16. LBAG_MsrB and VL8_MsrB showed higher reduction rates compared to control 1 (See chromatograms, Figure 4.14). Notably, the reductase activity of VL8_MsrB was remarkable because it was able to reduce almost all of the glucoraphanin to glucoerucin.



b

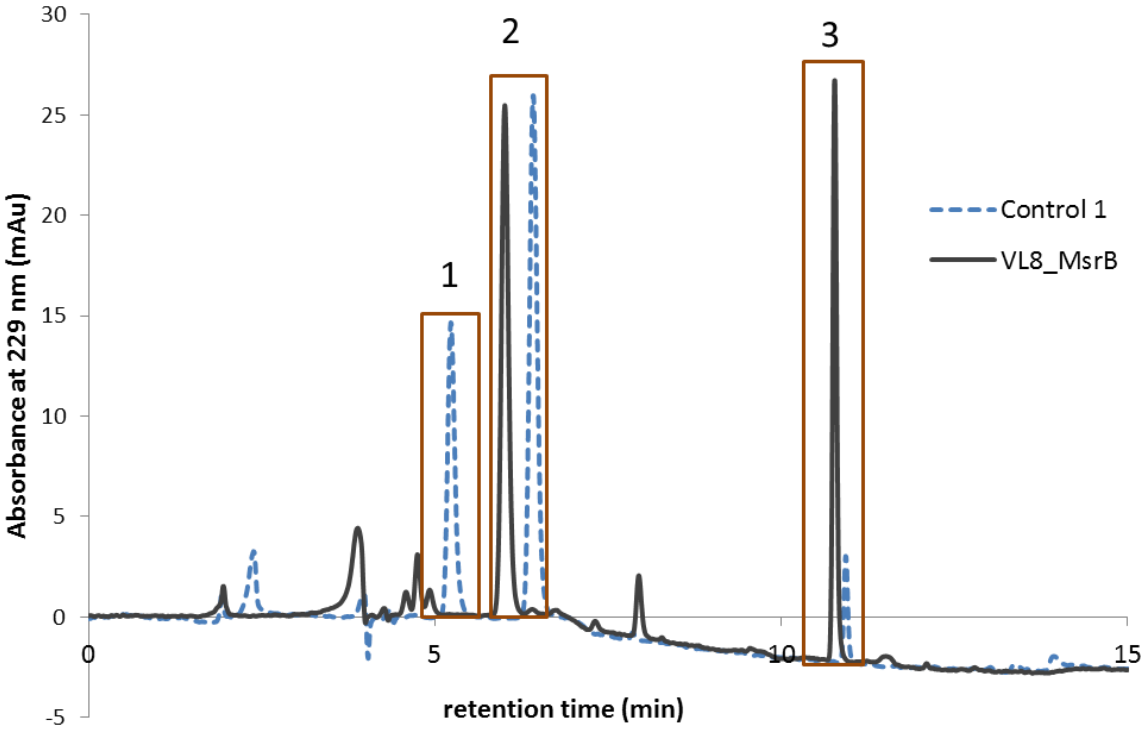


Figure 4.14 HPLC chromatograms showing reduction of glucoraphanin to glucoerucin by CFEs of *E. coli* BL21(DE3) cells expressing pET15b-LBAG_*msr*B (a) and pET15b-VL8_*msr*B (b). 1; glucoraphanin, 2; sinigrin (internal standard), 3; glucoerucin.

Ni-NTA purified, dialysed eluates of LBAG_MsrB and VL8_MsrB were used to determine enzyme activity in the presence of DTT or NADPH as reducing agent. These reactions were used in the sample preparation for HPLC and analysed on HPLC to check the formation of glucoerucin. The reduction rates were based on glucoerucin formation and are given in Table 4.8. In addition, the concentrations of glucoraphanin left and glucoerucin formed is shown in Figure 4.15. Controls including NADPH or DTT showed a reduction rate of 0.13 ± 0.01 nmol glucoerucin/µg protein and 0.51 nmol ± 0.02 nmol/glucoerucin/µg protein respectively. Reduction rate of LBAG_MsrB was 0.14 ± 0.01 nmol glucoerucin/µg protein and 1.03 ± 0.11 nmol glucoerucin/µg protein using NADPH and DTT as reducing agent respectively. VL8_MsrB showed a reduction rate of 0.15 ± 0.01 nmol glucoerucin/µg protein when NADPH was used as reducing agent. However, it reduced glucoraphanin to glucoerucin with a rate of 5.42 ± 0.18 nmol glucoerucin/µg protein when DTT was used instead of NADPH.

	Reduction rate (nmol glucoerucin/µg protein)	
Samples	NADPH	DTT
Control	0.13 ± 0.01	0.51 ± 0.01
LBAG_MsrB	0.14 ± 0.01	1.03 ± 0.11
VL8_MsrB	0.15 ± 0.01	5.42 ± 0.18

Table 4.8 The reduction rate of glucoraphanin to glucoerucin by Ni-NTA purified LBAG_MsrB and VL8_MsrB using NADPH or DTT as reducing agent. The reactions were set up with 1 mM NADPH or 20 mM DTT as reducing agent. Assays were set up using same amount of protein (10 µg). Results were given as mean of triplicates ± standard deviation.

VL8_MsrB could reduce glucoraphanin to glucoerucin with a high rate in presence of DTT as reducing agent. The reduction by VL8_MsrB was very low when NADPH used as a reducing agent (Figure 4.15).

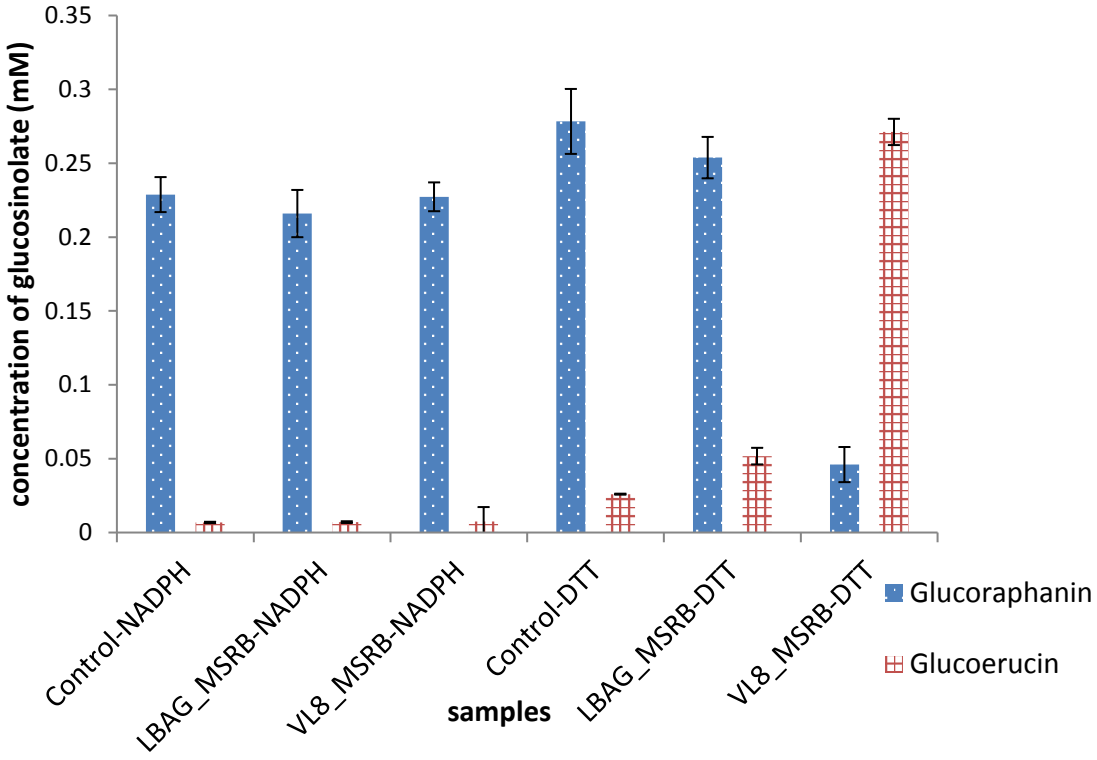


Figure 4.15 Reduction of glucoraphanin by Ni-NTA purified LBAG_MsrB and VL8_MsrB. Control indicates the glucoraphanin control without any eluate. The reactions were set up with 1 mM NADPH or 20 mM DTT as reducing agent. Samples and controls contained 0.5 mM glucoraphanin originally. Assays were set up using same amount of protein (10 µg). Error bars = standard deviation (n=3).

4.5 DISCUSSION

Significant inter-individual variations have been observed in the degradation of glucosinolates to ITCs by human gut bacteria [246]. The myrosinase activity of gut bacteria against different glucosinolates may vary and gut bacteria may use alternative pathways to degrade glucosinolates. For instance, some gut isolates were able to degrade methylsulfinylalkyl glucosinolates (such as glucoraphanin) directly to its corresponding isothiocyanate (sulforaphane for glucoraphanin) while others require a reduction step that converts methylsulfinylalkyl glucosinolates into methylthioalkyl glucosinolates followed by hydrolysis to isothiocyanates [55, 56]. Therefore this reduction step is essential in understanding the diverse routes of glucosinolate metabolism to maximize the health promoting effects of ITCs.

E. coli VL8 and *E. coli* FC44 converted the glucoraphanin in the media to its reduced counterpart, glucoerucin so these strains were further studied for their reductase activity. CFEs of *E. coli* VL8 and *E. coli* FC44 prepared from cells grown overnight with or without glucosinolate supplementation were able to reduce glucoraphanin to glucoerucin. Luang-In et al. (2014) reported that reductase activity in CFE of *E. coli* VL8 was inducible by glucoraphanin and CFE of *E. coli* VL8 was unable to reduce glucoraphanin without a glucosinolate supplementation to the media used in CFE preparation [56]. However, no induction effect was found for reductase activity in CFEs of *E. coli* VL8 and *E. coli* FC44. The same culture (nutrient broth) to grow cells and same buffer (0.1 M citrate phosphate buffer pH 7.0) to prepare CFE were used in our study. The differences were the method to prepare CFEs and the source of the substrate (glucoraphanin). Instead of using a cell-disrupter, sonication was used in our study. Glucoraphanin used in our study was either supplied to us from a human intervention study in IFR (purchased from Intertek) or purchased from a company (Cambridge Biosciences). However, Luang-In et al. (2014) have prepared their own glucoraphanin by oxidation of glucoerucin with hydrogen peroxide [56]. There was no comment about ITC or other degradation products in their study. If glucoraphanin was reduced to glucoerucin in uninduced CFEs, glucoerucin might have been degraded by myrosinase activity rapidly and not detected in their samples. Another possibility is that the glucoraphanin used in our study might be more amenable to reductase activity. Our results suggested that reductase activity in CFEs of *E. coli* VL8 and *E. coli* FC44 may be constitutive.

Methionine sulphoxide reductase A (MsrA) was suggested to be responsible for reducing methylsulfinylalkyl glucosinolates (such as glucoraphanin) into methylthioalkyl glucosinolates (such as glucoerucin) [55]. To test this hypothesis, partially purified MsrAs were tested for reductase activity against glucoraphanin (Intertek) in the presence of different reducing agents but the degradation rate of glucoraphanin to glucoerucin was low. Increasing incubation time and protein amount was also tested but these effects did not improve the reductase activity of MsrAs. This low glucoraphanin reduction rate was in accordance with the low glucoraphanin degradation rate by *L. agilis* R16 reported previously because *L. agilis* R16 degraded only 10 ± 2% of the glucoraphanin in the media after 24 h [56].

It was also thought that different types of Msr enzymes can have a role in reduction of glucosinolates because they can reduce different epimers of methionine sulphoxide. For instance, MsrAs can only reduce free and protein-based *S*-epimers of methionine sulphoxide.

In the literature, glucoraphanin extracted from *B. oleracea* var *italica* (broccoli) and *A. thaliana* was reported to be a pure epimer of *R* configuration at the sulphoxide group [247]. No information was provided to us by suppliers about configuration of the methyl sulphoxide of glucoraphanin, they might be a pure epimer (*R* or *S*) or a mix of these two epimers. Therefore, different Msr enzymes were investigated in our study including MsrA, MsrB and fRMsr of *L. agilis* R16, *E. coli* VL8 and *E. coli* FC44.

FC44_fRMsr failed to bind the anti-His tag during Western Blot. This might be a problem about protein production or may suggest that there had been protein degradation. Lin et al. (2007) studied the fRMsr of *E. coli* sharing the same amino acid with fRMsr of *E. coli* FC44. They reported the initiation of the protein synthesis 18 amino acids upstream from the first methionine and presence of larger version of fRMsr. Moreover, it was reported that the purified protein was a mix of both versions of the fRMsr showing the same Msr activity [245]. There was no indication of signal peptide. Maybe, the full length of protein has better stability or solubility so it was needed to clone the larger version of the protein for a successful protein expression with 6XHis-Tag.

When CFEs of *E. coli* BL21(DE3) expressing *msr* genes were tested for reductase activity, it was found that the sum of the glucoraphanin amount left and glucoerucin amount formed did not match with the original glucoraphanin concentration. This effect was seen in controls without any reducing agent and protein. The reason might be the purity of glucoraphanin stock used in the study. It should be noted that glucoraphanin stock (from Cambridge Biosciences) used in this part of the study was different from the one used for testing reductase activity of LBAG_MsrA1 and LBAG_MsrA2 (from Intertek). Notably, vector control which included CFE of *E. coli* BL21(DE3) cells expressing empty pET15b vector had a low glucoraphanin amount left compared to controls without any glucoerucin formation. It is not known if this effect might be caused by the metabolism of glucoraphanin by CFE. Another issue is the formation of glucoerucin in controls including only DTT without any protein extract. This might suggest a non enzymatic reduction and the longer incubation time might be the reason for this effect. When purified LBAG_MsrA1 and LBAG_MsrA2 were tested for reductase activity, samples and controls were incubated for 30 min and there was no glucoerucin formation in controls. To overcome this, the effect of shorter incubation times could be tested.

As expression of FC44_ fRMsr with 6XHis-tag failed, it is not possible to comment on its role in reductase activity of *E. coli* FC44. However, it is unlikely that LBAG_fRMsr and VL8_MsrA are responsible for reduction of glucoerucin because CFEs consisting these proteins showed similar reduction rates compared to control. However, shortening the incubation time might eliminate the reduction of glucoraphanin by control and give a better idea about the roles of these proteins in reductase activity.

Our results indicate that MsrB may be responsible for reduction glucoraphanin to glucoerucin because CFEs consisting LBAG_MsrB and VL8_MsrB showed remarkable reduction of glucoraphanin to glucoerucin compared to control. Luang-In et al. (2014) reported that reductase activity in CFEs of *E. coli* VL8 was oxygen-independent, NADPH and Mg^{+2} dependent [56]. When Ni-NTA purified LBAG_MsrB and VL8_MsrB were tested for reductase activity in presence of 1 mM NADPH with 1 mM $MgCl_2$, reduction rates were low and similar to control. Using 20 mM DTT as a reducing agent resulted in higher reduction rates and showed that DTT can be used as a reducing agent to test reductase activity *in vitro*. The reduced thioredoxin was reported to be the reducing agent *in vivo* in reduction of methionine sulphoxides and DTT was shown to substitute it *in vitro* [141, 163]. It also suggested that MsrB may be the responsible enzyme for reduction of glucoraphanin to glucoerucin in *L. agilis* R16 and *E. coli* VL8. The role of MsrB in reductase activity suggested that the glucoraphanin stock used was mainly included *R*-epimer of sulphoxide group as has been suggested to occur in plants [247]. It should be taken into account that MsrA has specificity for the *S*-epimer, MsrB and fRMsr have specificities for *R*-epimer of methionine sulphoxide and it is important to use a substrate that mimics the substrate found in the plant to determine which enzyme is responsible for activity. When reductase activity of purified MsrA1 and MsrA2 of *L. agilis* R16 was tested, a different glucoraphanin stock was used. If this previous glucoraphanin stock contained small amount of *S*-epimer of sulphoxide group, this might be the reason for the low levels of MsrA activity. It is not known why LBAG_fRMsr could not reduce glucoraphanin as fRMsr has specificity for free form of *R* epimer of methyl sulphoxide but this could be related to low levels of protein production in the soluble CFE.

When reductase activity of LBAG_MsrB and VL8_MsrB were compared to each other, it explained the difference in glucoraphanin degradation rates of *L. agilis* R16 and *E. coli* VL8. As reported previously, *L. agilis* R16 was degraded 10 ± 2% of glucoraphanin in the media but *E. coli* VL8 was able to degrade 91 ± 7% of the glucoraphanin in the media after 24 h incubation

[56]. This showed that *E. coli* VL8 has more efficient reductase activity compared to *L. agilis* R16, resulting in higher glucoraphanin degradation rates.

Unfortunately, we could not determine the optimum conditions for reductase activity of LBAG_MsrB and VL8_MsrB. The effect of temperature, buffer, pH, other reducing agents, etc. should be investigated. As reductase activity in CFEs of *E. coli* VL8 was reported to be oxygen-independent, all assays were performed under aerobic conditions. However, the effect of oxygen can be still tested for confirmation. Sulforaphane was a substrate for reductase activity of *E. coli* VL8 CFE [137], it was expected that MsrB can also reduce sulforaphane and further examination of the substrate range will be interesting.

In summary, it was found that MsrB is likely to be the responsible enzyme in reduction of glucoraphanin possessing *R*-epimer of sulphoxide in its structure. This study was the first to confirm the role of Msr enzymes in reduction of glucosinolates with a sulphoxide group in their side chain. This finding provides us a better understanding about the diverse routes of glucosinolate metabolism by gut bacteria.

CHAPTER FIVE

5 CHARACTERISATION OF A BACTERIAL MYROSINASE OF SOIL ORIGIN

5.1 INTRODUCTION

The first bacterial myrosinase was purified in 1974 from *Enterobacter cloacae* but it was not extended to further identification of the gene or amino acid sequence [231]. Since then a limited number of studies have investigated the metabolism of glucosinolates by different bacteria. These studies attempted to identify and characterise the enzyme but with limited success [137, 232] Some studies reported the glucosinolate metabolism by human gut bacteria *in vitro* [56], in rat models [112, 213] or using in vitro fermentation models [165, 212]. These studies contributed to our understanding of the glucosinolate metabolism by human gut bacteria. However, identification of the responsible enzyme or enzyme groups and their characterisation could not be achieved. To maximize the health promoting effects of glucosinolate degradation products, isothiocyanates, we need a better understanding of bacterial myrosinases. In this context, identification and characterisation of bacterial myrosinases is crucial.

Recently, a *C.* WYE1 strain was isolated from soil and this strain was reported to have glucosinolate degrading ability. An enrichment method using the sinigrin as sole carbon source was used to isolate this bacterium. The enzyme responsible for myrosinase activity was purified by a chromatography approach that combined ion exchange chromatography and gel filtration. Later, N-terminal and internal amino acid sequence of the myrosinase was determined and enzyme activity partially characterised using cell-free extracts. It was found to be a periplasmic β-glucosidase from GH3 family with a signal peptide [118]. So far, identified myrosinases in plants [119], aphids [132] or beetles [248] were reported to belong to the GH1 family so this myrosinase was the first of its own kind. It was found be activated by ascorbate like plant myrosinases [118]. The identification of this myrosinase changed our perspective in this study and we then focused on β-glucosidases from GH3 too to identify a bacterial myrosinase from the human gut. Rossiter Group at Imperial College attempted to clone and express this myrosinase gene in *E. coli* (unpublished data). Unfortunately, their attempt to clone the myrosinase gene was not successful and they did not pursue this further. This chapter focuses on cloning and expression of *C.* WYE1 myrosinase (*cmyr*) and characterisation of this enzyme.

5.2 HYPOTHESIS

C. WYE1 showed myrosinase activity. The reported myrosinase activity of *C.* WYE1 is the product of the expression of a periplasmic β-glucosidase from GH3 family.

5.3 MATERIALS AND METHODS

5.3.1 Strain Identification and Preparation of Phylogenetic Tree

The strain *C.* WYE1 was kindly provided to by John Rossiter from Imperial College London. The genome of *C.* WYE1 was sequenced, assembled and annotated as described in Chapter 2-section 2.2.13. 16S rDNA sequencing was performed to assign the strain to a specific species, method was performed as described in Chapter 2- section 2.2.12. To determine the phylogeny of *C.* WYE1, a phylogenetic tree was constructed as described in Chapter 2-section 2.2.14.

5.3.2 Comparison of *C.* WYE1 Myrosinase with Other Myrosinases

Amino acid sequences of different myrosinases from different origins were selected using the Uniprot database. These myrosinases were aligned with cmyr using Clustal Omega [237].

5.3.3 Cloning and Expression of Myrosinase Gene

As cmyr was reported to have a signal peptide, a cloning vector which has 6XHis-tag site in both N-terminal and C-terminal site (pET28b) was used. Primers were designed to introduce new restriction sites by changing nucleotides and to enable in-frame translation of C-terminal 6XHis-tag by insertion of a new nucleotide to the reverse primer. This allowed the expression of the protein with two 6XHis-tags at N-terminal and C-terminal sites. Then it was possible to purify protein by Ni-NTA column even if the N-terminal tag is being cleaved off. The primers were designed to amplify the *cmyr* are given in Table 5.1, the restriction sites are underlined. The nucleotide sequence of the cloned gene is shown in the Appendix 2. PCR insert was amplified (PCR conditions are given in Table 5.2) and double restricted by NdeI/XhoI and ligated to NdeI/XhoI double restricted pET28b vector (Novagen) then this gene construct was transformed into *E. coli* DH5α cells as described in Chapter 2-section 2.2.10 and the cells were plated on L agar with kanamycin (30 µg/ml, final concentration).

Name	Primer Sequence (5``-3`)	Restriction Site
CMYR-F	GGAA**CA**TATGCTCACTGCTTTTAAGA	NdeI
CMYR-R	CAGA**CTCG**AG**A**CGTGTCAGTCCGAAT	XhoI

Table 5.1 Primers used in the PCR experiments to amplify the *cmyr*. The restriction sites are underlined. The altered nucleotides to introduce the restriction sites are in bold. The inserted nucleotide is highlighted.

Name	T_A, time, cycles	T_E, time	Product size
CMYR-F	54°C, 30 s, 5X	72°C, 1 min	1998 bp
CMYR-R	68°C, 30 s, 20X		

Table 5.2 Conditions used in the PCR experiments to amplify the *cmyr* gene. T_A; annealing temperature, cycles; cycles used in annealing step of PCR, T_E; extension temperature and time used in PCR, Product size; the expected PCR product sizes for β-glucosidase genes. The annealing step was performed using T_m of 100% sequence match part of primers first (5X) and then using the T_m of entire primer (20X).

Colonies were screened by colony PCR using primers T7P2 (5'-TGAGCGGATAACAATTCCC) and T7T (5'-GCTAGTTATTGCTCAGCGG) to select the positive transformants. The clones that gave positive result for PCR were selected and inoculated in L Broth with kanamycin (30 µg/ml final concentration) to grow overnight at 37°C, with shaking at 250 rpm. The recombinant plasmids were extracted using EZNA Plasmid Mini Kit II (Omega Bio-Tek) and sequenced using T7P2 (5'-3': TGAGCGGATAACAATTCCC), T7T (5'-3': GCTAGTTATTGCTCAGCGC) and an internal primer (5'-3': GACCCTGAGCTGGCTAAA). After confirmation that there were no mutations on cloned insert, the recombinant plasmid was transformed into *E. coli* BL21 (DE3) for protein expression (Method is in Chapter 2-section 2.2.10).

E. coli BL21 (DE3) expressing pET28b-*cmyr* was induced with 0.5 mM IPTG. Protein expression induction was carried out in two ways either at 25°C for 24 h or at 37°C for 3 h. After protein induction, CFEs were prepared by sonication as described in Chapter 2-section 2.3.1. Soluble and insoluble fractions were run on SDS-PAGE gel as described in Chapter 2-section 2.3.3.

5.3.4 Purification of Myrosinase and Measurement of Enzyme Activity

The CFEs were checked for the presence of the expressed protein by Western blot using antibodies raised against 6X-His tag (Chapter 2-section 2.3.4) and then cmyr was purified by affinity purification using Ni-NTA column (Chapter 2-section 2.3.5). Partially purified cmyr was

dialysed to remove excessive amount of imidazole from Ni-NTA column purification. Spectra/Por porous membrane tubing with 500-1000Da cut off (Spectrum Labs) was used for dialysis. The protein was dialysed against 20 mM citrate phosphate buffer pH 6.0 for 18 h at 4°C. The protein concentration of the dialysed fractions was quantified by Bradford assay and God-Perid assay was used to determine the enzyme activity of protein. These methods were explained in detail in Chapter 2 sections 2.3.2 and 2.3.6 respectively.

5.3.5 Dialysis and Filtration of Purified Protein for Crystallography

Ni-NTA purified cmyr was dialysed against 50 mM Tris-HCl 20 mM NaCl pH 7.0 buffer for 18 h at 4°C. Later, cmyr was filtered and concentrated for crystallography studies. Amicon Ultra 4-50k filters (Millipore) were used according to supplier's advice. The sample was loaded on to the filter unit and spun down at 4000 x *g* for 20 min at 4°C then concentrated samples were removed from filter device. The concentration of cmyr was adjusted to 10 mg/ml and sent to Imperial College London for crystallography studies.

5.3.6 Intact Mass Analysis

Dialysed protein sample was concentrated to 50 µM, this protein sample was analysed by LCMS to determine the intact mass of the protein. Protein intact mass was analysed on a Synapt G2-Si mass spectrometer coupled to an Acquity UPLC system (Waters, UK). An aliquot of ~300 pmoles of protein was injected onto an Aeris Widepore 3.6u C4 Column 50 x 2.1 mm (Phenomenex). Elution was performed by a gradient of 5-90% acetonitrile in 5 min with a flow rate of 0.4 ml/min. Masslynx 4.1 software (Waters) was used to control the spectrometer and it was operated in positive MS-TOF and resolution mode with a capillary voltage of 2.5 kV and a cone voltage of 40 V. Leu-enkephalin peptide (1 ng/ml, Waters) was infused at 3 µl/min as a lock mass and measured every 20 s. The spectra were produced by combining a number of scans, and deconvoluted using the MaxEnt1 tool in Masslynx. LCMS run was performed by Dr. Gerhard Saalbach (JIC, Norwich).

5.3.7 Characterisation of Myrosinase

The myrosinase activity of *C.* WYE1 was assayed using cell-free extracts previously by Albaser et al. (2016)[118]. In this study, the protein with 6XHis-tag was partially purified and further characterisation was performed. The optimum pH, temperature and buffer for myrosinase activity were studied and enzyme kinetics experiments were performed to determine K_m and V_{max} for sinigrin.

The optimum pH was tested in 20 mM citrate phosphate buffer (pH range of 3.6-7.6). Dialysed cmyr was added to the 300 µl mixture of sinigrin (2 mM, final), protein (24 ng) in 20 mM citrate phosphate buffer at a pH range of 3.0 - 7.6. The assay mixture was incubated at 37°C for 1 h and reaction was stopped by boiling the tubes for 5 min to inactivate the myrosinase. God-Perid assay was used to determine the glucose released. The optimum temperature was tested over a range of temperatures (5-70°C). The reaction mixtures were set up in 300 µl of volume consisting sinigrin (2 mM), protein (25 ng) in citrate phosphate buffer pH 6.0, incubated for 30 min and glucose release was measured by God-Perid assay.

To assess the the Michaelis-Menten constant (K_m) and maximum velocity (V_{max}) for sinigrin, 0.1-5 mM range of sinigrin was incubated with cmyr (24 ng) in 300 µl of citrate phosphate buffer pH 6.0 at 25°C for 10 min then activity was determined by God-Perid assay. The Lineweaver-Burk plot method was used to determine K_m and V_{max}. The myrosinase activity of cmyr was tested using 2 mM of different glucosinolates (sinigrin, glucoiberin, progoitrin, glucoerucin, gluroraphanin, glucotrapaeolin, gluconasturtiin and glucobrassicin) by God-Perid assay (Chapter 2-section 2.3.6). The assay was undertaken at optimum temperature and buffer conditions for myrosinase activity. In addition, the stability of myrosinase activity of cmyr was tested in 20 mM citrate phosphate buffer pH 6.0 over 5 weeks at 4°C.

5.4 RESULTS

5.4.1 Strain Identification and Preparation of Phylogenetic Tree

Rossiter Group at Imperial College isolated a strain with myrosinase activity from soil. We found that this soil isolate belongs to the family Enterobacteriaceae and the genus *Citrobacter*. 16s rDNA sequencing results matched with 20 known *Citrobacter* species. It was isolated from soil nearby Wye (Kent, UK) so named as *C.* WYE1 to show the origin of the isolate by Dr. John Rossiter before. 16s rDNA could not assign the strain to a specific species so a phylogenetic tree was constructed (Figure 5.1). The results suggested that *C.* WYE1 is likely to be a new species.

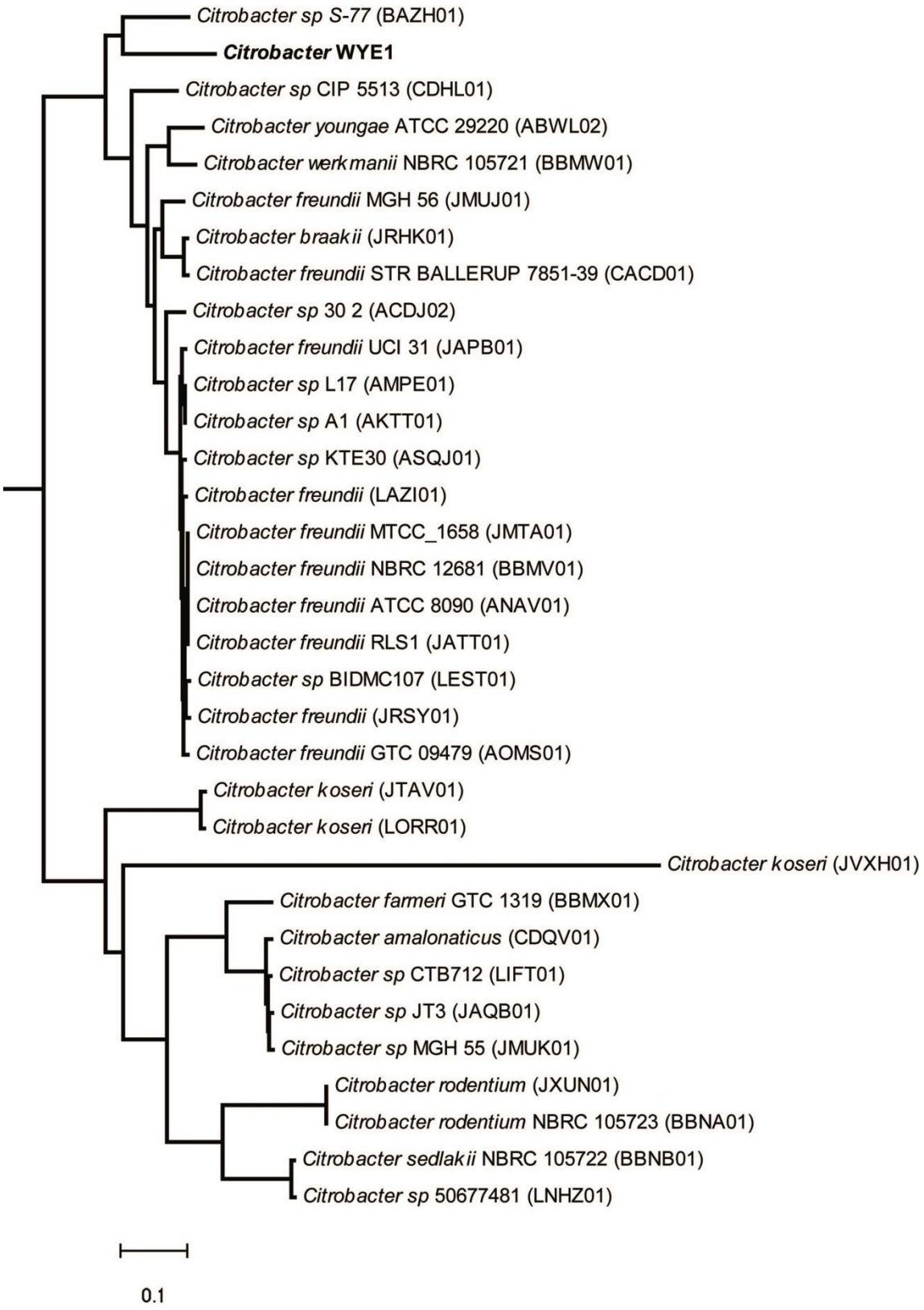


Figure 5.1 Phylogenetic tree of *C.* WYE1 compared with other *Citrobacter* species. Draft genomes of different *Citrobacter* species were obtained from Genbank whole genome shotgun projects. Core genome SNPs in draft genomes and *C.* WYE1 were used to prepare the phylogenetic tree using the parSNP program Version 1.2 by Dr. Arnoud Van Vliet (IFR).

5.4.2 The Relevance of *C.* WYE1 Myrosinase to Other Myrosinases

The protein sequence of cmyr was aligned with other myrosinases identified so far (Figure 5.2). There is a high homology among plant myrosinases from *A. thaliana*, *Eutrema japonicum*, *S. alba*, *B. napus* and *B. oleracea* var. *alboglabra*. The proton donor and nucleophile were reported to be Gln 207 and Glu 426 residues respectively in *S. alba* myrosinase obtained from Uniprot database and conserved among plant myrosinases aligned below (Figure 5.2). In *B. brassicae* myrosinase (aphid myrosinase), Glu 374 was reported to be the nucleophile and Glu 167 was reported as the proton donor [133]. These two key amino acid residues were not conserved in cmyr and it was found to be quite distinct from aphid and plant myrosinases. The importance of the conserved motif 'SDW' in cmyr possessing aspartic acid (D) as the nucleophile was suggested by Albaser et al. (2016) [118].

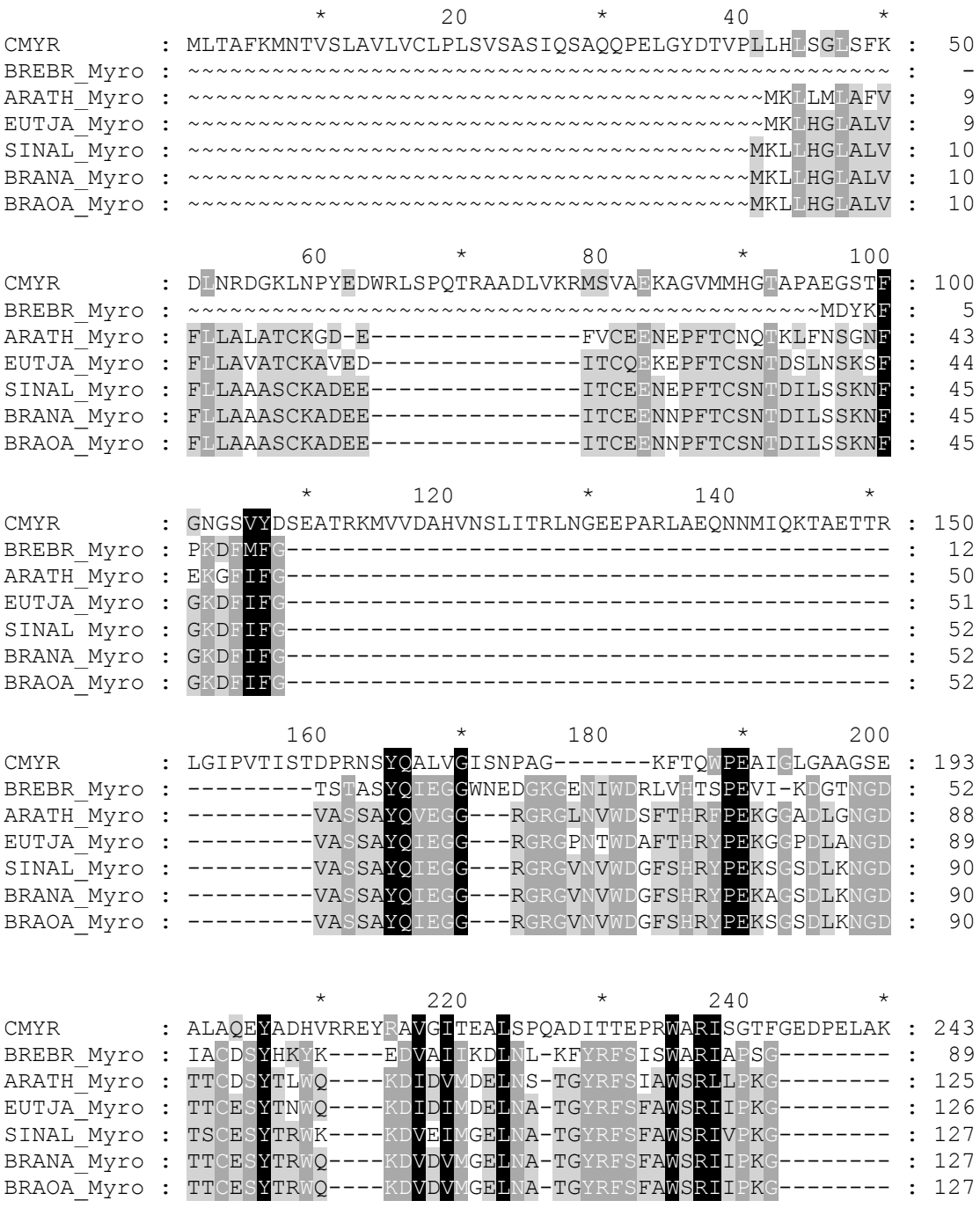


Figure 5.2 The multiple alignment of cmyr with some of the identified myrosinases. The abbreviations were used for myrosinase from: CMYR; *C.* WYE1, BREBR (Q95X01); *B. brassicae* (aphid), ARAT (P37702); *A. thaliana*, EUTJA (Q4AE75); *E. japonicum*, SINAL (P29092); *S. alba*, BRANA (Q00326); *B. napus*, BRAO (A6XG32); *B. oleracea var. alboglabra*. The default settings were used for Clustal Omega. The entry numbers of sequences at Uniprot database are given in brackets. SDW motif was highlighted in red. The proton donor and nucleophile in *S. alba* myrosinase are in Gln 207 (Q) and Glu 426 (E) positions, respectively and they are highlighted in yellow.

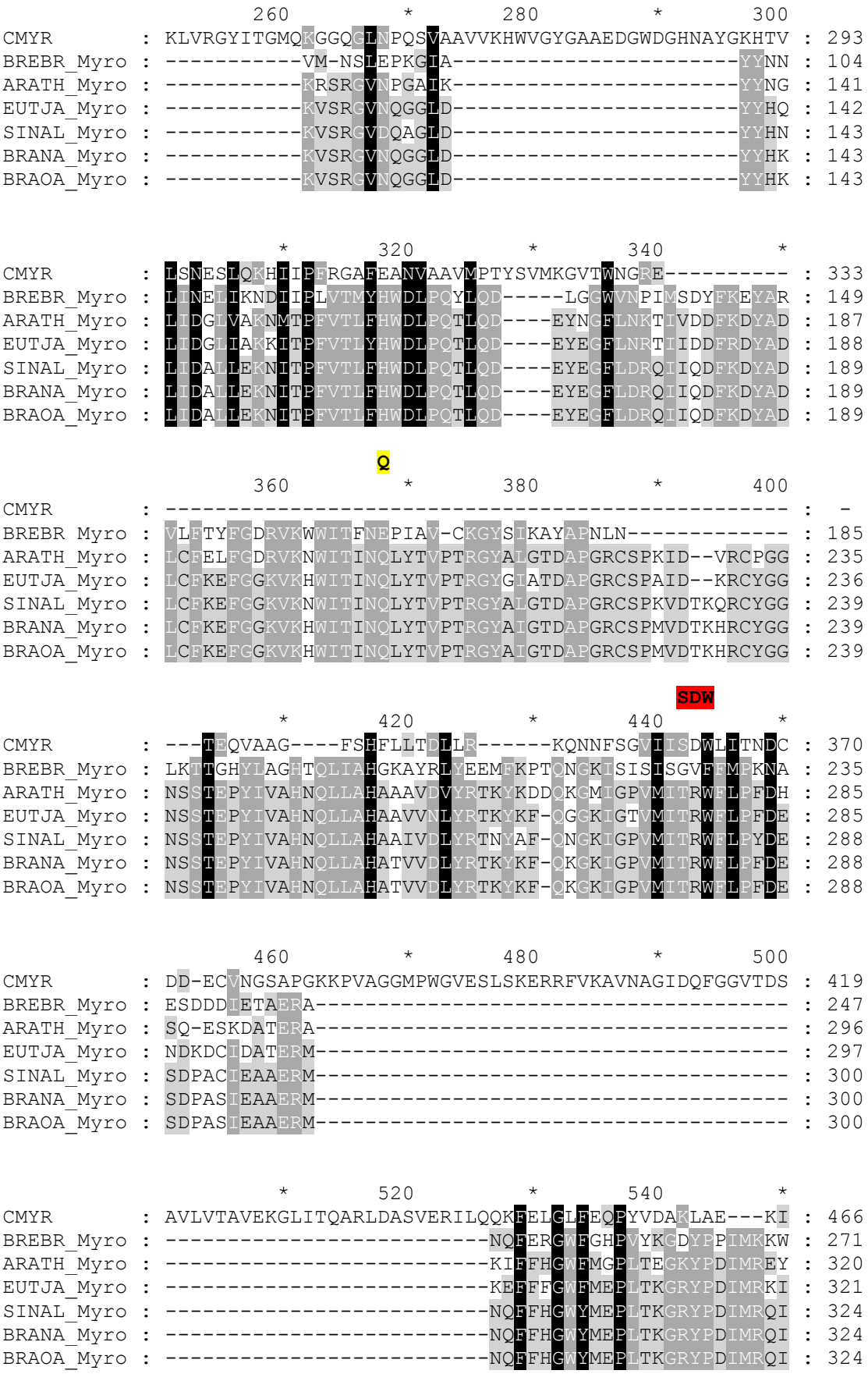

260 * 280 * 300
CMYR : KLVRGYITGMQKGGQGLNPQSVAAVVKHWVGYGAAEDGWDGHNAYGKHTV : 293
BREBR_Myro : ----------VM-NSLEPKGIA----------YYNN : 104
ARATH_Myro : ----------KRSRGVNPGAIK----------YYNG : 141
EUTJA_Myro : ----------KVSRGVNQGGLD----------YYHQ : 142
SINAL_Myro : ----------KVSRGVDQAGLD----------YYHN : 143
BRANA_Myro : ----------KVSRGVNQGGLD----------YYHK : 143
BRAOA_Myro : ----------KVSRGVNQGGLD----------YYHK : 143
* 320 * 340 *
CMYR : LSNESLQKHIIPFRGAFEANVAAVMPTYSVMKGVTWNGRE---------- : 333
BREBR_Myro : LTNELIKNDIIPLVTMYHWDLPQYLQD-----LGGWVNPIMSDYFKEYAR : 149
ARATH_Myro : LIDGLVAKNMTPFVTLFHWDLPQTLQD----EYNGFLNKTIVDDFKDYAD : 187
EUTJA_Myro : LIDGLIAKKITPFVTLYHWDLPQTLQD----EYEGFLNRTIIDDFRDYAD : 188
SINAL_Myro : LIDALLEKNITPFVTLFHWDLPQTLQD----EYEGFLDRQIIQDFKDYAD : 189
BRANA_Myro : LIDALLEKNITPFVTLFHWDLPQTLQD----EYEGFLDRQIIQDFKDYAD : 189
BRAOA_Myro : LIDALLEKNITPFVTLFHWDLPQTLQD----EYEGFLDRQIIQDFKDYAD : 189
Q
360 * 380 * 400
CMYR : ---------- : -
BREBR_Myro : VLFTYFGDRVKWWITFNEPIAV-CKGYSIKAYAPNLN---------- : 185
ARATH_Myro : LCFELFGDRVKNWITINQLYTVPTRGYALGTDAPGRCSPKID--VRCPGG : 235
EUTJA_Myro : LCFKEFGGKVKHWITINQLYTVPTRGYGIATDAPGRCSPAID--KRCYGG : 236
SINAL_Myro : LCFKEFGGKVKNWITINQLYTVPTRGYALGTDAPGRCSPKVDTKQRCYGG : 239
BRANA_Myro : LCFKEFGGKVKHWITINQLYTVPTRGYAIGTDAPGRCSPMVDTKHRCYGG : 239
BRAOA_Myro : LCFKEFGGKVKHWITINQLYTVPTRGYAIGTDAPGRCSPMVDTKHRCYGG : 239
SDW
* 420 * 440 *
CMYR : ---TEQVAAG----FSHFLLTDLLR------KQNNFSGVIISDWLITNDC : 370
BREBR_Myro : LKTTGHYLAGHTQLIAHGKAYRLYEEMFKPTQNGKISISISGVFFMPKNA : 235
ARATH_Myro : NSSTEPYIVAHNQLLAHAAAVDVYRTKYKDDQKGMIGPVMITRWFLPFDH : 285
EUTJA_Myro : NSSTEPYIVAHNQLLAHAAVVNLYRTKYKF-QGGKIGTVMITRWFLPFDE : 285
SINAL_Myro : NSSTEPYIVAHNQLLAHAAIVDLYRTNYAF-QNGKIGPVMITRWFLPYDE : 288
BRANA_Myro : NSSTEPYIVAHNQLLAHATVVDLYRTKYKF-QKGKIGPVMITRWFLPFDE : 288
BRAOA_Myro : NSSTEPYIVAHNQLLAHATVVDLYRTKYKF-QKGKIGPVMITRWFLPFDE : 288
460 * 480 * 500
CMYR : DD-ECVNGSAPGKKPVAGGMPWGVESLSKERRFVKAVNAGIDQFGGVTDS : 419
BREBR_Myro : ESDDDIETAERA---------- : 247
ARATH_Myro : SQ-ESKDATERA---------- : 296
EUTJA_Myro : NDKDCIDATERM---------- : 297
SINAL_Myro : SDPACIEAAERM---------- : 300
BRANA_Myro : SDPASIEAAERM---------- : 300
BRAOA_Myro : SDPASIEAAERM---------- : 300
* 520 * 540 *
CMYR : AVLVTAVEKGLITQARLDASVERILQQKFELGLFEQPYVDAKLAE---KI : 466
BREBR_Myro : ----------NQFERGWFGHPVYKGDYPPIMKKW : 271
ARATH_Myro : ----------KIFFHGWFMGPLTEGKYPDIMREY : 320
EUTJA_Myro : ----------KEFFFGWFMEPLTKGRYPDIMRKI : 321
SINAL_Myro : ----------NQFFHGWYMEPLTKGRYPDIMRQI : 324
BRANA_Myro : ----------NQFFHGWYMEPLTKGRYPDIMRQI : 324
BRAOA_Myro : ----------NQFFHGWYMEPLTKGRYPDIMRQI : 324

Figure 5.2-Continued

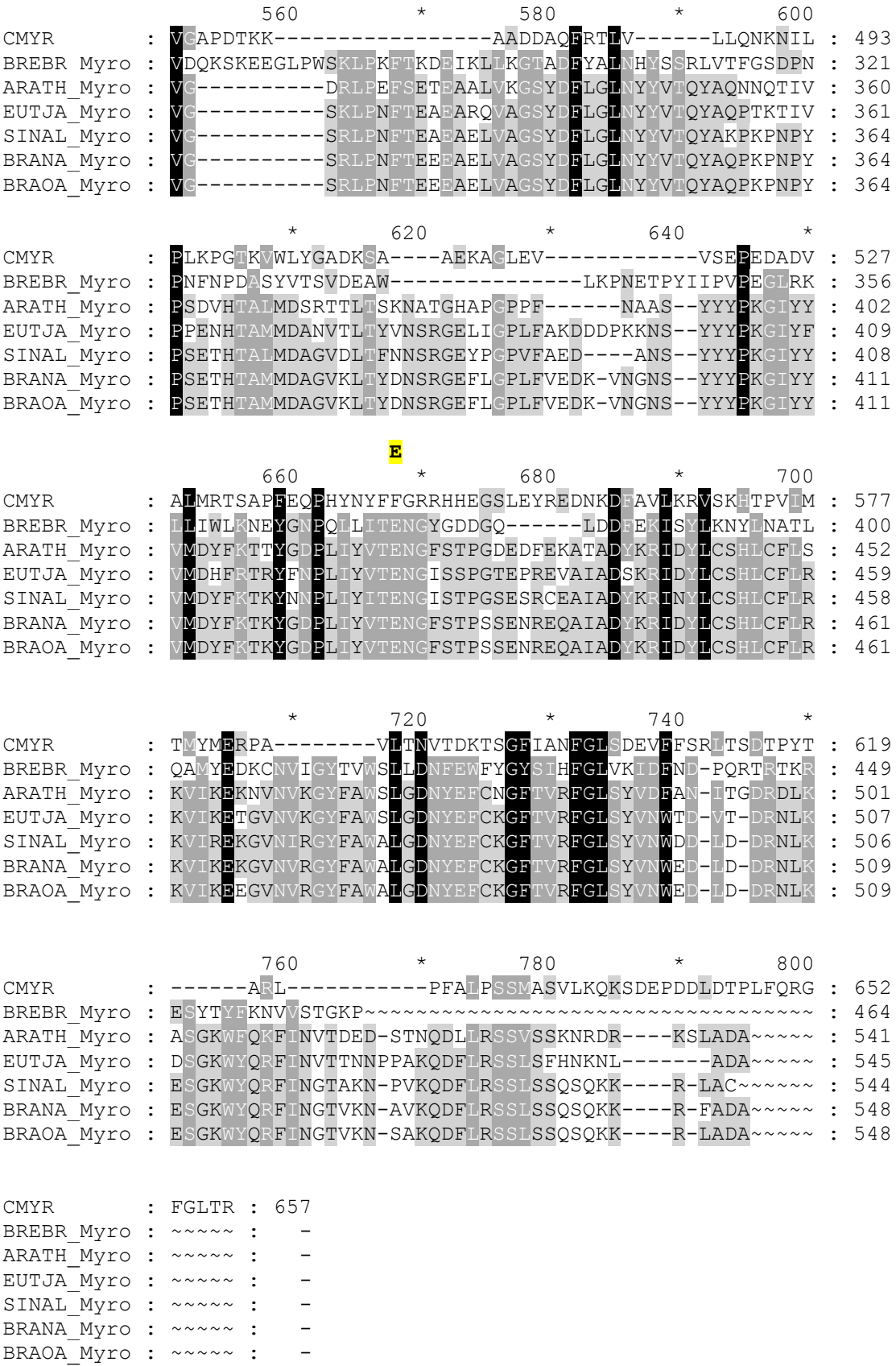


Figure 5.2 Continued

5.4.3 Cloning and Expression of Myrosinase Gene

The gene insert was amplified by PCR. The predicted size of the insert was 1998 bp as shown by gel electrophoresis (See Figure 5.3). In addition, double digested pET28b vector was run on DNA gel (Figure 5.3). After transformation of recombinant vector into *E. coli* DH5α cells, transformants on plates (10 clonies were picked) were checked by colony PCR, results are given in Figure 5.4.

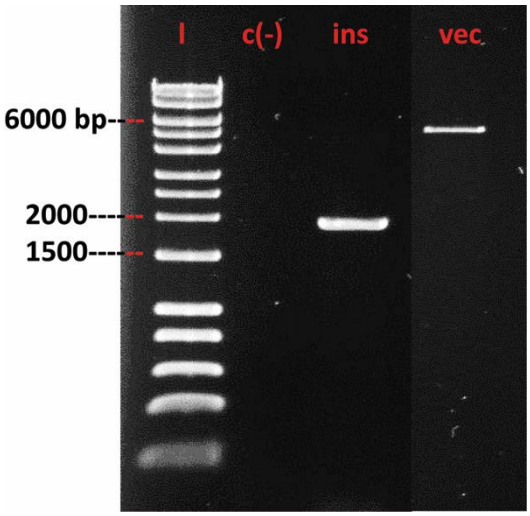


Figure 5.3 Agarose gel electrophoresis of cmyr insert and double digested pET28b vector. Samples were run on 1% agarose gel. l; hyperladder, c (-); negative control for PCR (PCR was set up without any genomic DNA), ins; cmyr insert, vec; Nde/XhoI digested pET28b.

The expected size of colony PCR products with correct gene insert was 2190 bp and expected sizes of the product bands from all transformants were observed (Figure 5.4). One of the positive transformants was used for the recombinant plasmid preparation. This recombinant plasmid was sequenced and confirmed to have no mutation in the gene insert. This recombinant plasmid was transformed into *E. coli* BL21(DE3) and the induction of protein expression was performed by addition of IPTG.

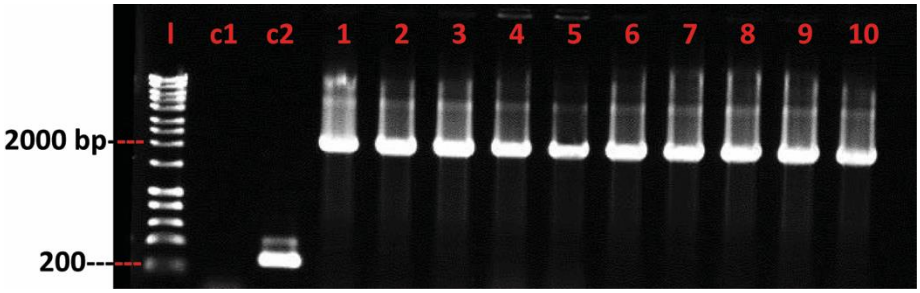


Figure 5.4 Agarose gel electrophoresis for colony PCR experiments. PCR products were run on 1% agarose gel. l; hyperladder, c1; negative control for PCR (PCR was set up without any genomic DNA), c2; positive control for PCR (PCR was set up with pET28b vector), 1-10; PCR products of selected colonies.

Soluble and insoluble extracts from protein expression were run on SDS-PAGE gel (Figure 5.5). Protein expression at 25°C for 24 h resulted in yield of expressed cmyr in the soluble fraction (Figure 5.5- the protein band is shown) compared to protein expression at 37°C for 3 h. Protein expression at 37°C for 3 h results in cmyr to be in the insoluble fraction (Figure 5.5- Lane 8).

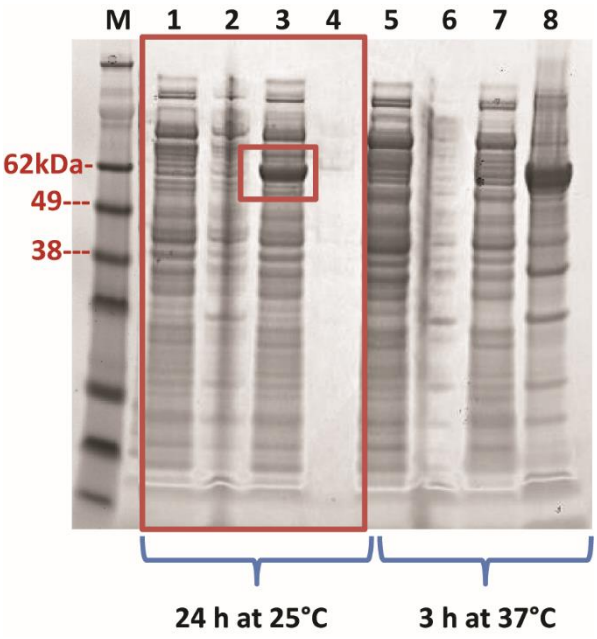


Figure 5.5 SDS-PAGE of the CFEs from *E. coli* BL21 (DE3) expressing pET28b-*cmyr*. 10 µg of protein was loaded per lane. M: See Blue Protein Marker, 1-2: soluble and insoluble CFEs obtained from uninduced *E. coli* BL21 (DE3) expressing pET28b-*cmyr* grown at 25°C for 24 h respectively, 3-4: soluble and insoluble CFEs obtained from induced (at 25°C for 24 h) *E. coli* BL21 (DE3) expressing pET28b-*cmyr* respectively, 5-6: soluble and insoluble CFEs obtained from uninduced *E. coli* BL21 (DE3) expressing pET28b-*cmyr* grown at 37°C for 3 h respectively. 7-8: soluble and insoluble CFEs obtained from induced (at 37°C for 3 h) *E. coli* BL21 (DE3) expressing pET28b-*cmyr* respectively. Protein gel was run on 4-12% Bis-Tris gel under denaturing, reducing conditions in MES buffer. Insoluble CFEs were difficult to load on gel so the protein bands are not as clear as soluble ones.

5.4.4 Western Blotting and Purification of Myrosinase

The CFEs were prepared and checked for the presence of 6X-His tag by Western blot (Figure 5.6) using a His-tag antibody (Novagen). When protein expression was performed at 25°C for 4 h instead of 24 h and it also produced a better yield of cmyr protein. The expected protein size of cmyr with N- and C-terminal 6X-His tags was 75 kDa. Recombinant proteins with expected protein sizes were found to be recognised by the His-tag antibody so purification of cmyr using Ni-NTA column was performed.

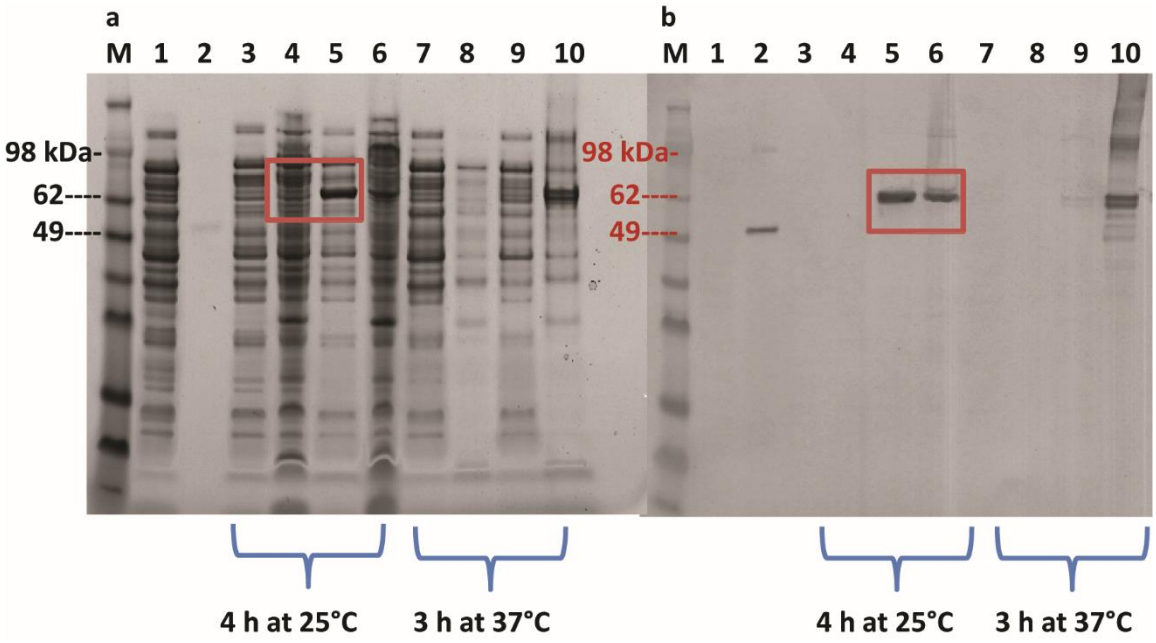


Figure 5.6 SDS-PAGE of soluble and insoluble extracts from *E. coli* BL21(DE3) expressing pET28b-*cmyr*. a) Protein gel where CFEs were run on 4-12% Bis-Tris gel under denaturing, reducing conditions in MES buffer. b) Western blot (using anti-His tag antibody) within the same order of protein gel in a. 10 µg of protein was loaded per lane. 1; *E. coli* BL21(DE3) pET28b soluble CFE, 2; A protein with 6XHis-tag (positive control for Western blotting), 3-4; Soluble and insoluble CFEs of uninduced cells grown at 25°C for 4 h , 5-6; Soluble and insoluble CFEs of induced (25°C, 4 h) cells, 7-8; Soluble and insoluble CFEs of uninduced cells grown at 37°C for 3 h, 9-10; Soluble and insoluble CFEs of induced (37°C, 3 h) cells.

The cmyr was purified by Ni-NTA column (Figure 5.7) and the protein levels were quantified by the Bradford assay.

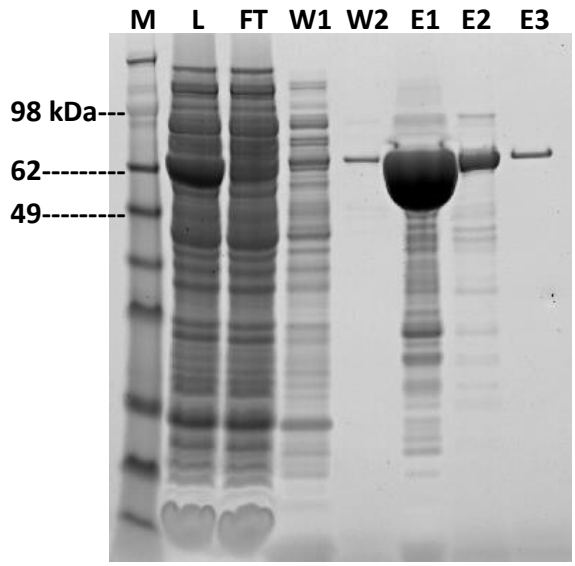


Figure 5.7 SDS-PAGE of the Ni-NTA purified Cmyr protein from *E. coli* BL21 (DE3) expressing pET28b-*cmyr*. 3.25 µl of fractions were loaded per lane. Proteins were run on 4-12% Bis-Tris Gel under denaturing, reducing conditions in MES buffer. L: Lysate, FT: Flow through, W: Wash, E: Eluate, M; Protein marker. The expected protein size of cmyr with N- and C-terminal 6X-His tags was 75 kDa.

The presence of two strong bands in the Western Blot (Figure 5.6 b- Lane 10) was observed. It was hypothesized that these bands might present the whole protein and the protein with signal peptide cleaved off. This was further investigated by intact mass analysis after purification of cmyr. When cmyr has two 6XHis-tags at N and C terminal site, the predicted size of the protein is ~75 kDa. However if the signal peptide was cleaved off then the predicted size is ~70 kDa. Intact mass analysis showed that cmyr has a measured mass of 70,332 Da so, it was suggesting that the signal peptide was cleaved off correctly (See Figure 5.8 for MS chromatogram) as predicted size of recombinant cmyr without signal peptide is ~70 kDa. Purified, dialysed and filtered cmyr was adjusted to a concentration of 10 mg/ml and the sample was sent to Imperial College London for crystallography. This part of the study is still ongoing.

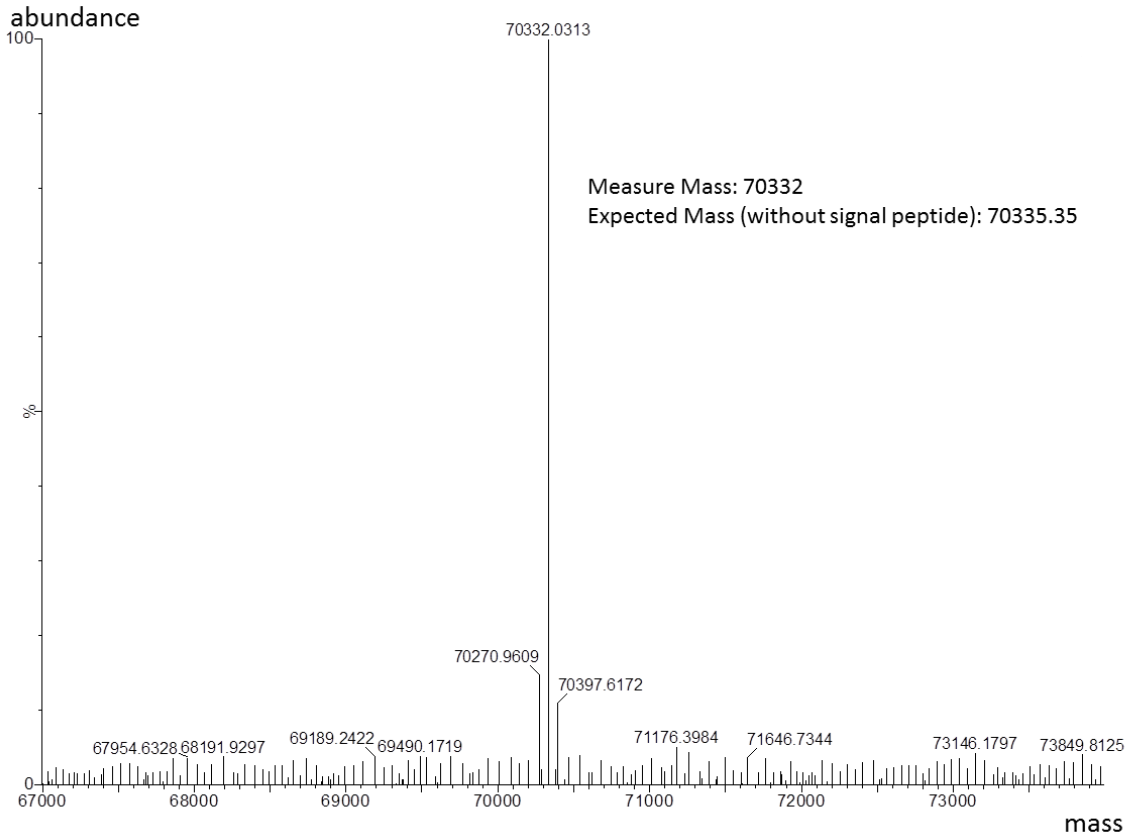


Figure 5.8 The MS spectrum for cmyr. Appropriate fragments of cmyr were determined in positive MS-TOF.

5.4.5 Myrosinase Activity of Recombinant cmyr

The myrosinase activity of dialysed and undialysed proteins were tested by God-Perid assay. The results showed that recombinant cmyr was active against sinigrin. Dialysis had a positive effect on myrosinase activity of proteins. When the eluate was used in the assay without dialysis, it showed an activity of 25.4 ± 0.74 µmol glucose/min/mg enzyme. Dialysed eluate of Cmyr had a specific enzyme activity of 57.2 ± 9.4 µmol glucose/min/mg enzyme. Dialysed protein showed more than two fold higher enzyme activity compared to the enzyme activity of undialysed protein. *S. alba* myrosinase (Sigma Aldrich) was used as a positive control and gave a positive result for myrosinase activity (0.33 ± 0.05 µmol glucose/min/mg enzyme). Negative control (without protein) did not show any myrosinase activity.

When myrosinase activity of cmyr was tested in 20 mM citrate phosphate buffer within a pH range of 3.6-7.6, pH 6.0 was found to be optimum for myrosinase activity (Figure 5.9). The myrosinase activity was found to be gradually reduced above pH 6.0.

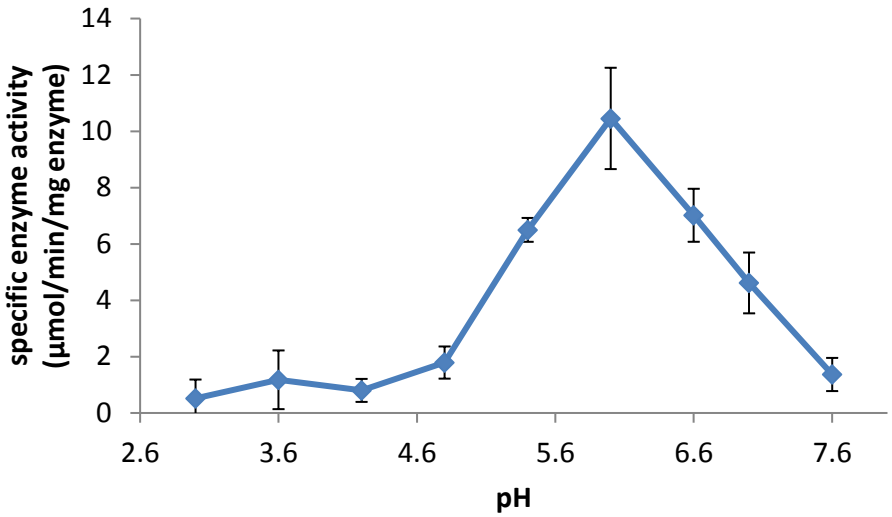


Figure 5.9 pH optimum range of recombinant cmyr myrosinase activity in 20 mM citrate phosphate buffer. The myrosinase activity was determined at 37°C for 1 h by God-Perid assay. Error bars = standard deviation (n=3).

When myrosinase activity of recombinant cmyr was tested for its stability over a range of temperatures (5-70°C), the optimum temperature was found to be 25°C under the conditions tested (Figure 5.10). The enzyme activity was reduced significantly over 30°C and further diminished over 50°C.

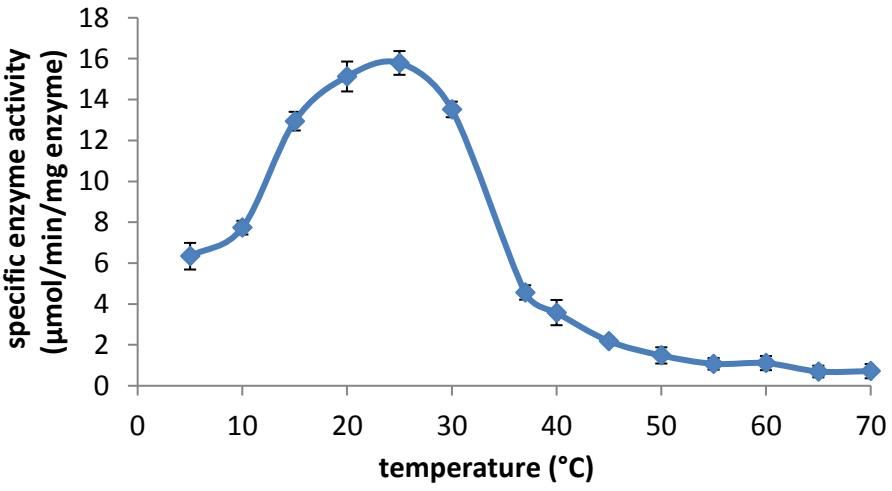


Figure 5.10 Temperature stability of *C.* WYE1 myrosinase in 20 mM citrate phosphate buffer pH 6.0. The myrosinase activity was determined by God-Perid assay. Error bars = standard deviation (n=3).

The optimum temperature of cmyr for myrosinase activity was different from plant myrosinases [118, 253] which their optimum temperatures are generally higher (around 45-50°C). The aphid (B. brassicae) [137] and plant (A. rusticana) myrosinase [118] were reported to retain their activity at 4°C for months but it was not reported what percentage of the activity remained. When the stability of myrosinase activity over time was analysed, myrosinase lost 97% of its original enzyme activity in 5 weeks when kept at 4°C. This result required us to use fresh made Ni-NTA purified proteins for assays each time.

5.4.6 Enzyme Activity and Kinetics

The kinetic parameters of the purified cmyr were determined using sinigrin as substrate. K_m and V_{max} were estimated as 0.2 mM and 0.033 mmol/min/mg for sinigrin at pH 6.0 and 25°C (Figure 5.11).

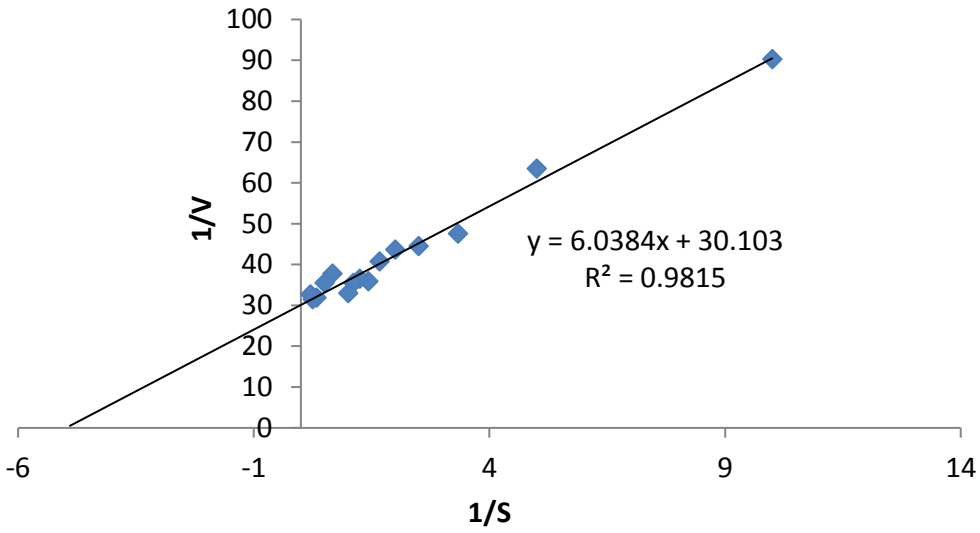


Figure 5.11 Kinetic analysis of cmyr using sinigrin as substrate. Specific activity (mmol/min/mg enzyme) of cmyr (24 ng incubated) at different concentrations of sinigrin (0.1-5 mM) was used to prepare Lineweaver Burk plot.

5.4.7 Substrate Specificity of Recombinant cmyr

The enzyme activity of cmyr against different glucosinolates including sinigrin, glucoiberin, progoitrin, glucoerucin, gluroraphanin, glucotrapaeolin, gluconasturtiin and glucobrassicin were tested by God-Perid assay (Figure 5.12). Cmyr showed activity towards all of the glucosinolates except gluconasturtiin and glucobrassicin. The specific activity of cmyr for sinigrin was highest, followed by glucoiberin > progoitrin > glucoerucin > glucoraphanin > glucotrapaeolin > gluconasturtiin=0. The cmyr had similar specific activity for glucoerucin,

glucoiberin and progoitrin but the activity of glucoraphanin and glucotrapaeolin was lower compared to those. The cmyr was not tested for activity against gluconasturtiin and glucobrassicin previously [118] and tested in this study but there was no activity towards gluconasturtiin under the conditions tested and glucobrassicin stock used in the assay gave false positive result.

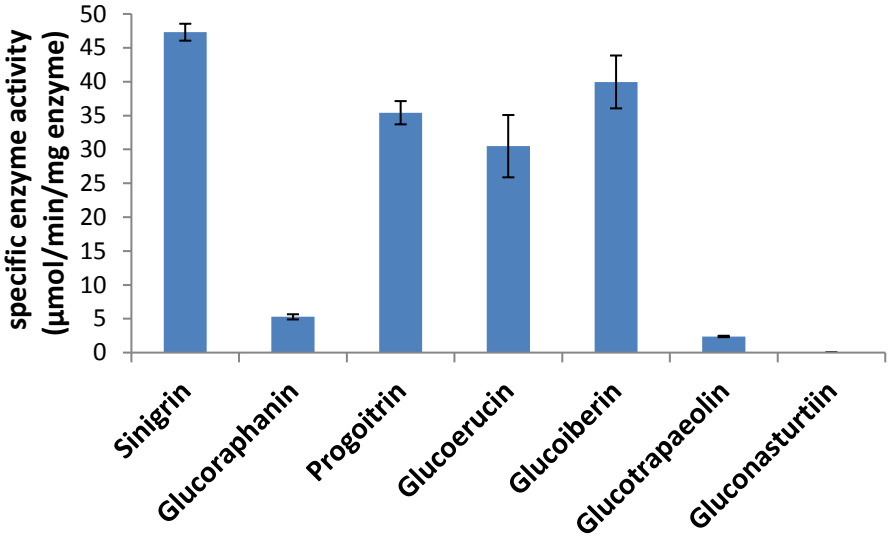


Figure 5.12 The myrosinase activity of cmyr towards different glucosinolates. The myrosinase activity was determined at 25°C in 20 mM citrate phosphate buffer by God-Perid assay. Error bars = standard deviation (n=3).

5.4.8 Stability of Myrosinase Activity of Recombinant cmyr

Dialysed and filtered cmyr was tested for the stability of its myrosinase activity in 20 mM citrate phosphate buffer pH 6.0 at 4°C. According to results, cmyr lost half of its original myrosinase activity in a week. In 5 weeks, cmyr could only keep 3% of the original activity. This data was especially important and used in the planning of the characterisation studies.

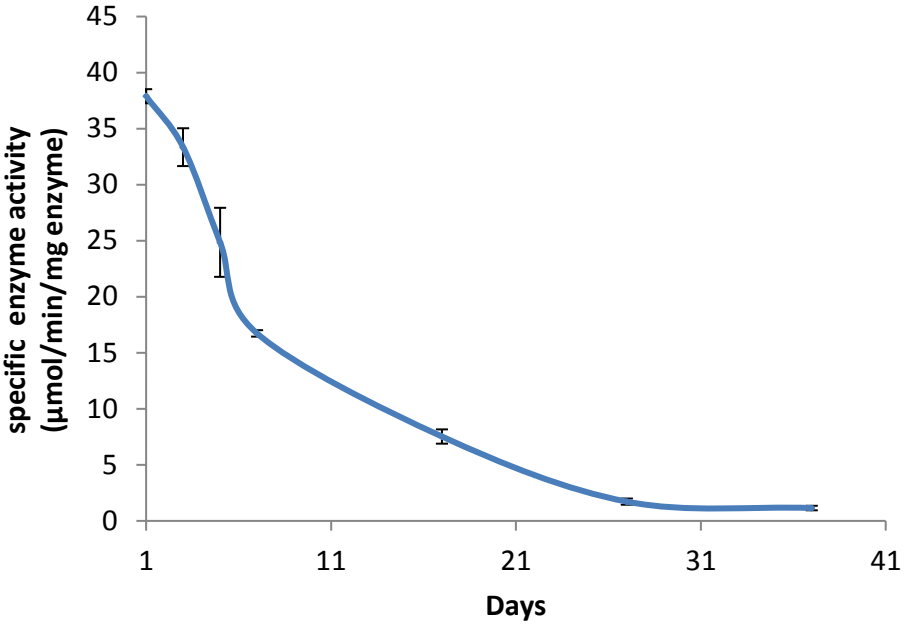


Figure 5.13 Stability of myrosinase activity of cmyr in 20 mM citrate phosphate buffer pH 6.0 at 4°C over time. The myrosinase activity was determined at 25°C in 20 mM citrate phosphate buffer by God-Perid assay. Error bars = standard deviation (n=3).

5.5 DISCUSSION

The publication of cmyr activity during the course of this study led to changes in the experimental plans for the investigation of bacterial myrosinases from human gut. As bacterial myrosinases were suggested to be more closely related to aphid myrosinases than plant myrosinases [56], the search for potential myrosinases in the genomes of human gut bacteria was initially based on aphid myrosinase amino acid sequences. Aphid myrosinases belong to GH1 enzyme family so we were unfortunately excluding the enzymes from GH3 and others. As cmyr is a periplasmic β-glucosidase from GH3 family, we also included enzymes from GH3 for cloning experiments to identify a β-glucosidase responsible for myrosinase activity.

The myrosinase activity of *C.* WYE1 was characterised using CFEs and the myrosinase gene (*cmyr*) was identified previously by Albaser et al. [118]. In this study, *cmyr* gene was cloned and expressed in *E. coli* with both N and C-terminal 6XHis-tags. The protein with 6xHis-tag was purified and further characterisation was carried out. Dialysis of Ni-NTA purified protein enhanced myrosinase activity of cmyr. The optimum pH, temperature and buffer for

myrosinase activity were found to be same with the study of Albaser et al.(2016) [118]. K_m and V_{max} values were reported to be 0.46 mM and 4.91 mmol dm^{-3} min^{-1} mg^{-1} respectively by Albaser *et al*. (2016). Our results are different from their results but this was expected as they used CFE in their study. K_m value of aphid (*Brevicoryne brassicae*) myrosinase for sinigrin was investigated by two different research groups and it was found to be 0.41 mM (at pH 4.5, 30°C)[136] and 0.61 mM (at pH 5.5, 30°C)[132]. The K_m value of the cmyr (0.2 mM) for sinigrin was lower than these values indicating a higher affinity for sinigrin and similar to K_m value of plant (*Lepidium latifolium L.)* myrosinase (0.17 mM, at pH 6.0, 37°C) [249].

The optimum pH of 6 for cmyr myrosinase activity was found to be similar to some plant myrosinases such as from *Brassica napus* [135], *Armoracia rusticana* [117] and *Lepidium latifolum* L. [249]. It was lower than fungous myrosinase [250] and higher than aphid myrosinase [133]. The myrosinase activity of cmyr was found to be gradually reduced above pH 6.0 that was different from plant myrosinases such as *Brassica napus* [135] and *Armoracia rusticana* [117] myrosinases. Li et al. (2005) reported that optimum pH was 5.7 for horseradish (*Armoracia rusticana*) myrosinase and myrosinase activity was reported to be retained 80% of the maximum activity between pH 5.0-8.0 values. However, there was a significant decrease in myrosinase activity above optimum pH especially above pH 7.0 for cmyr. This might be explained by the origin of the *C.* WYE1 as an adaptation to soil pH. It was isolated from soil from Kent (UK) and UK soil was reported to be acidic in general (https://commons.wikimedia.org/wiki/File:World_Soil_pH.svg).

The cmyr showed a variable specificity to different glucosinolates. The preference of aliphatic glucosinolates (glucoiberin, progoitrin, glucoerucin and glucoraphanin) over aromatic glucosinolates (glucotropaeolin and gluconasturtiin) might indicate the importance of the side chain for myrosinase activity of cmyr. Selective glucosinolate breakdown profile was seen in many myrosinases including aphid (*B. brassicae*) myrosinase [132, 136], plant myrosinases [125, 126, 249]. For instance, aphid (*B. brassicae*) myrosinase was assessed to have higher specificity for sinigrin than glucotropaeolin in two seperate studies [132, 136]. A plant (*Lepidium latifolum* L.) myrosinase showed a better myrosinase activity against sinigrin compared to glucoraphanin [249]. Two myrosinase isoenzymes isolated by James and Rossiter (1991) showed better specificity for aliphatic glucosinolates rather than indole glucosinolates [125], while another study reported the existance of a myrosinase that was highly specific for epi-progoitrin [126].

Ascorbate is reported to be an activator for plant myrosinases [117, 119, 249]. However, aphid myrosinases of *B. brassicae* and *Lipaphis erysini* are reported to be unaffected by the presence of ascorbate [133]. The effect of ascorbate on cmyr myrosinase activity was not determined in this study but it was already reported that the enzyme was activated with ascorbate by a factor of 1.67 [118].

Successful cloning and expression of myrosinase made it possible to produce large amounts of protein for further characterisation. The importance of a conserved amino acid sequence, SDW, was reported for the enzyme activity of cmyr and aspartic acid (D) was suggested as the nucleophile by Albaser et al. (2016) [118]. However, crystallography studies for the enzyme in complex with glucosinolates are needed for confirmation of this and identification of the acid catalyst, nucleophile and other important amino acid residues for myrosinase activity. The crystallisation of this myrosinase is being performed in collaboration with Imperial College London.

In summary, the identification of cmyr is important to have a better understanding of glucosinolate metabolism by bacteria as all identified and characterised myrosinases are from GH1 family so far. The cmyr was also used as a control positive in the other parts of the study. In this chapter, recombinant cmyr was characterised and will be crystallised to determine its structure in the future. Crystallography studies about cmyr may give us a clearer idea of the nature and activity of this new GH3 class bacterial myrosinases.

CHAPTER SIX

6 CONCLUSIONS AND FUTURE WORK

6.1 CONCLUSIONS AND FUTURE WORK

Several studies have investigated the glucosinolate metabolism by human gut bacteria so far but the mechanisms are still not well understood [56, 57, 112, 165, 211-213, 216]. Especially, very little is known about the enzymes involved in glucosinolate metabolism. This study mainly aimed to identify and characterise these enzymes involved in the metabolism of glucosinolates including the specific roles of some myrosinases and reductases. To achieve this, the candidate genes were cloned and expressed then enzymatic assays were performed to assess the activity of these proteins. To identify the first bacterial myrosinase from human gut, a cloning approach and a FPLC approach which is a combination of ion exchange chromatography and gel filtration was used. The bacterial strains that were examined for their myrosinase or reductase activity are given in Table 6.1.

Bacterial Strain	Source of the strain	Activity
L. agilis R16 [56]	Human gut (provided to us)	Myrosinase, reductase
E. coli VL8 [56]	Human gut (provided to us)	Reductase
E. coli FI10944	Human gut (provided to us)	Myrosinase
E. casseliflavus CP1 [56]	Human gut (provided to us)	Myrosinase
E. coli FC44	Human gut (isolated in this study)	Reductase
C. freundii FC50	Human gut (isolated in this study)	Myrosinase
C. WYE1 [118]	Soil (provided to us)	Myrosinase

Table 6.1The list of bacterial strains studied in this research.

Three strains (*E. coli* FI10944, *E. coli* VL8, *E. casseliflavus* CP1) were previously isolated by our collaborator, Dr. Vijitra Luang In, using a sinigrin enrichment method. *C.* WYE1 was isolated from soil by Dr. Abdulhadi Albaser using a sinigrin enrichment method. *L. agilis* R16 was included to the study because it was reported to have myrosinase activity previously [210]. *E. coli* VL8, *L. agilis* R16 and *E. casseliflavus* CP1 were found to be capable of metabolising glucosinolates with different efficiencies [56]. In addition to these previously isolated strains, a glucoraphanin enrichment method was performed in our study and isolated 98 more human gut bacteria with glucosinolate metabolism. Fifteen of these isolates were determined to degrade all of the glucoraphanin supplied in the media and ten of them were identified by 16S rDNA sequencing at the genus level. These isolates were identified with myrosinase activity but not all of them were further studied. Two of the human gut strains, *E. coli* FC44

and *C. freundii* FC50, were further studied for their reductase and myrosinase activities respectively.

In Chapter 3, *L. agilis* R16 and *E. casseliflavus* CP1 were tested for their glucosinolate degrading ability. Interestingly, the glucoraphanin degradation rate by *E. casseliflavus* CP1 was lower in this study compared to the degradation rate reported previously by Luang-In et al. (2014)[56]. This was possibly an effect of long time storage of the strain away from its original environment. *L. agilis* R16 showed a low glucoraphanin degradation rate which was similar to the degradation rate reported previously [56]. It was thought that pre-culturing in media with glucosinolates might have an effect of glucoraphanin degradation ability of *L. agilis* R16. However, pre-culturing with glucosinolates did not cause the anticipated increase in glucoraphanin or sinigrin degradation rates by this strain under the conditions tested.

When *E. coli* FI10944 and *C. freundii* FC50 were screened for myrosinase activity, myrosinase activity was detected in CFEs of both strains. Therefore, it was concluded that the myrosinase activity in these strains was not cell-associated. This result was contradictory to the previous studies reporting the necessity of intact cells for myrosinase activity from *E. coli* VL8 [56], *L. agilis* R16 [56, 210], *E. casseliflavus* CP1 [56] and *E. coli* O157:H7 [217]. As myrosinase activity was reported to be cell-associated and not identified in CFEs of bacteria so far, it was proposed that the phoshotransferase system (PTS) is involved in uptake and phosphorylation of glucosinolates by β-glucoside kinase and it was suggested that the disruption of cell-integrity destroyed the action of PTS and myrosinase activity could not be achieved [56]. This can be possible for *E. casseliflavus* CP1 as no myrosinase activity of CFEs or recombinant β-glucosidases was observed. However, our results showed that myrosinase activity is not always cell wall-associated. The myrosinase activity in CFEs of *E. coli* FI10944 and *C. freundii* FC50 was examined to see if it could be induced by adding glucosinolate to the original growth media. It was found that both induced and uninduced CFEs of these two strains had myrosinase activity. The results suggested that the myrosinase activity in CFEs of *E. coli* FI10944 and *C. freundii* FC50 may be constitutively expressed.

To identify the genes encoding myrosinase activity, putative β-glucosidase genes of *E. coli* FI10944 (*ecolg*3) and *E. casseliflavus* CP1 (*ecg*4, *ecg*39, *ecg*44) were cloned and the proteins expressed successfully (Chapter 3). However, none of them were identified with myrosinase activity. Two of them (*ecg*4, *ecg*39) showed β-glucosidase activity against 4-nitrophenyl β-D-glucopyranoside so it was concluded that these two enzymes were active but did not act on

glucosinolates. As we hypothesized that specific β-glucosidases are responsible for myrosinase activity, either these enzymes were not the right β-glucosidase for myrosinase activity or glucosinolate metabolism by gut bacteria is more complex than we originally thought. Desulfo-glucosinolates were reported to be degraded by *E. coli* VL8, *L. agilis* R16 and *E. casseliflavus* CP1 to form nitriles as the only final products [57, 137]. Nitrile formation is not the main focus in this study but this bioconversion is also important because it suggests the presence of an alternative pathway for glucosinolate metabolism by gut bacteria (Figure 6.1). This alternative pathway includes sulfatase activity to form desulfo-glucosinolates and then β-glucosidase activity to produce nitriles. The sulfatases from *E. coli* VL8 were cloned, expressed and characterised by our collaborator previously, where they found that the crude extracts of recombinant sulfatase from *E. coli* VL8 was found to have sulfatase activity [137]. A recombinant β-glucosidase from *Caldocellum saccharolyticum* was able to transform desulfo-glucosinolates into nitriles [251]. Considering these studies, ecg4 and ecg39 could be checked for their activity towards desulfo-glucosinolates. This can shed light on the proposed metabolism of desulfo-glucosinolates by gut bacteria such as *E. casseliflavus* CP1 and confirm the existence of an alternative pathway (Figure 6.1). Setting up a GC-MS method for identification of nitriles was achieved but ecg4 and ecg39 was not examined for their ability to degrade desulfo-glucosinolates.

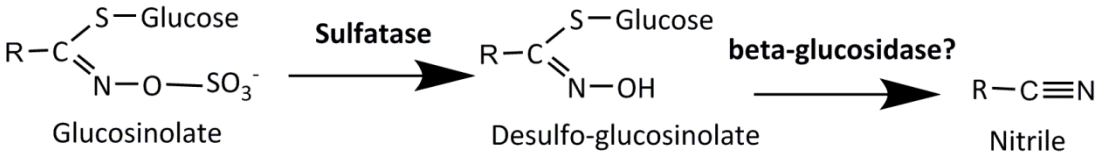


Figure 6.1 Proposed metabolism of desulfo-glucosinolates by human gut bacteria. R indicates the side chain in glucosinolate structure.

A study showed that the disruption of 3 β-glucosidases (*bglA, ascB, chbF*) in *E. coli* O157:H7 affected its myrosinase activity [240]. They concluded that *bglA* and *ascB* were involved in myrosinase activity. However, disruption of these genes did not completely abolish the myrosinase activity [240]. *E. casseliflavus* CP1 and *E. coli* FI10944 have 40 and 6 potential glucosidases in their genome respectively. In this study, we cloned and expressed the four most likely β-glucosidases based on their sequence similarities with identified bacterial and aphid myrosinases. This number of β-glucosidases we cloned represents only a small fraction of all putative glucosidases. Disruption of these genes and screening the mutants for myrosinase activity would take time but it could be tried. The putative glucosidase genes in *E. casseliflavus* CP1 and *E. coli* FI10944 could be disrupted then myrosinase activity of the

mutant *E. casseliflavus* CP1 and *E. coli* FI10944 strains could be assessed. If any correlation would be found between the disruption of genes and the decrease in the myrosinase activity, the genes could be cloned and expressed in *E. coli* and their myrosinase activity could be determined.

When purification of bacterial myrosinase in *E. coli* FI10944 by FPLC was attempted, only cytoplasmic α-amylase and periplasmic trehalase were identified in fractions showing myrosinase activity. These enzymes can act on α-glucosidic bonds. Therefore, it is unlikely that these enzymes can play a role in degradation of β-thioglucosides (glucosinolates). The FPLC fraction with myrosinase activity was analysed by LCMS/MS and the enzymes found in high abundance were phosphoglycerate kinases and triosephosphate isomerases which belong to the glycolysis pathway. Their presence is possibly associated with utilisation of glucose in glucosinolates, not with glucosinolate degradation by human gut bacteria. In brief, identification of bacterial myrosinase from human gut could not have been achieved using cloning and chromatography approaches. However, the results from chapter 3 emphasized the different capacity and diverse routes of glucosinolate metabolism by human gut bacteria.

A myrosinase from the bacterial strain of *C.* WYE1 of soil origin was identified when it was grown in M9 minimal media including sinigrin as the sole carbon source [118]. This experiment was carried out in our study with *E. coli* FI10944 in minimal media with 1 mM sinigrin but *E. coli* FI10944 failed to grow in these conditions. Using a different glucosinolate as a sole carbon source could support the growth of *E. coli* FI10944, this culture might be used for the purification of the myrosinase of *E. coli* FI10944. *C.* freundii FC50 is also a good candidate for a further study thanks to its myrosinase activity. Its genome was assembled and annotated. A candidate gene (a periplasmic β-glucosidase) has 25% amino acid sequence identity with *C.* WYE1 myrosinase (cmyr). This gene might be a good candidate to clone and express. Besides, this strain can be used to purify the myrosinase by chromatographic methods as described for *C.* WYE1 by Albaser et al. (2016) [118].

In chapter 4, *E. coli* VL8 and *E. coli* FC44 were studied for their ability to reduce glucoraphanin to glucoerucin. The CFEs of both strains were able to reduce glucoraphanin to glucoerucin so CFEs were tested to see whether the reductase activity in CFEs was inducible or not by glucoraphanin supplied to the media. It was found that glucoraphanin induced or uninduced CFEs of *E. coli* VL8 and *E. coli* FC44 had similar reductase activity. This suggested that reductase activity in CFEs of *E. coli* VL8 and *E. coli* FC44 may be constitutive which contradicts

previous work where it was reported that uninduced CFE of *E. coli* VL8 was unable to reduce glucoraphanin [56].

In addition, the difference in glucoraphanin degradation rates between *L. agilis* R16 *and E. coli* VL8 was investigated and we aimed to identify the reductase responsible for converting methylsulfinylalkyl glucosinolates (such as glucoraphanin) into methylthioalkyl glucosinolates (glucoerucin). To identify the reductases involved in glucosinolate metabolism, the study was focused on methionine sulphoxide reductases because MsrA was proposed to be responsible for reduction of glucosinolates [55]. There was no information about stereoisomerism of sulphoxide group in the glucoraphanin used in our study, it might consist of glucoraphanin with *S* or *R*-epimer of sulphoxide group. However, glucoraphanin extracted from *Brassica oleracea* var *italica* (broccoli) and *Arabidopsis thaliana* was found to be a pure epimer of *R* configuration at the sulphoxide group [247]. As MsrA can only act on the *S*-epimer of methyl sulphoxide, MsrB and fRMsr which can act on *R*-epimers of methyl sulphoxide were also investigated for their role in reducing glucoraphanin. MsrB of *L. agilis* R16 and *E. coli* VL8 were identified with their ability to reduce glucoraphanin to glucoerucin. Especially, MsrB of *E. coli* VL8 showed 5 times higher reduction rate compared to MsrB of *L. agilis* R16. This difference also explained the high glucoraphanin degradation rate by *E. coli* VL8 compared to *L. agilis* R16. Under the conditions tested, MsrB was likely to be the responsible enzyme for reduction of glucoraphanin. However, it was also noted that stereochemistry of sulphoxide group in glucoraphanin side chain can alter the role of MsrB in reduction.

MsrB was the very first reductase identified with glucosinolate reducing ability. Identification of this enzyme provided a clearer understanding of the metabolism of methylsulfinylalkyl glucosinolates such as glucoraphanin by human gut bacteria (See Figure 6.2). Future work should have included a GC-MS method to identify ITCs as degradation products as well. In this way, the proposed mechanisms (Figure 6.2) in glucosinolate metabolism could be studied in greater detail.

Many experiments are needed to get a better understanding of the role of MsrB in glucosinolate metabolism. First, reductase activity of MsrB on different substrates can be tested. The reductase activity of MsrB on other methylsulfinylalkyl glucosinolates such as glucoiberin or glucoallysin and ITCs such as sulforaphane, iberin or allysin can be tested. As *E. coli* VL8 was able to reduce sulforaphane to erucin [56], MsrB of *E. coli* VL8 is expected to reduce sulforaphane to erucin as well. It is also likely and should be examined that if MsrB can

reduce sulforaphane nitrile to erucin nitrile. These experiments will enlighten the proposed glucosinolate metabolism by gut bacteria (Figure 6.2). Stereochemistry of sulphoxide group should be taken into account. The role of MsrA on reduction of glucosinolates possessing *S*-epimer of sulphoxide can be tested. Secondly, MsrB can be added to media to favor the degradation of methylsulfinylalkyl glucosinolates such as glucoraphanin. Strains like *L. agilis* R16 which show low glucoraphanin degradation rates (~12% after 24 h) can be tested for myrosinase activity when MsrB is supplemented to the media. If MsrB addition can increase the glucoraphanin degradation rate, further animal model studies or human intervention studies investigating the effect of MsrB on glucosinolate metabolism can be performed. Encapsulation of MsrB or bacteria that constitutively express MrsB might enable the transfer of active MsrB to the gut to increase the glucoraphanin transformation by gut bacteria and enhance the health benefits of ITCs.

Thirdly, MsrB can be investigated for oxidation of methylthioalkyl glucosinolates (such as glucoerucin) to methylsulfinylalkyl glucosinolates (such as glucoraphanin). In addition to its role in reduction of methionine sulphoxide to methionine, MsrA was also found to be a methionine oxidase to catalyse the oxidation of methionine to methionine sulfoxide in proteins and peptides when reducing agent was absent or inaccessible [252]. This phenomenon can be tested for MsrA and MsrB using glucosinolates and ITCs. Interconversion between sulforaphane and erucin (reduced form of sulforaphane) (See Figure 6.2) was reported in mice [115] and in humans [253]. It was suggested that this interconversion can take place post absorption, in tissues [253]. Based on our results, MsrB is likely to be responsible for reduction of glucoraphanin and MsrB is also found in mammals including mice and humans. If interconversion between sulforaphane and erucin is possible in tissues and MsrB can show oxidase activity then MsrB may be the responsible enzyme for this interconversion *in vivo* as well.

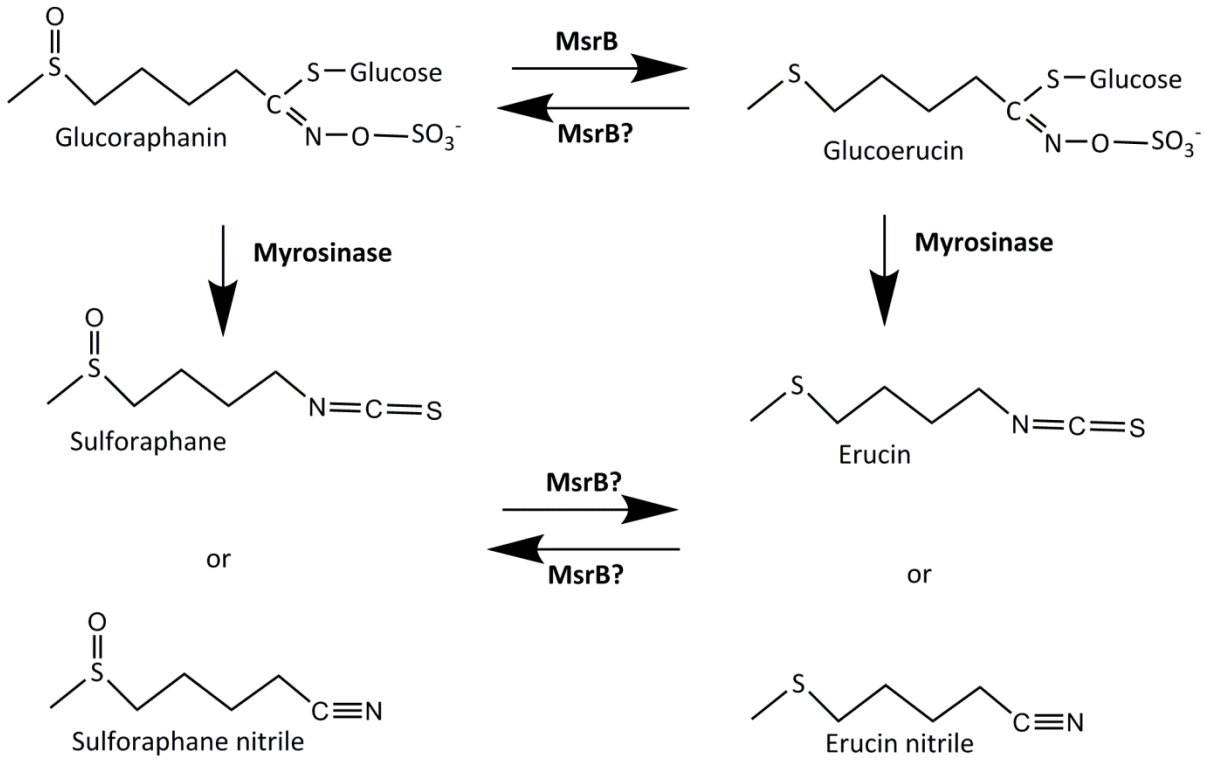


Figure 6.2 Proposed glucosinolate metabolism pathway of methylsulfinylalkyl glucosinolates by human gut bacteria.

In chapter 5, myrosinase of *C.* WYE1 (cmyr) of soil origin was investigated. The myrosinase activity of *C.* WYE1 was characterised using CFE and the myrosinase gene was identified previously by Albaser et al. (2016)[118]. In the current study, this gene was cloned and expressed in *E. coli* with a 6XHis-tag. The protein with the His-tag was purified and further characterised. The same optimum pH, temperature and buffer conditions for myrosinase activity of cmyr were obtained as in the study of Albaser et al. (2016)[118]. When the K_m value of the cmyr for sinigrin was compared to other myrosinases, it was found that cmyr has a higher affinity for sinigrin compared to aphid myrosinase [132, 136] and a similar affinity compared to plant myrosinase [249]. This emphasizes the huge potential of cmyr for myrosinase activity. In addition to sinigrin, cmyr showed a variable affinity to different glucosinolates such as glucoraphanin and progoitrin. Identification of this bacterial myrosinase was an important step for a better understanding of the bacterial myrosinases. There are many questions about cmyr that need to be answered. The effect of ascorbate was studied by Albaser et al. (2016) using CFE of *C.* WYE1 and it was found that ascorbate is an activator for cmyr [118]. This effect was not tested using purified recombinant cmyr. When the effect of many inorganic salts (such as $MgCl_2$ and $FeCl_2$) on myrosinase activity of

Enterobacter cloacae was studied, different inorganic salts were reported to have mostly diverse inhibition effects on myrosinase activity [231]. For instance, only 32% of the original activity was kept when 1 mM $FeCl_2$ was added to the assays [231]. Therefore, the effect of different inorganic salts can be tested. As the human gut also harbors simple sugars as carbon source for bacteria [254], the effect of different sugars (such as glucose or fructose) on myrosinase activity of cmyr should be also investigated.

Even if cmyr is from *C.* WYE1 of soil origin, human gut also harbors *Citrobacter* strains such as *C. freundii* FC50 isolated from human gut in this study. Therefore, structural characterisation of cmyr can answer lots of questions about bacterial myrosinases from human gut. Large amounts of cmyr were obtained by successful cloning and expression to be used for further characterisation. To achieve structural characterisation, the crystallisation of this myrosinase is already being performed in collaboration with Imperial College London.

In conclusion, a better understanding of bacterial myrosinases and reductases involved in glucosinolate metabolism is required to maximise the health benefits of ITCs. Such understanding can enable the identification of a probiotic strain with myrosinase activity from human gut. This would improve the efficiency of consumed glucosinolates for people who are unable to utilise glucosinolates and benefit efficiently from them. Our study focused on identification and characterisation of myrosinases and reductases and achieved the identification of MsrB with glucosinolate reducing ability. In addition, findings about myrosinase and reductase activity of studied strains which were contradictory to previous studies were obtained and new human gut isolates with high myrosinase activities such as *C. freundii* FC50 was identified.

LIST OF REFERENCES

REFERENCES

1. Fahey, J.W., A.T. Zalcmann, and P. Talalay, *The chemical diversity and distribution of glucosinolates and isothiocyanates among plants.* Phytochemistry, 2001. **56**(1): p. 5-51.
2. Verkerk, R. and M. Dekker, *Bioactive Compounds in Foods*, in *Glucosinolates*, J. Gilbert and H.Z. Senyuva, Editors. 2008, Blackwell Publishing: Oxford, UK. p. 31-51.
3. Wittstock, U. and M. Burow, *Glucosinolate breakdown in Arabidopsis: mechanism, regulation and biological significance.* Arabidopsis Book, 2010. **8**: p. e0134.
4. Textor, S. and J. Gershenzon, *Herbivore induction of the glucosinolate-myrosinase defense system: major trends, biochemical bases and ecological significance.* Phytochemistry Reviews, 2009. **8**(1): p. 149-170.
5. Koroleva, O.A., et al., *Identification of a new glucosinolate-rich cell type in Arabidopsis flower stalk.* Plant Physiol, 2000. **124**(2): p. 599-608.
6. Kelly, P.J., A. Bones, and J.T. Rossiter, *Sub-cellular immunolocalization of the glucosinolate sinigrin in seedlings of Brassica juncea.* Planta, 1998. **206**(3): p. 370-7.
7. Koroleva, O.A., et al., *Glucosinolate-accumulating S-cells in Arabidopsis leaves and flower stalks undergo programmed cell death at early stages of differentiation.* Plant J, 2010. **64**(3): p. 456-69.
8. Halkier, B.A. and J. Gershenzon, *Biology and Biochemistry of glucosinolates.* Annual Review of Plant Biology, 2006. **57**: p. 303-333.
9. Bennett, R.N., et al., *Glucosinolate Biosynthesis (Further Characterization of the Aldoxime-Forming Microsomal Monooxygenases in Oilseed Rape Leaves).* Plant Physiol, 1995. **109**(1): p. 299-305.
10. Fahey, J.W., Y. Zhang, and P. Talalay, *Broccoli sprouts: an exceptionally rich source of inducers of enzymes that protect against chemical carcinogens.* Proc Natl Acad Sci U S A, 1997. **94**(19): p. 10367-72.
11. Bellostas, N., J.C. Sorensen, and H. Sorensen, *Qualitative and quantitative evaluation of glucosinolates in cruciferous plantsduring their life cycles.* Agroindustria, 2004. **3**(3): p. 5-10.
12. Rosa, E.A.S. and A.S. Rodrigues, *Total and individual glucosinolate content in 11 broccoli cultivars grown in early and late seasons.* Hortscience, 2001. **36**(1): p. 56-59.
13. Cartea, M.E., et al., *Seasonal variation in glucosinolate content in Brassica oleracea crops grown in northwestern Spain.* Phytochemistry, 2008. **69**(2): p. 403-10.
14. Booth, E.J. and K.C. Walker, *The Effect of Site and Foliar Sulfur on Oilseed Rape - Comparison of Sulfur Responsive and Nonresponsive Seasons.* Phyton-Annales Rei Botanicae, 1992. **32**(3): p. 9-13.
15. Brown, P.D., et al., *Variation of glucosinolate accumulation among different organs and developmental stages of Arabidopsis thaliana.* Phytochemistry, 2003. **62**(3): p. 471-81.

16. Del Carmen Martinez-Ballesta, M., D.A. Moreno, and M. Carvajal, *The physiological importance of glucosinolates on plant response to abiotic stress in Brassica.* Int J Mol Sci, 2013. **14**(6): p. 11607-25.

17. Cartea, M.E. and P. Velasco, *Glucosinolates in Brassica foods: bioavailability in food and singinificance for human health.* Phytochemistry, 2008. **7**: p. 213-229.

18. Borek, V., et al., *Toxicity of rapeseed meal and methyl isothiocyanate to larvae of the black vine weevil.* J. Econ Entomol, 1997: p. 109-112.

19. Lazzeri, L., et al., *Effects of glucosinolates and their enzymatic hydrolysis products via myrosinase on the root-knot nematode Meloidogyne incognita (Kofoid et White) Chitw.* J Agric Food Chem, 2004. **52**(22): p. 6703-7.

20. Giamoustaris, A. and R. Mithen, *The effect of modifying the glucosinolate content of leaves of oilseed rape (Brassica napus ssp. oleifera) on its interaction with specialist and general;ist pests.* Ann appl. Biol, 1995. **126**: p. 347-363.

21. Bones, A.M. and J.T. Rossiter, *The myrosinase-glucosinolate system, its organisation and biochemistry.* Physiologia Plantarum, 1996. **97**: p. 194-208.

22. Bridges, M., et al., *Spatial organization of the glucosinolate-myrosinase system in brassica specialist aphids is similar to that of the host plant.* Proc Biol Sci, 2002. **269**(1487): p. 187-91.

23. Kazana, E., et al., *The cabbage aphid: a walking mustard oil bomb.* Proc Biol Sci, 2007. **274**(1623): p. 2271-7.

24. Holst, B. and G. Williamson, *A critical review of the bioavailability of glucosinolates and related compounds.* Nat Prod Rep, 2004. **21**(3): p. 425-47.

25. Searle, L.M., et al., *The Conversion of 3-Indolylmethylglucosinolate to 3-Indolylacetonitrile by Myrosinase, and Its Relevance to the Clubroot Disease of the Cruciferae.* Journal of Experimental Botany, 1982. **33**(136): p. 935-942.

26. Bartel, B. and G.R. Fink, *Differential regulation of an auxin-producing nitrilase gene family in Arabidopsis thaliana.* Proc Natl Acad Sci U S A, 1994. **91**(14): p. 6649-53.

27. Bak, S., et al., *CYP83B1, a cytochrome P450 at the metabolic branch paint in auxin and indole glucosinolate biosynthesis in Arabidopsis.* Plant Cell, 2001. **13**(1): p. 101-111.

28. Mikkelsen, M.D., P. Naur, and B.A. Halkier, *Arabidopsis mutants in the C-S lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis.* Plant J, 2004. **37**(5): p. 770-7.

29. Redovnikovic, I.R., et al., *Glucosinolates and their potential role in plant.* Periodicum Biologorum, 2008. **110**(4): p. 297-309.

30. Mari, M., et al., *Antifungal vapour-phase activity of allyl-isothiocyanate against Penicillium expansum on pears.* Plant Pathology, 2002. **51**(2): p. 231-236.

31. Smolinska, U., et al., *Isothiocyanates produced by Brassicaceae species as inhibitors of Fusarium oxysporum.* Plant Disease, 2003. **87**(4): p. 407-412.

32. Aires, A., et al., *Initial in vitro evaluations of the antibacterial activities of glucosinolate enzymatic hydrolysis products against plant pathogenic bacteria.* J Appl Microbiol, 2009. **106**(6): p. 2096-105.

33. Aires, A., et al., *The antimicrobial effects of glucosinolates and their respective enzymatic hydrolysis products on bacteria isolated from the human intestinal tract.* J Appl Microbiol, 2009. **106**(6): p. 2086-95.

34. Manici, L.M., L. Lazzeri, and S. Palmieri, *In vitro fungitoxic activity of some glucosinolates and their enzyme-derived products toward plant pathogenic fungi.* Journal of Agricultural and Food Chemistry, 1997. **45**(7): p. 2768-2773.

35. Sarwar, M., et al., *Biofumigation potential of brassicas - III. In vitro toxicity of isothiocyanates to soil-borne fungal pathogens.* Plant and Soil, 1998. **201**(1): p. 103-112.

36. Tierens, K.F., et al., *Study of the role of antimicrobial glucosinolate-derived isothiocyanates in resistance of Arabidopsis to microbial pathogens.* Plant Physiol, 2001. **125**(4): p. 1688-99.

37. Matsuo, M. and M. Yamazaki, *Biosynthesis of sinigrin. VI. Incorporation from homomethionine (2-14C, 15N) and some labelled compounds into sinigrin.* Chem Pharm Bull (Tokyo), 1968. **16**(6): p. 1034-9.

38. Kutacek, M. and M. Kralova, *Biosynthesis of Glucobrassicin Aglycone from C-14 and N-15 Labeled L-Tryptophan Precursors.* Biologia Plantarum, 1972. **14**(4): p. 279-&.

39. Wetter, L.R., *Biosynthesis of Mustard Oil Glucosides .2. The Administration of Sulphur-35 Compounds to Horse-Radish Leaves.* Phytochemistry, 1964. **3**(1): p. 57-64.

40. Rechner, A.R., et al., *Colonic metabolism of dietary polyphenols: influence of structure on microbial fermentation products.* Free Radic Biol Med, 2004. **36**(2): p. 212-25.

41. Sonderby, I.E., F. Geu-Flores, and B.A. Halkier, *Biosynthesis of glucosinolates--gene discovery and beyond.* Trends Plant Sci, 2010. **15**(5): p. 283-90.

42. Geu-Flores, F., et al., *Cytosolic gamma-glutamyl peptidases process glutathione conjugates in the biosynthesis of glucosinolates and camalexin in Arabidopsis.* Plant Cell, 2011. **23**(6): p. 2456-69.

43. Wiesner, M., et al., *Induced production of 1-methoxy-indol-3-ylmethyl glucosinolate by jasmonic acid and methyl jasmonate in sprouts and leaves of pak choi (Brassica rapa ssp. chinensis).* Int J Mol Sci, 2013. **14**(7): p. 14996-5016.

44. Traka, M.H., et al., *Genetic regulation of glucoraphanin accumulation in Beneforte (R) broccoli.* New Phytologist, 2013. **198**(4): p. 1085-1095.

45. Bell, L. and C. Wagstaff, *Glucosinolates, myrosinase hydrolysis products, and flavonols found in rocket (Eruca sativa and Diplotaxis tenuifolia).* J Agric Food Chem, 2014. **62**(20): p. 4481-92.

46. Banerjee, A., et al., *Aliphatic glucosinolate synthesis and gene expression changes in gamma-irradiated cabbage.* Food Chem, 2016. **209**: p. 99-103.

47. Pellegrini, N., et al., *Effect of different cooking methods on color, phytochemical concentration, and antioxidant capacity of raw and frozen brassica vegetables.* J Agric Food Chem, 2010. **58**(7): p. 4310-21.

48. Miller, A.B., et al., *Fruits and vegetables and lung cancer: Findings from the European Prospective Investigation into Cancer and Nutrition.* Int J Cancer, 2004. **108**(2): p. 269-76.

49. Voorrips, L.E., et al., *Vegetable and fruit consumption and risks of colon and rectal cancer in a prospective cohort study: The Netherlands Cohort Study on Diet and Cancer.* Am J Epidemiol, 2000. **152**(11): p. 1081-92.

50. Joseph, M.A., et al., *Cruciferous vegetables, genetic polymorphisms in glutathione S-transferases M1 and T1, and prostate cancer risk.* Nutr Cancer, 2004. **50**(2): p. 206-13.

51. Halkier, B.A. and J. Gershenzon, *Biology and biochemistry of glucosinolates.* Annu Rev Plant Biol, 2006. **57**: p. 303-33.

52. Wittstock, U. and M. Burow, *Tipping the scales - Specifier proteins in glucosinolate hydrolysis.* Iubmb Life, 2007. **59**(12): p. 744-751.

53. Bones, A.M. and J.T. Rossiter, *The enzymic and chemically induced decomposition of glucosinolates.* Phytochemistry, 2006. **67**(11): p. 1053-67.

54. Song, L.J., et al., *Analysis of glucosinolates, isothiocyanates, and amine degradation products in vegetable extracts and blood plasma by LC-MS/MS.* Analytical Biochemistry, 2005. **347**(2): p. 234-243.

55. Mullaney, J.A., et al., *Lactic acid bacteria convert glucosinolates to nitriles efficiently yet differently from enterobacteriaceae.* J Agric Food Chem, 2013. **61**(12): p. 3039-46.

56. Luang-In, V., et al., *The metabolism of methylsulfinylalkyl- and methylthioalkyl-glucosinolates by a selection of human gut bacteria.* Mol Nutr Food Res, 2014. **58**(4): p. 875-83.

57. Luang-In, V., et al., *Glucosinolate and Desulfo-glucosinolate Metabolism by a Selection of Human Gut Bacteria.* Curr Microbiol, 2016. **73**(3): p. 442-51.

58. Tookey, H.L., *Crambe thioglucoside glucohydrolase (EC 3.2.3.1): separation of a protein required for epithiobutane formation.* Can J Biochem, 1973. **51**(12): p. 1654-60.

59. Foo, H.L., et al., *Purification and characterisation of epithiospecifier protein from Brassica napus: enzymic intramolecular sulphur addition within alkenyl thiohydroximates derived from alkenyl glucosinolate hydrolysis.* FEBS Lett, 2000. **468**(2-3): p. 243-6.

60. Bernardi, R., et al., *Isolation of the epithiospecifier protein from oil-rape (Brassica napus ssp. oleifera) seed and its characterization.* FEBS Lett, 2000. **467**(2-3): p. 296-8.

61. Burow, M., et al., *Glucosinolate hydrolysis in Lepidium sativum--identification of the thiocyanate-forming protein.* Plant Mol Biol, 2007. **63**(1): p. 49-61.

62. Kissen, R. and A.M. Bones, *Nitrile-specifier proteins involved in glucosinolate hydrolysis in Arabidopsis thaliana.* J Biol Chem, 2009. **284**(18): p. 12057-70.

63. Matusheski, N.V. and E.H. Jeffery, *Comparison of the bioactivity of two glucoraphanin hydrolysis products found in broccoli, sulforaphane and sulforaphane nitrile.* Journal of Agricultural and Food Chemistry, 2001. **49**(12): p. 5743-5749.

64. Matusheski, N.V., J.A. Juvik, and E.H. Jeffery, *Heating decreases epithiospecifier protein activity and increases sulforaphane formation in broccoli.* Phytochemistry, 2004. **65**(9): p. 1273-81.

65. Chandler, J.D. and B.J. Day, *Thiocyanate: a potentially useful therapeutic agent with host defense and antioxidant properties.* Biochem Pharmacol, 2012. **84**(11): p. 1381-7.

66. Gumz, F., et al., *The crystal structure of the thiocyanate-forming protein from Thlaspi arvense, a kelch protein involved in glucosinolate breakdown.* Plant Mol Biol, 2015. **89**(1-2): p. 67-81.

67. Kuchernig, J.C., et al., *A thiocyanate-forming protein generates multiple products upon allylglucosinolate breakdown in Thlaspi arvense.* Phytochemistry, 2011. **72**(14-15): p. 1699-709.

68. Eisenbrand, G. and H.P. Gelbke, *Assessing the potential impact on the thyroid axis of environmentally relevant food constituents/contaminants in humans.* Arch Toxicol, 2016. **90**(8): p. 1841-57.

69. Walters, D.G., et al., *Cruciferous vegetable consumption alters the metabolism of the dietary carcinogen 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) in humans.* Carcinogenesis, 2004. **25**(9): p. 1659-69.

70. Verhoeven, D.T., et al., *Epidemiological studies on brassica vegetables and cancer risk.* Cancer Epidemiol Biomarkers Prev, 1996. **5**(9): p. 733-48.

71. Cohen, J.H., A.R. Kristal, and J.L. Stanford, *Fruit and vegetable intakes and prostate cancer risk.* J Natl Cancer Inst, 2000. **92**(1): p. 61-8.

72. Seow, A., H. Vainio, and M.C. Yu, *Effect of glutathione-S-transferase polymorphisms on the cancer preventive potential of isothiocyanates: an epidemiological perspective.* Mutat Res, 2005. **592**(1-2): p. 58-67.

73. Meng, X., et al., *A new hypothesis for the cancer mechanism.* Cancer Metastasis Rev, 2012. **31**(1-2): p. 247-68.

74. Jones, S., et al., *Core signaling pathways in human pancreatic cancers revealed by global genomic analyses.* Science, 2008. **321**(5897): p. 1801-6.

75. Knockaert, L., B. Fromenty, and M.A. Robin, *Mechanisms of mitochondrial targeting of cytochrome P450 2E1: physiopathological role in liver injury and obesity.* FEBS J, 2011. **278**(22): p. 4252-60.

76. Fimognari, C. and P. Hrelia, *Sulforaphane as a promising molecule for fighting cancer.* Mutat Res, 2007. **635**(2-3): p. 90-104.

77. Steck, S.E. and J.R. Hebert, *GST polymorphism and excretion of heterocyclic aromatic amine and isothiocyanate metabolites after Brassica consumption.* Environ Mol Mutagen, 2009. **50**(3): p. 238-46.

78. Gupta, P., et al., *Molecular targets of isothiocyanates in cancer: recent advances.* Mol Nutr Food Res, 2014. **58**(8): p. 1685-707.

79. Jancova, P., P. Anzenbacher, and E. Anzenbacherova, *Phase II drug metabolizing enzymes.* Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub, 2010. **154**(2): p. 103-16.

80. Zhang, Y., *Cancer-preventive isothiocyanates: measurement of human exposure and mechanism of action.* Mutat Res, 2004. **555**(1-2): p. 173-90.

81. Novio, S., et al., *Effects of Brassicaceae Isothiocyanates on Prostate Cancer.* Molecules, 2016. **21**(5).

82. Zhang, Y., *Role of glutathione in the accumulation of anticarcinogenic isothiocyanates and their glutathione conjugates by murine hepatoma cells.* Carcinogenesis, 2000. **21**(6): p. 1175-82.

83. Thornalley, P.J., *Isothiocyanates: mechanism of cancer chemopreventive action.* Anticancer Drugs, 2002. **13**(4): p. 331-8.

84. Senanayake, G.V., et al., *The dietary phase 2 protein inducer sulforaphane can normalize the kidney epigenome and improve blood pressure in hypertensive rats.* Am J Hypertens, 2012. **25**(2): p. 229-35.

85. Armah, C.N., et al., *Diet rich in high glucoraphanin broccoli reduces plasma LDL cholesterol: Evidence from randomised controlled trials.* Mol Nutr Food Res, 2015. **59**(5): p. 918-26.

86. Ping, Z., et al., *Sulforaphane protects brains against hypoxic-ischemic injury through induction of Nrf2-dependent phase 2 enzyme.* Brain Res, 2010. **1343**: p. 178-85.

87. Mao, L., et al., *Transcription factor Nrf2 protects the spinal cord from inflammation produced by spinal cord injury.* J Surg Res, 2011. **170**(1): p. e105-15.

88. Chen, G., et al., *Role of the Nrf2-ARE pathway in early brain injury after experimental subarachnoid hemorrhage.* J Neurosci Res, 2011. **89**(4): p. 515-23.

89. Dinkova-Kostova, A.T. and R.V. Kostov, *Glucosinolates and isothiocyanates in health and disease.* Trends Mol Med, 2012. **18**(6): p. 337-47.

90. Mi, L., et al., *The role of protein binding in induction of apoptosis by phenethyl isothiocyanate and sulforaphane in human non-small lung cancer cells.* Cancer Res, 2007. **67**(13): p. 6409-16.

91. Boreddy, S.R. and S.K. Srivastava, *Pancreatic cancer chemoprevention by phytochemicals.* Cancer Lett, 2013. **334**(1): p. 86-94.

92. Herr, I., et al., *Sulforaphane and related mustard oils in focus of cancer prevention and therapy.* Wien Med Wochenschr, 2013. **163**(3-4): p. 80-8.

93. Xu, C., et al., *Suppression of NF-kappaB and NF-kappaB-regulated gene expression by sulforaphane and PEITC through IkappaBalpha, IKK pathway in human prostate cancer PC-3 cells.* Oncogene, 2005. **24**(28): p. 4486-95.

94. Pledgie-Tracy, A., M.D. Sobolewski, and N.E. Davidson, *Sulforaphane induces cell type-specific apoptosis in human breast cancer cell lines.* Mol Cancer Ther, 2007. **6**(3): p. 1013-21.

95. Kanematsu, S., et al., *Sulforaphane inhibits the growth of KPL-1 human breast cancer cells in vitro and suppresses the growth and metastasis of orthotopically transplanted KPL-1 cells in female athymic mice.* Oncol Rep, 2011. **26**(3): p. 603-8.

96. Bhattacharya, A., et al., *Allyl isothiocyanate-rich mustard seed powder inhibits bladder cancer growth and muscle invasion.* Carcinogenesis, 2010. **31**(12): p. 2105-10.

97. Melchini, A., et al., *Antiproliferative activity of the dietary isothiocyanate erucin, a bioactive compound from cruciferous vegetables, on human prostate cancer cells.* Nutr Cancer, 2013. **65**(1): p. 132-8.

98. Wagner, A.E., et al., *Myrosinase-treated glucoerucin is a potent inducer of the Nrf2 target gene heme oxygenase 1--studies in cultured HT-29 cells and mice.* J Nutr Biochem, 2015. **26**(6): p. 661-6.

99. Galuppo, M., et al., *Anti-inflammatory and anti-apoptotic effects of (RS)-glucoraphanin bioactivated with myrosinase in murine sub-acute and acute MPTP-induced Parkinson's disease.* Bioorg Med Chem, 2013. **21**(17): p. 5532-47.

100. Tripathi, M.K. and A.S. Mishra, *Glucosinolates in animal nutrition:A review.* Animal Feed Science and Technology, 2007(132): p. 1-27.

101. Luthy, J., et al., *Goitrin--a nitrosatable constitutent of plant foodstuffs.* Experientia, 1984. **40**(5): p. 452-3.

102. Burel, C., et al., *Dietary low-glucosinolate rapeseed meal affects thyroid status and nutrient utilization in rainbow trout (Oncorhynchus mykiss).* Br J Nutr, 2000. **83**(6): p. 653-64.

103. Tripathi, M.K., et al., *Effect of substitution of groundnut with high glucosinolate mustard (Brassica juncea) meal on nutrient utilization, growth, vital organ weight and blood composition of lambs.* Small Rumin Res, 2001. **39**(3): p. 261-267.

104. Fimognari, C., et al., *Natural isothiocyanates: genotoxic potential versus chemoprevention.* Mutat Res, 2012. **750**(2): p. 107-31.

105. Fernandez-Garcia, E., I. Carvajal-Lerida, and A. Perez-Galvez, *In vitro bioaccessibility assessment as a prediction tool of nutritional efficiency.* Nutr Res, 2009. **29**(11): p. 751-60.

106. Rungapamestry, V., et al., *Effect of cooking brassica vegetables on the subsequent hydrolysis and metabolic fate of glucosinolates.* Proc Nutr Soc, 2007. **66**(1): p. 69-81.

107. Vermeulen, M., et al., *Association between consumption of cruciferous vegetables and condiments and excretion in urine of isothiocyanate mercapturic acids.* J Agric Food Chem, 2006. **54**(15): p. 5350-8.

108. Cai, C., et al., *Effects of industrial pre-freezing processing and freezing handling on glucosinolates and antioxidant attributes in broccoli florets.* Food Chem, 2016. **210**: p. 451-6.

109. Martinez-Hernandez, G.B., et al., *Induced changes in bioactive compounds of kailan-hybrid broccoli after innovative processing and storage.* Journal of Functional Foods, 2013. **5**(1): p. 133-143.

110. Maskell, I. and R. Smithard, *Degradation of glucosinolates during in vitro incubations of rapeseed meal with myrosinase (EC 3.2.3.1) and with pepsin (EC 3.4.23.1)-hydrochloric acid, and contents of porcine small intestine and caecum.* Br J Nutr, 1994. **72**(3): p. 455-66.

111. Michaelsen, S., et al., *Absorption and Degradation of Individual Intact Glucosinolates in the Digestive-Tract of Rodents.* Acta Agriculturae Scandinavica Section a-Animal Science, 1994. **44**(1): p. 25-37.

112. Elfoul, L., et al., *Formation of allyl isothiocyanate from sinigrin in the digestive tract of rats monoassociated with a human colonic strain of Bacteroides thetaiotaomicron.* FEMS Microbiol Lett, 2001. **197**(1): p. 99-103.

113. Zhang, Y., *Molecular mechanism of rapid cellular accumulation of anticarcinogenic isothiocyanates.* Carcinogenesis, 2001. **22**(3): p. 425-31.

114. Freig, A.A.H., L.D. Campbell, and N.E. Stanger, *Fate of Ingested Glucosinolates in Poultry.* Nutrition Reports International, 1987. **36**(6): p. 1337-1345.

115. Bricker, G.V., et al., *Isothiocyanate metabolism, distribution, and interconversion in mice following consumption of thermally processed broccoli sprouts or purified sulforaphane.* Mol Nutr Food Res, 2014. **58**(10): p. 1991-2000.

116. Oliviero, T., et al., *In vivo formation and bioavailability of isothiocyanates from glucosinolates in broccoli as affected by processing conditions.* Mol Nutr Food Res, 2014. **58**(7): p. 1447-56.

117. Li, X. and M.M. Kushad, *Purification and characterization of myrosinase from horseradish (Armoracia rusticana) roots.* Plant Physiol Biochem, 2005. **43**(6): p. 503-11.

118. Albaser, A., et al., *Discovery of a Bacterial Glycoside Hydrolase Family 3 (GH3) beta-Glucosidase with Myrosinase Activity from a Citrobacter Strain Isolated from Soil.* J Agric Food Chem, 2016. **64**(7): p. 1520-7.

119. Burmeister, W.P., et al., *The crystal structures of Sinapis alba myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase.* Structure, 1997. **5**(5): p. 663-75.

120. Lenman, M., et al., *Characterization of a Brassica napus myrosinase pseudogene: myrosinases are members of the BGA family of beta-glycosidases.* Plant Mol Biol, 1993. **21**(3): p. 463-74.

121. Ohtsuru, M. and H. Kawatani, *Studies on the Myrosinase from Wasabia-Japonica - Purification and Some Properties of Wasabi Myrosinase.* Agricultural and Biological Chemistry, 1979. **43**(11): p. 2249-2255.

122. Ketudat Cairns, J.R. and A. Esen, *beta-Glucosidases.* Cell Mol Life Sci, 2010. **67**(20): p. 3389-405.

123. Andersson, D., et al., *Myrosinases from root and leaves of Arabidopsis thaliana have different catalytic properties.* Phytochemistry, 2009. **70**(11-12): p. 1345-54.

124. Burmeister, W.P., et al., *High resolution X-ray crystallography shows that ascorbate is a cofactor for myrosinase and substitutes for the function of the catalytic base.* J Biol Chem, 2000. **275**(50): p. 39385-93.

125. James, D.C. and J.T. Rossiter, *Development and characteristics of myrosinase in Brassica napus during early seedling growth.* Physiologia Plantarum, 1991. **82**: p. 163-170.

126. Bernardi, R., et al., *Isolation and biochemical characterization of a basic myrosinase from ripe Crambe abyssinica seeds, highly specific for epi-progoitrin.* J Agric Food Chem, 2003. **51**(9): p. 2737-44.

127. Hennig, K., et al., *A metabolomics approach to identify factors influencing glucosinolate thermal degradation rates in Brassica vegetables.* Food Chem, 2014. **155**: p. 287-97.

128. Volden, J., et al., *Kinetics of changes in glucosinolate concentrations during long-term cooking of white cabbage (Brassica oleracea L. ssp. capitata f. alba).* J Agric Food Chem, 2008. **56**(6): p. 2068-73.

129. Oliviero, T., et al., *Effect of water content and temperature on inactivation kinetics of myrosinase in broccoli (Brassica oleracea var. italica).* Food Chem, 2014. **163**: p. 197-201.

130. Oerlemans, K., et al., *Thermal degradation of glucosinolates in red cabbage.* Food Chem, 2006. **163**: p. 19-29.

131. Rakariyatham, N., et al., *Screening of filamentous fungi for production of myrosinase.* Brazilian Journal of Microbiology, 2005. **36**(3): p. 242-245.

132. Jones, A.M., et al., *Purification and characterisation of a non-plant myrosinase from the cabbage aphid Brevicoryne brassicae (L.).* Insect Biochem Mol Biol, 2001. **31**(1): p. 1-5.

133. Jones, A.M., et al., *Characterization and evolution of a myrosinase from the cabbage aphid Brevicoryne brassicae.* Insect Biochem Mol Biol, 2002. **32**(3): p. 275-84.

134. Husebye, H., et al., *Crystal structure at 1.1 Angstroms resolution of an insect myrosinase from Brevicoryne brassicae shows its close relationship to beta-glucosidases.* Insect Biochem Mol Biol, 2005. **35**(12): p. 1311-20.

135. Chen, S. and B.A. Halkier, *Functional expression and characterization of the myrosinase MYR1 from Brassica napus in Saccharomyces cerevisiae.* Protein Expr Purif, 1999. **17**(3): p. 414-20.

136. Pontoppidan, B., et al., *Purification and characterization of myrosinase from the cabbage aphid (Brevicoryne brassicae), a brassica herbivore.* European Journal of Biochemistry, 2001. **268**(4): p. 1041-1048.

137. Luang-In, V., *Influence of Human Gut Microbiota on the Metabolic Fate of Glucosinolates*, in *Faculty of Natural Sciences*. 2013, Imperial College London: London.

138. Luang-In, V., et al., *Identification of Proteins Possibly Involved in Glucosinolate Metabolism in L. agilis R16 and E. coli VL8.* Protein J, 2015. **34**(2): p. 135-46.

139. Rahman, M.A., et al., *Cloning, sequencing, and expression of the Escherichia coli peptide methionine sulfoxide reductase gene.* J Biol Chem, 1992. **267**(22): p. 15549-51.

140. Moskovitz, J., et al., *Escherichia coli peptide methionine sulfoxide reductase gene: regulation of expression and role in protecting against oxidative damage.* J Bacteriol, 1995. **177**(3): p. 502-7.

141. Achilli, C., A. Ciana, and G. Minetti, *The discovery of methionine sulfoxide reductase enzymes: An historical account and future perspectives.* Biofactors, 2015. **41**(3): p. 135-52.

142. Alamuri, P. and R.J. Maier, *Methionine sulfoxide reductase in Helicobacter pylori: interaction with methionine-rich proteins and stress-induced expression.* J Bacteriol, 2006. **188**(16): p. 5839-50.

143. Aachmann, F.L., et al., *Structural and biochemical analysis of mammalian methionine sulfoxide reductase B2.* Proteins, 2011. **79**(11): p. 3123-31.

144. Moskovitz, J., et al., *Methionine sulfoxide reductase (MsrA) is a regulator of antioxidant defense and lifespan in mammals.* Proc Natl Acad Sci U S A, 2001. **98**(23): p. 12920-5.

145. Moskovitz, J., et al., *Overexpression of peptide-methionine sulfoxide reductase in Saccharomyces cerevisiae and human T cells provides them with high resistance to oxidative stress.* Proc Natl Acad Sci U S A, 1998. **95**(24): p. 14071-5.

146. Kumar, R.A., et al., *Reaction mechanism, evolutionary analysis, and role of zinc in Drosophila methionine-R-sulfoxide reductase.* J Biol Chem, 2002. **277**(40): p. 37527-35.

147. Romero, H.M., et al., *Investigations into the role of the plastidial peptide methionine sulfoxide reductase in response to oxidative stress in Arabidopsis.* Plant Physiol, 2004. **136**(3): p. 3784-94.

148. Moskovitz, J., et al., *The yeast peptide-methionine sulfoxide reductase functions as an antioxidant in vivo.* Proc Natl Acad Sci U S A, 1997. **94**(18): p. 9585-9.

149. Antoine, M., et al., *Characterization of the amino acids from Neisseria meningitidis MsrA involved in the chemical catalysis of the methionine sulfoxide reduction step.* J Biol Chem, 2006. **281**(51): p. 39062-70.

150. Kim, Y.K., et al., *Structural and kinetic analysis of an MsrA-MsrB fusion protein from Streptococcus pneumoniae.* Mol Microbiol, 2009. **72**(3): p. 699-709.

151. Lowther, W.T., et al., *The mirrored methionine sulfoxide reductases of Neisseria gonorrhoeae pilB.* Nat Struct Biol, 2002. **9**(5): p. 348-52.

152. Vogt, W., *Oxidation of methionyl residues in proteins: tools, targets, and reversal.* Free Radic Biol Med, 1995. **18**(1): p. 93-105.

153. Boschi-Muller, S., A. Gand, and G. Branlant, *The methionine sulfoxide reductases: Catalysis and substrate specificities.* Archives of Biochemistry and Biophysics, 2008. **474**(2): p. 266-273.

154. Ezraty, B., L. Aussel, and F. Barras, *Methionine sulfoxide reductases in prokaryotes.* Biochim Biophys Acta, 2005. **1703**(2): p. 221-9.

155. Lee, B.C. and V.N. Gladyshev, *The biological significance of methionine sulfoxide stereochemistry.* Free Radic Biol Med, 2011. **50**(2): p. 221-7.

156. Ruan, H., et al., *High-quality life extension by the enzyme peptide methionine sulfoxide reductase.* Proc Natl Acad Sci U S A, 2002. **99**(5): p. 2748-53.

157. Ejiri, S.I., H. Weissbach, and N. Brot, *The purification of methionine sulfoxide reductase from Escherichia coli.* Anal Biochem, 1980. **102**(2): p. 393-8.

158. Tete-Favier, F., et al., *Crystal structure of the Escherichia coli peptide methionine sulphoxide reductase at 1.9 A resolution.* Structure, 2000. **8**(11): p. 1167-78.

159. Taylor, A.B., et al., *Structure of Mycobacterium tuberculosis methionine sulfoxide reductase A in complex with protein-bound methionine.* J Bacteriol, 2003. **185**(14): p. 4119-26.

160. Rouhier, N., et al., *Functional and structural aspects of poplar cytosolic and plastidial type a methionine sulfoxide reductases.* J Biol Chem, 2007. **282**(5): p. 3367-78.

161. Lowther, W.T., et al., *Structure and mechanism of peptide methionine sulfoxide reductase, an "anti-oxidation" enzyme.* Biochemistry, 2000. **39**(44): p. 13307-12.

162. Spector, D., et al., *New membrane-associated and soluble peptide methionine sulfoxide reductases in Escherichia coli.* Biochem Biophys Res Commun, 2003. **302**(2): p. 284-9.

163. Weissbach, H., L. Resnick, and N. Brot, *Methionine sulfoxide reductases: history and cellular role in protecting against oxidative damage.* Biochim Biophys Acta, 2005. **1703**(2): p. 203-12.

164. Etienne, F., et al., *Reduction of Sulindac to its active metabolite, sulindac sulfide: assay and role of the methionine sulfoxide reductase system.* Biochem Biophys Res Commun, 2003. **312**(4): p. 1005-10.

165. Saha, S., et al., *Isothiocyanate concentrations and interconversion of sulforaphane to erucin in human subjects after consumption of commercial frozen broccoli compared to fresh broccoli.* Mol Nutr Food Res, 2012. **56**(12): p. 1906-16.

166. Guarner, F. and J.R. Malagelada, *Gut flora in health and disease.* Lancet, 2003. **361**(9356): p. 512-9.

167. Xu, X., et al., *Gut microbiota, host health, and polysaccharides.* Biotechnol Adv, 2013. **31**(2): p. 318-37.

168. Lampe, J.W., *The Human Microbiome Project: getting to the guts of the matter in cancer epidemiology.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(10): p. 2523-4.

169. Goel, A., M. Gupta, and R. Aggarwal, *Gut microbiota and liver disease.* J Gastroenterol Hepatol, 2014. **29**(6): p. 1139-48.

170. Clemente, J.C., et al., *The Impact of the Gut Microbiota on Human Health: An Integrative View.* Cell, 2012. **148**(6): p. 1258-1270.

171. Browne, H.P., et al., *Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation.* Nature, 2016. **533**(7604): p. 543-6.

172. Ley, R.E., D.A. Peterson, and J.I. Gordon, *Ecological and evolutionary forces shaping microbial diversity in the human intestine.* Cell, 2006. **124**(4): p. 837-48.

173. Koenig, J.E., et al., *Succession of microbial consortia in the developing infant gut microbiome.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**: p. 4578-4585.

174. Maslowski, K.M. and C.R. Mackay, *Diet, gut microbiota and immune responses.* Nat Immunol, 2011. **12**(1): p. 5-9.

175. Turnbaugh, P.J., et al., *The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice.* Sci Transl Med, 2009. **1**(6): p. 6ra14.

176. Li, F., et al., *Human gut bacterial communities are altered by addition of cruciferous vegetables to a controlled fruit- and vegetable-free diet.* J Nutr, 2009. **139**(9): p. 1685-91.

177. Benson, A.K., et al., *Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors.* Proc Natl Acad Sci U S A, 2010. **107**(44): p. 18933-8.

178. McKnite, A.M., et al., *Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits.* PLoS One, 2012. **7**(6): p. e39191.

179. Fan, W., et al., *Impact of diet in shaping gut microbiota revealed by a comparative study in infants during the six months of life.* J Microbiol Biotechnol, 2014. **24**(2): p. 133-43.

180. Palmer, C., et al., *Development of the human infant intestinal microbiota.* Plos Biology, 2007. **5**(7): p. 1556-1573.

181. Turnbaugh, P.J., et al., *A core gut microbiome in obese and lean twins.* Nature, 2009. **457**(7228): p. 480-U7.

182. Abreu, M.T., *Toll-like receptor signalling in the intestinal epithelium: how bacterial recognition shapes intestinal function.* Nat Rev Immunol, 2010. **10**(2): p. 131-44.

183. Patterson, J.A. and K.M. Burkholder, *Application of prebiotics and probiotics in poultry production.* Poult Sci, 2003. **82**(4): p. 627-31.

184. Mead, G.C., *Prospects for 'competitive exclusion' treatment to control salmonellas and other foodborne pathogens in poultry.* Vet J, 2000. **159**(2): p. 111-23.

185. Nurmi, E., L. Nuotio, and C. Schneitz, *The competitive exclusion concept: development and future.* Int J Food Microbiol, 1992. **15**(3-4): p. 237-40.

186. Edens, F.W., et al., *Principles of ex ovo competitive exclusion and in ovo administration of Lactobacillus reuteri.* Poult Sci, 1997. **76**(1): p. 179-96.

187. Seksik, P., et al., *Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon.* Gut, 2003. **52**(2): p. 237-42.

188. Morgan, X.C., et al., *Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment.* Genome Biol, 2012. **13**(9): p. R79.

189. Ott, S.J., et al., *Dynamics of the mucosa-associated flora in ulcerative colitis patients during remission and clinical relapse.* J Clin Microbiol, 2008. **46**(10): p. 3510-3.

190. Castellarin, M., et al., *Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma.* Genome Res, 2012. **22**(2): p. 299-306.

191. Ley, R.E., et al., *Obesity alters gut microbial ecology.* Proc Natl Acad Sci U S A, 2005. **102**(31): p. 11070-5.

192. Albenberg, L.G., J.D. Lewis, and G.D. Wu, *Food and the gut microbiota in inflammatory bowel diseases: a critical connection.* Curr Opin Gastroenterol, 2012. **28**(4): p. 314-20.

193. Moore, W.E. and L.H. Moore, *Intestinal floras of populations that have a high risk of colon cancer.* Appl Environ Microbiol, 1995. **61**(9): p. 3202-7.

194. Koppel, N. and E.P. Balskus, *Exploring and Understanding the Biochemical Diversity of the Human Microbiota.* Cell Chem Biol, 2016. **23**(1): p. 18-30.

195. Fraher, M.H., P.W. O'Toole, and E.M. Quigley, *Techniques used to characterize the gut microbiota: a guide for the clinician.* Nat Rev Gastroenterol Hepatol, 2012. **9**(6): p. 312-22.

196. Fritz, J.V., et al., *From meta-omics to causality: experimental models for human microbiome research.* Microbiome, 2013. **1**(1): p. 14.

197. Li, Z., et al., *Pomegranate ellagitannins stimulate growth of gut bacteria in vitro: Implications for prebiotic and metabolic effects.* Anaerobe, 2015. **34**: p. 164-8.

198. Morais, C.A., et al., *Anthocyanins as inflammatory modulators and the role of the gut microbiota.* J Nutr Biochem, 2016. **33**: p. 1-7.

199. Tourlomoussis, P., *The effect of genotype and environment on the composition of the gastroinstestinal microbiota in pigs and chickens as revealed by molecular profiling* in *Faculty of Medical and Veterinary Sciences*. 2009, University of Bristol: Bristol.

200. Lin, H.V., et al., *Butyrate and propionate protect against diet-induced obesity and regulate gut hormones via free fatty acid receptor 3-independent mechanisms.* PLoS One, 2012. **7**(4): p. e35240.

201. Lampe, J.W. and J.L. Chang, *Interindividual differences in phytochemical metabolism and disposition.* Seminars in Cancer Biology, 2007. **17**(5): p. 347-353.

202. Couteau, D., et al., *Isolation and characterization of human colonic bacteria able to hydrolyse chlorogenic acid.* Journal of Applied Microbiology, 2001. **90**(6): p. 873-881.

203. Tsangalis, D., et al., *Enzymic transformation of isoflavone phytoestrogens in soymilk by beta-glucosidase-producing bifidobacteria.* Journal of Food Science, 2002. **67**(8): p. 3104-3113.

204. Hur, H. and F. Rafii, *Biotransformation of the isoflavonoids biochanin A, formononetin, and glycitein by Eubacterium limosum.* FEMS Microbiol Lett, 2000. **192**(1): p. 21-5.

205. Matthies, A., et al., *Conversion of daidzein and genistein by an anaerobic bacterium newly isolated from the mouse intestine.* Appl Environ Microbiol, 2008. **74**(15): p. 4847-52.

206. Rowland, I., et al., *Bioavailability of phyto-oestrogens.* Br J Nutr, 2003. **89 Suppl 1**: p. S45-58.

207. Possemiers, S., et al., *The prenylflavonoid isoxanthohumol from hops (Humulus lupulus L.) is activated into the potent phytoestrogen 8-prenylnaringenin in vitro and in the human instestine.* J. Nutrition, 2006: p. 1862-1867.

208. Tani, N., M. Ohtsuru, and T. Hata, *Studies on Bacterial Myrosinase .1. Isolation of Myrosinase Producing Microorganism.* Agricultural and Biological Chemistry, 1974. **38**(9): p. 1617-1622.

209. Brabban, A.D. and C. Edwards, *Isolation of glucosinolate degrading microorganisms and their potential for reducing the glucosinolate content of rapemeal.* FEMS Microbiol Lett, 1994. **119**(1-2): p. 83-8.

210. Palop, M.L., J.P. Smiths, and B. Brink, *Degradation of sinigrin by Lactobacillus agilis strain R16.* International Journal of Food Microbiology, 1995. **26**: p. 219-229.

211. Combourieu, B., et al., *Identification of new derivatives of sinigrin and glucotropaeolin produced by the human digestive microflora using 1H NMR spectroscopy analysis of in vitro incubations.* Drug Metab Dispos, 2001. **29**(11): p. 1440-5.

212. Krul, C., et al., *Metabolism of sinigrin (2-propenyl glucosinolate) by the human colonic microflora in a dynamic in vitro large-intestinal model.* Carcinogenesis, 2002. **23**(6): p. 1009-16.

213. Rouzaud, G., et al., *Influence of plant and bacterial myrosinase activity on the metabolic fate of glucosinolates in gnotobiotic rats.* Br J Nutr, 2003. **90**(2): p. 395-404.

214. Cheng, D.L., K. Hashimoto, and Y. Uda, *In vitro digestion of sinigrin and glucotropaeolin by single strains of Bifidobacterium and identification of the digestive products.* Food Chem Toxicol, 2004. **42**(3): p. 351-7.

215. Lai, R.H., M.J. Miller, and E. Jeffery, *Glucoraphanin hydrolysis by microbiota in the rat cecum results in sulforaphane absorption.* Food Funct, 2010. **1**(2): p. 161-6.

216. Li, F., et al., *Variation of glucoraphanin metabolism in vivo and ex vivo by human gut bacteria.* Br J Nutr, 2011. **106**(3): p. 408-16.

217. Luciano, F.B., J. Belland, and R.A. Holley, *Microbial and chemical origins of the bactericidal activity of thermally treated yellow mustard powder toward Escherichia coli O157:H7 during dry sausage ripening.* Int J Food Microbiol, 2011. **145**(1): p. 69-76.

218. Herzallah, S., M.L. Lledo, and R. Holley, *Influence of NaCl and NaNO3 on sinigrin hydrolysis by foodborne bacteria.* J Food Prot, 2011. **74**(12): p. 2162-8.

219. Olaimat, A.N., B. Sobhi, and R.A. Holley, *Influence of temperature, glucose, and iron on sinigrin degradation by Salmonella and Listeria monocytogenes.* J Food Prot, 2014. **77**(12): p. 2133-8.

220. Michaelsen, S., et al., *Adsorption and Degradation of Invidual Intact Glucosinolates in the Digestive Tract of Rodents.* Acta Agriculturae Scandinavica, Section A — Animal Science 1994. **44**(1): p. 25-37.

221. Sambrook, J., E.F. Fritsch, and T. Maniatis, *Molecular Cloning*, S. J., F. EF., and M. T., Editors. 1989, Coldspring Harbor Laboratory Press: New York, USA.

222. Baker, G.C., J.J. Smith, and D.A. Cowan, *Review and re-analysis of domain-specific 16S primers.* J Microbiol Methods, 2003. **55**(3): p. 541-55.

223. Joshi, N. and J. Fass, *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files [Software]. Available at https://github.com/najoshi/sickle.* 2011.

224. Simpson, J.T., et al., *ABySS: a parallel assembler for short read sequence data.* Genome Res, 2009. **19**(6): p. 1117-23.

225. Huang, X. and A. Madan, *CAP3: A DNA sequence assembly program.* Genome Res, 1999. **9**(9): p. 868-77.

226. Nurk, S., et al., *Assembling single-cell genomes and mini-metagenomes from chimeric MDA products.* J Comput Biol, 2013. **20**(10): p. 714-37.

227. Rutherford, K., et al., *Artemis: sequence visualization and annotation.* Bioinformatics, 2000. **16**(10): p. 944-5.

228. Overbeek, R., et al., *The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST).* Nucleic Acids Res, 2014. **42**(Database issue): p. D206-14.

229. Treangen, T.J., et al., *The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes.* Genome Biol, 2014. **15**(11): p. 524.

230. Clarke, D.B., *Glucosinolates, structures and analysis in food.* Analytical Methods, 2010. **2**(4): p. 310-325.

231. Tani, N., M. Ohtsuru, and T. Hata, *Studies on Bacterial Myrosinase .2. Purification and General Characteristics of Bacterial Myrosinase Produced by Enterobacter-Cloacae.* Agricultural and Biological Chemistry, 1974. **38**(9): p. 1623-1630.

232. Mullaney, J.A., *The biotransformation of glucosinolates: A bacterial perspective. PhD thesis*. 2013, Massey University: New Zealand.

233. Cantarel, B.L., et al., *The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics.* Nucleic Acids Res, 2009. **37**(Database issue): p. D233-8.

234. Cole, J.R., et al., *Ribosomal Database Project: data and tools for high throughput rRNA analysis.* Nucleic Acids Res, 2014. **42**(Database issue): p. D633-42.

235. Matteotti, C., et al., *New glucosidase activities identified by functional screening of a genomic DNA library from the gut microbiota of the termite Reticulitermes santonensis.* Microbiol Res, 2011. **166**(8): p. 629-42.

236. Ash, A., et al., *Effect of calcium ions on in vitro pellicle formation from parotid and whole saliva.* Colloids Surf B Biointerfaces, 2013. **102**: p. 546-53.

237. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Mol Syst Biol, 2011. **7**: p. 539.

238. Hatziioanou, D., *Discovery and analysis of novel bacteriocins from gut bacteria, PhD thesis*, in *Faculty of Natural Sciences*. 2011, University of East Anglia Norwich.

239. Missiakas, D. and S. Raina, *Protein folding in the bacterial periplasm.* J Bacteriol, 1997. **179**(8): p. 2465-71.

240. Cordeiro, R.P., et al., *Role of glycoside hydrolase genes in sinigrin degradation by E. coli O157:H7.* Int J Food Microbiol, 2015. **205**: p. 105-11.

241. Boschi-Muller, S., A. Gand, and G. Branlant, *The methionine sulfoxide reductases: Catalysis and substrate specificities.* Arch Biochem Biophys, 2008. **474**(2): p. 266-73.

242. Brunell, D., et al., *Studies on the metabolism and biological activity of the epimers of sulindac.* Drug Metab Dispos, 2011. **39**(6): p. 1014-21.

243. Moskovitz, J., et al., *Identification and characterization of a putative active site for peptide methionine sulfoxide reductase (MsrA) and its substrate stereospecificity.* J Biol Chem, 2000. **275**(19): p. 14167-72.

244. Boschi-Muller, S. and G. Branlant, *Methionine sulfoxide reductase: chemistry, substrate binding, recycling process and oxidase activity.* Bioorg Chem, 2014. **57**: p. 222-30.

245. Lin, Z., et al., *Free methionine-(R)-sulfoxide reductase from Escherichia coli reveals a new GAF domain function.* Proc Natl Acad Sci U S A, 2007. **104**(23): p. 9597-602.

246. Fahey, J.W., et al., *Protection of humans by plant glucosinolates: efficiency of conversion of glucosinolates to isothiocyanates by the gastrointestinal microflora.* Cancer Prev Res (Phila), 2012. **5**(4): p. 603-11.

247. Vergara, F., et al., *Determination of the absolute configuration of the glucosinolate methyl sulfoxide group reveals a stereospecific biosynthesis of the side chain.* Phytochemistry, 2008. **69**(15): p. 2737-42.

248. Beran, F., et al., *Phyllotreta striolata flea beetles use host plant defense compounds to create their own glucosinolate-myrosinase system.* Proc Natl Acad Sci U S A, 2014. **111**(20): p. 7349-54.

249. Bhat, R., et al., *Purification and Characterization of a Novel Redox-Regulated Isoform of Myrosinase (beta-Thioglucoside Glucohydrolase) from Lepidium latifolium L.* J Agric Food Chem, 2015. **63**(47): p. 10218-26.

250. Ohtsuru, M., I. Tsuruo, and T. Hata, *Studies on Fungous Myrosinase .I. Production, Purification and Some Characteristics.* Agricultural and Biological Chemistry, 1969. **33**(9): p. 1309-&.

251. Wathelet, J.P., et al., *A recombinant beta-O-glucosidase from Caldocellum saccharolyticum to hydrolyse desulfo-glucosinolates.* Biotechnology Letters, 2001. **23**(6): p. 443-446.

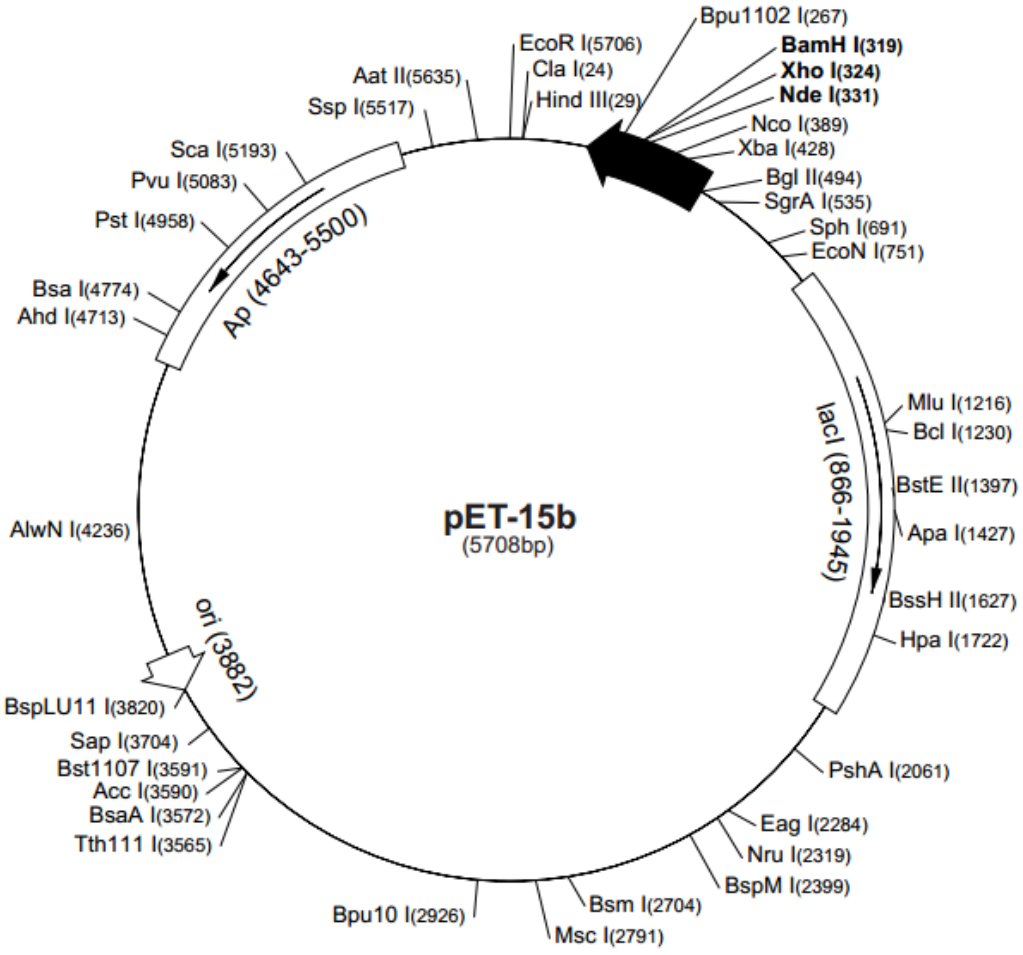
252. Lim, J.C., et al., *Methionine sulfoxide reductase A is a stereospecific methionine oxidase.* Proc Natl Acad Sci U S A, 2011. **108**(26): p. 10472-7.

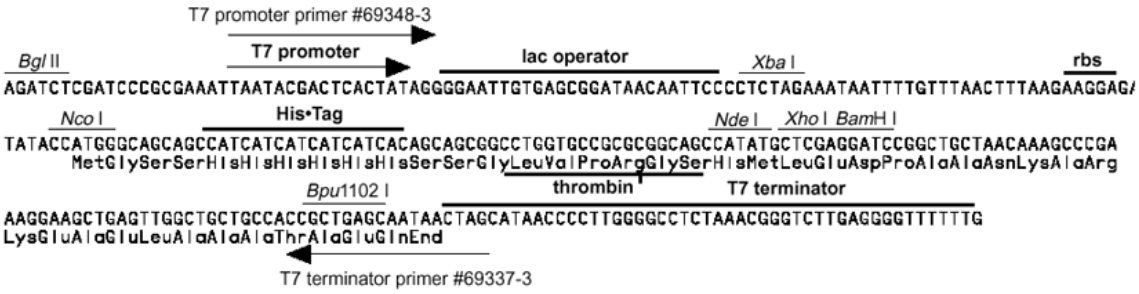
253. Clarke, J.D., et al., *Bioavailability and inter-conversion of sulforaphane and erucin in human subjects consuming broccoli sprouts or broccoli supplement in a cross-over study design.* Pharmacol Res, 2011. **64**(5): p. 456-63.

254. Conlon, M.A. and A.R. Bird, *The Impact of Diet and Lifestyle on Gut Microbiota and Human Health.* Nutrients, 2015. **7**(1): p. 17-44.

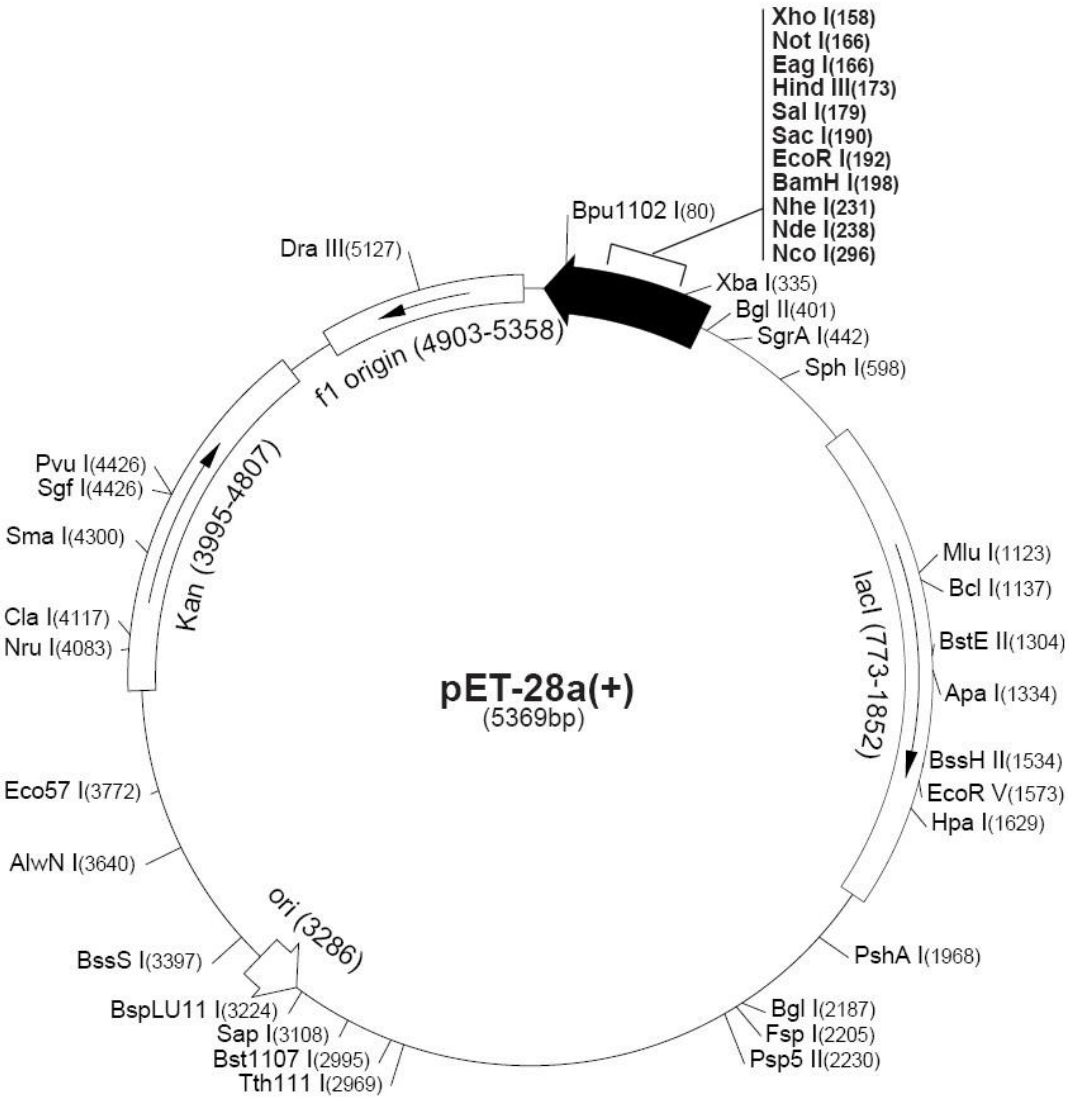
APPENDICES

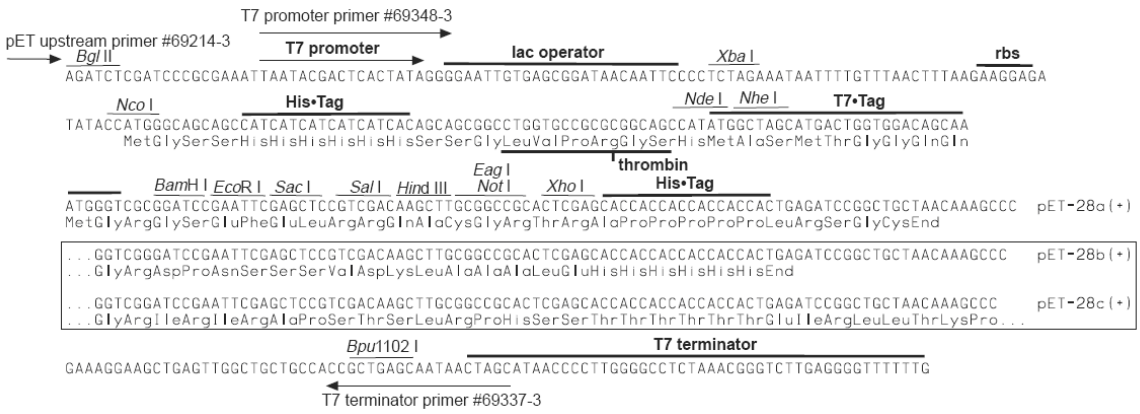
APPENDIX 1 : Maps of the Cloning Vectors





pET-15b cloning/expression region





pET-28a (+) cloning/expression region

APPENDIX 2 : The Nucleotide Sequences of the Cloned Genes

*ecg*4

gtggagaccgtaaaactaaaagaactgattgaccaaatgaccgtagctgaaaaagtgggtcagttagtccaactga
cgccagatttttttgcacaaggaggagaaatcacgggtccgatgcagcaatggcaaatgagtgaagaccaattattt
acgattggttctgtgttgggcactcatactgctgaacaagtatatgccatccaaacgtcttatcttgagaaaagtcgttt
aaagattccccttgtttttatggcagatgttatccatgggtatgaaacgattttcccaattccgcttgccttagcctcttc
ctttgatgcgcagttagtggaagaggttgctaaagcttcagcgaaagaagcggcagaagctggtgtccatgtcacgt
tttcaccgatggcggatcatgtcaaagacgcgcgctggggaagggtcttggaaagtaatggggaagatcccaccttta
tcttctgtattgacggctgcttatgtacgagggtatcaaggaactgcttcccttgcagaaaataaacagcgaattgccg
cttgtgtcaaacactttatcggctatggcgctgcagaaggtggacgggactaccataccgttgatttgtccgatctgga
gctctatcagaattatttgcccgcgtttcaggctgcgattgaagcaggggcacaattagtaatgacctcctttaatacg
atccacggtgttcctgccaccgcaaatcaacccttattgcaggaagtattacgacaaaagctgggattcgacggcttg
atcatctcagattgggcagcggtagcggaattgatggcgcatcgtgtcgctgctgatcgaaaggaagccgcaaaga
aagcttttactgctggcgtggaaatggatatgatgagcgattgctatctgaatgcattggaacaaatcattcatgcag
aacctgaaatgaacgagcagctaaataaggctgtctttcatgtgttaagcttaaaaaataagttgggtctttttgaag
acccttttagaggattagtgcagggcaacttagcaccaaaattagcgcaggaaacacgccacctttcacgagaagc
agcaaagaaaacaacggttcttctgaaaaaccaagcactgttaccgttgcagaaaaagcaaaaggtcgcgttgatc
ggccccaaagcagcttcgaaggatcttttgggggcttggtcgtggatcggcaaaccggaggctgcggtctccttagcg
gagggtctttctcaaaaagaaattgacttaactgtcttatcgtatcctgatggggaaatgctgacgagtacctttattg
atgaggcgtgtaaagttgctaaagaagctgaggttgttttattggctctgggggaaactagcgaggaagcgggaga
agcggcaagtctgaccaaactttctttatcaagaaagcaagaagcattgattgaagcagtaagtcaggtcaatccta
ggatcgcaacgattcttttctgtggcagaccactggttttgaccgccatcgagcctctttgccaagccatcatgattgcc
tggttccctggatcagaaggaggcaatgcgttggcagatttgttgatgggcgagtctgaaccgcaaggacgtttagct
atgagttttcctcgagcagaaggccaattaccgatgacctatgctcaactctctacagggcgtccactgactgaagag
aatagtgatcaaaaatacatttcgcgttatatggatgaaaaaaatgagccattgttttcttttggctcggggatgggct
atggtcgatgcacgatcacccacacccaaatcattaatcctcaggaaaaggatgcaccatttgaggtcagctgcacg
ttgaaaaatgagggagatataccgcataccacgacactgcaactctatagcagagacgatgttacagaagtagcgc
gaccgatgcgggaattaaaacaatggcaaaaaatccatcttgctgcagccgaagaaaggcagcattcattctatgtt
acaccagaggattttgcttatgttcattccgatttctcgacaaaaagtgatcctggtacgatttctttataccttggcttt
gctgctgcctcagcgaagttgattggtacgctcacgatttaa

*ecg*39

atgaaaattgatgtgaatatgagaaaaaatcggctccaatatacggaagcagctgagaagattttgaagcaaatgaccc
taaaagaaaaagtgcacttgatgagcggcaatatggttttagcagatgtaagaagggatatggcaaatgggcagcatta
caacgaatttccttatgaggccggtggaaatgaacgattgaatgttccctccatgaaatttgtagacggtcctcgcggtgc
agtaaccggcagagacaaaacaacttgttttccagtttctatggcaaggggcgcttccttcgacactgagttagaaaagc
aaattggaatagcgattgcgaaagaaatcaaagatagtggtgggaatttttttggcggcgtgtgtattaatttaccttata
atcccggttgggggcgcagtcaagaagtctacggagaagattcattccttctcggtgcgatgggatctgccctttcagaag
gtgttcaatctgaaagtgtggttgcttgcgtcaagcattatgcatttaacagtatggaaaatgctcgttttaaagtcgatgtt
catgctagcaaaagaacagaacgagaaatttatttatcgcattttaaagatgttattgattctggtgcagcttcggtaatg
acttcttataataaatattttggcgaacatactggtaatagtacttatttagttcgagatgtactaaaaaatgaatggaattt
cgacggctttgtcattagtgatttcgttaacggaacgagagacactgtaaaagcagcattggcaggattagatattgaaa
tgcatgtaacaaatcattacggagaaaaattagaaaaagctgtcgaagatggattagtacctgtagaaacaattgacga
tgctgcattaagaattattcgtaccctactcgtctttgaagaagctaggaaaaacgatgattccaaagaaaccattgatta
caaaaaacatatacgtctagctttagatgctgcagaacaatcgattactttgataaaaaatgaaaacaactgtcttccttt
aaacgcaaatattaaaaaagttgctatttttggaaaattagcgactgaagaaaatacaggcgatcgtggttctagtcgag
tttatcctccctatgttatcagttatttagaaggattaaaaaaatactctccacacattcaggttatttataacgaaggatca
gatattgaattagcaaaaacgattgccagagaagcggacgcagcgatttttgttgttggatataattatgatgatgaaggt
gaatatactggtgaagcagaaaacagagaggttcctgctggtgcgatttttgacgctaaaggcggtgatcgaaaagaat
cattagaactccattctgaagatattttactaataaaccaagttggcccagaaaacaataattcagttgttgctttagttgg
tggtaacacaattatgattgaagaatggaaaaatgcagtatctgcgattctctttacctactactcaggaatggaaggtgg
cacagcgctggcaaaaatcctctttggagaagccaatccgagtggaaaactcccatttgtaatcccttacaaagagaaa
gatctgccacagattgattggaatgcagatcaaatcacctatgaatattaccatggatatgcgaagttggaaaaggaagg
aaaagtgccttcgattccttttggttatgggctttcttataccaattttgatatttcaaaccatgattttacaatagataaaaa
gcttttgacagcaaaatgtacagttgaaaacatagggaaactacctggagcagaagttgttcaattatatgtaggattta
gtcactcgcaaatcgatcgtcccatcaaagtattacgcgggttcaagaaagtgcatctacttcctggagaaaaaaagaat
attacgattaaatgtcctatcgaaaaattgaaatggtacaatcctgattctgaacaatgggagctagaagaaattccttac
gaagtctatctaggaaacagcagttcaccaacagatttaataaaagaaactataactgtttaa

*ecg*44

ttggatcataaacaactaaaagaattccccaatgatttttatggggttcagcatcagcagcctatcaagtagaaggcgca
tggcaagaagacggcaaaggcgcttcggtttgggacgattttgtccgcattcccggaaaaaccctttaaagcaactaacgg
agatgtcgcggtcgaccattaccaccgcttcaaagaagatgttgccttgatgaaagaacaaggattgaaaacctatcgtt
tttccattgcttggacacggatttttccagaaggacgaggagaagtcaatcaagcagggttagacttttatttagcattgat
cgatgaactgataaaagcagggatcgaaccgatggtgaccttgtaccattgggatctgccacgagctttgcaagaagaa
tacggtggctgggaatctcgcaaaatcattgaagatttcaccaactatgcagctgtattgtttgaagcctttcgcggcaaa
gtccactattgggtcagcttgaatgaacaaaatatctttacttcgttaggctatcttttagctgcccatccaccaggagtga
ccgatcctaaacggatgtatgaagtcaatcatatcgctaacttggcgaatgcttcggtgatcaacaagttccatgagatga
aaatcccagggaaaatcggaccaagtttcgcctactcaccaaactacccaatcaacagcgacccgaaaaatattttggc
agctgaaaacgcggaagacttgatggcgcattattggttagacgtctatctatggggcgaatacccgattgcggcaatga
attacttgaaagagcaagggatcgcaccaacgatcgaaccaggagatatggacttgctgcgttcagcaaaaccagattt
cttagggatcaactattaccaaacagcaacaaatgcctacaatccattagatggtgtcggtgcggggaaaatgaacacc
actgggaaaaaaggcagctccgaagaaaccgggacaccaggaatgttcaaaaaagccgaaaatccatttgtggaacg
gacgaattgggattgggagatcgatcctcaagggctgcggatcgctttacgacgaatcactagccgttatcgggtgcctgt
gatcattacggaaaacggacttggtgaatacgacaaactaaccgacaatcatcaaatccatgaccaataccgcatcgat
tatttagcagggcatgtccatgccatcaaagaagcgatcagcgacggcgcggaagttcttggctattgcacctggtccttc
actgatttgctaagctggttgaacggctaccaaaaacgctatggctttgtctatgtcgaccaagatgaaacgcaagaagg
atctttggcacgctacaagaaagacagcttctactggtaccaagagctgatcaaaaccaacgggcaagagtgctag

*ecolg*3

atgaaatggctatgttcagtaggaatcgcggtgagtctggccctgcagccagcactggcggatgatttattcggcaacca
tccattaacgcccgaagcgcgggatgcgttcgtcaccgaactgcttaagaaaatgacagttgatgagaaaattggtcag
ctgcgcttaatcagcgtcggcccggataacccgaaagaggcgatccgcgagatgatcaaagacggtcaggttggggcg
attttcaacaccgtaacccgtcaggatatccgcgccatgcaggatcaggtgatggaattaagccgcctgaaaattcctct
tttctttgcttacgacgtgctgcacggtcagcgcacggtgttcccgattagcctcggtctggcctcgtcttttaacctcgatg
cagtgaaaacggtcggacgtgtctctgcttatgaagcggcagatgatggcctgaatatgacctgggcaccgatggtcga
tgtctcgcgcgatccgcgctggggacgtgcttccgaaggttttggcgaagatacgtatctcacctcaacaatgggtaaaa
ccatggtggaagcgatgcagggtaaaagcccggcagatcgctactcggtgatgaccagcgtcaaacactttgccgcata
cggcgcggtagaaggcggtaaagagtacaacaccgtcgatatgagtccgcagcgcctgtttaatgattatatgccgccg
tacaaagcggggctggacgcaggcagcggcgcggtgatggtggcgctgaactcgctgaacggcacgccagccacctcc
gattcctggctgctgaaagatgttctgcgcgaccagtggggctttaaaggcatcaccgtttccgatcacggtgcaatcaa
agagctgattaaacatggcacggcggcagacccggaagatgcggtgcgcgtggcgctgaaatccggaatcaacatga
gtatgagcgacgagtattactcgaagtatctgcctgggttgatcaaatccggcaaagtgacgatggaagagctggacga
tgctgcccgccatgtactgaacgttaaatatgatatgggattgtttaacgacccatacagccatttggggccgaaagagt
ctgacccggtggataccaatgccgaaagccgcctgcaccgtaaagaagcgcgtgaagtggcgcgcgaaagcctggtgt
tgctgaaaaaccgtctcgaaacgttaccgctgaaaaaatcggccaccattgcggtggttgggccactggcggacagtaa
acgtgacgtgatgggcagctggtccgcagccggtgttgccgatcaatccgtgaccgtactgaccgggatcaaaaatgcc
gtcggtgaaaacggtaaagtgctgtatgccaaaggggcgaacgttaccagtgacaaaggcattatcgatttcctgaatc
agtatgaagaagcggtcaaagtcgatccgcgttcgccgcaagagatgattgatgaagcggtgcagacggcgaaacaat
ctgatgtggtggtggctgtagtcggtgaagcacaggggatggcgcacgaagcctccagccggaccgatatcactattcc
gcaaagccaacgtgacttgattgcggcgctgaaagccaccggtaaaccgctggtgctggtgctgatgaacgggcgtccg
ctggcgctggtgaaagaagatcagcaggctgatgcgattctggaaacctggtttgcggggactgaaggcggtaatgca
attgccgatgtattgtttggcgattacaacccgtccggcaagctgccaatgtccttcccgcgttctgtcgggcagatcccg
gtgtactacagccatctgaataccggtcgcccgtataatgccgacaagccgaacaaatacacttcgcgttattttgatga
agctaacggggcgttgtatccgttcggctatgggctgagctacaccactttcaccgtctctgatgtgaaactttctgcgccg
accatgaagcgtgacggcaaagtgactgccagcgtgcaggtgacgaacaccggtaagcgcgaggggccacggtagt
gcagatgtacttgcaggatgtgacggcttccatgagtcgccctgtgaaacagctgaaaggctttgagaaaatcaccctg
aaaccgggcgaaactcagactgtcagcttcccgatcgatattgaggcgctgaagttctggaatcaacagatgaaatatg
acgccgagcctggcaagttcaatgtatttatcggcactgattccgcacgcgttaagaaaggcgagtttgagttgctgtaa

LBAG_ *msr*A1

atgacagagactgcaatatttgctggtggctgtttctggtgcatggtcaaaccgtttgatcaacagcccgggatcaagtc
agtcatttcaggttacactggtggtaccgtcgctaatccgacgtatgaacaagttgccagccacacgactgggcatactg
aagcggttaaaatcacttttgacccagatgtaatcagttatgctgacttggtcgaaatttactggcgtcagactgatccaa
ccgatgcttctggccaatttcaggaccgtggtgacagttatcgaccggtaatttttgttaatagtgaagctcaacgtcgaa
ctgcaactgcttcaagagatgcacttgcggctagcggtaagtttgctgagccgattgtgacaaccattgaagatgcgaag
ccattttatcccgcagaagccgagcaccaagacttttatcgacgcaatccattccgttatcaaatcgaagaaatgggtgg
tcgtgaggcgtttattaagcaacattggcaataa

LBAG_ *msr*A2

atggaaacagcaatttttgcgggtggatgtttttggtgcatggttcagccttttgatagtcaacccggaatcgacagtgttg
tctcgggttatactggtggacacacgaaaaatcccacttacgaagacgtcaaagcgcatacgacgggtcatactgaagc
tgtcaaaattacgtttgatcccgatatcattagttacacggaattaatcaatatttattggcatcaaacggatcctactgat
gcaatggggcaatttcaagatcgtggtgataattatcggccggtcatttttgtgaatggtccggaacaacggcggatcgc
ggaagcgtctaaacgggcattacaggtgagcgaacgatttagtaaaccaattgtaacggcgattgaggatgccaaacc
attttatccagccgaggcacgccaccaacgcttttataaaaataatccagtggtatttgcggaacaagaagcgggtggc
cgcgctgatttcatcgccgagcagtgggccgatgcacctaaagtggactaa

LBAG_ *msr*A3

atgaaaacaaaacctgaagccttaatgcgcgacttatacaacttaatcatcaacccagcgacgcgtgactgggaacgcc
atttactagtccaggccaagaataatcggcaacaactatcgccgactgggcaactaaaacagattgaagctgatttgcg
accattagctatgcggaataacttgactccggatgttatggacttttatcttgccatcactggcaacgcacccatgccgac
taaaaacaatgtatcggcttacccaacccttagttcgccttatgaagaacgggcagtctttgctggtgggtgtttttggtgc
atggttgagccctttgatcagcgtcccggtattaataatgtgatttctggctataccggtggttcgtgggcccaaccgacct
atgaacaagttagtggtcaatatacgggacacgttgaagccgtggaaatcagatacgattcccgtaaaatctcgtatca
aaatcttgtcgatatttattggcaactaattgatccgaccgaccgctttggacagatcaatgatcgcggtagccaataccg
gccagtcatttttttatgctaatcaacatcaattaaccattgcgcaagcatccaaacaagcgcttgtggactcgcaacgtta
tcaaaaaccaattgtcgtggccattgagcccttacagcaattttggccagcagaaaattatcatcaggatttttaccgcaa
acaaccccggcgctaccgacaactgaaaaaagcccatgatcactatctaaattggttgaaattcaaaaacaaattctag

LBAG_ *msr*B

atggataaacagcaaggtgaattacgacagcgtctaacgcctgaacaatacgccgttactcaagaggcagcgaccgaa
cggccatttagcggtgagtatgataacttttaccaggaaggtatttacgttgatgtagttagtggtcaagcactctttagct
cgcgtgacaaatatgatgcgggttgtgggtggccatccttcactaaaccaatcgatcagactaatcttaacgaacaccgt
gacgagtcgtttggcatgcatcgaacagaagtgactagcacacaagcaaactcacatttaggtcacgtctttccagatg
ggccacgagattgcggcgggcttcgttattgtattaattcagcagcactgaagttcattccagttgccgaccttgaaaagg
ctggttatggtcagtatcagagtttatttaaataa

LBAG_*frmsr*

atggctaccacggaaacatcacttatgaaccaacaactcgacgccctattgtttcaagaaactaacctggtcgcgaactt
agctaatgctagtgccctcctcaacagtacttatgacaaccttaactgggccggcttttacttatttaacgaacaaactgg
cgagctcgacttaggtcccttccaaggcaaagtcgcttgcatgcacatcaaagttggggctggtgtcgtcggaactgcttt
tgaaacgcaaacgaaccaacgggtagccgatgtccatcaattccccggccacatcgcttgtgacagtgccagcaactct
gagatcgtggtcccaatcaccaaggacggccaccaaatcggcgtactcgatgtcgactctccatcattagaccgattcaa
tgctgataacgaagctgaattgactgagtttgtggcgattctaacgtcacacattgactaa

VL8_*msr*A

atgagtttatttgataaaaagcatctggtttcccccgccgatgccctgcctggacgtaacaccccgatgcccgtagccacg
ctgcatgcggtcaacggtcactcaatgaccaatgtacctgacggaatggagattgccatttttgcgatgggttgtttctgg
ggtgtggagcgtctgttctggcagttacccggcgtttacagcaccgccgcaggctataccggaggctatacgccaaatcc
gacttatcgggaagtgtgctccggtgatacgggtcatgccgaagcggtacgcattgtttacgatccttccgtcatcagcta
tgagcagttgctacaggtattttgggagaatcacgatcccgcccagggaatgcgtcagggcaatgaccacggcacgca
gtatcgttcagcgatttatccgctgaccccagaacaggatgccgcagctcgcgccagtctggaacgttttcaggcggcga
tgcttgccgccgatgatgatcgtcacatcaccacggaaatcgctaacgccacaccgttttattatgccgaagatgaccac
cagcaatatctgcataaaaacccgtatggttactgtggaattggcggaattggcgtctgtctgccaccggaagcatag

VL8_*msr*B

atggctaataaaccttcggcagaagaactgaaaaaaaatttgtccgagatgcagttctacgtgacgcagaatcatggga
cagagccgccatttacgggtcgtttactgcataacaagcgtgacggcgtatatcactgtttgatctgcgatgccccgctgt
ttcattcccaaaccaagtatgattccggctgtggctggcccagtttctacgaaccggtgagtgaagaatccattcgttata
tcaaagacttgtcacatgggatgcagcgcatagaaattcgttgcggtaactgtgatgctcatctggggcatgtcttccccg
acgggccgcagccaacgggtgaacgttattgtgttaactcagcctctttacgctttaccgatggcgaaaacggcgaaga
aatcaacggttga

FC44_*frmsr*

atgaacaaaacagaattttacgcggatttaaatcgcgactttaacgcgctgatggcgggagaaaccagttttctggcaa
cgcttgcgaacaccagtgcgttgttatatgagcgtctcactgacataaactgggcaggtttttatttgcttgaggacgata
cactggtactcggaccatttcagggcaaaattgcctgtgtccggatacccgtcgggcgcggcgtgtgcggcactgcggtt
gcccgcaatcaagtgcagcgtatcgaggatgttcatgtgtttgacgggcatattgcctgtgatgcggcgagtaattctga
aattgttctgccgctggtggtgaaaaatcagattattggtgttctcgacatcgatagtaccgtcttcggtcgctttacagac
gaggacgagcaaggcttacgtcagcttgtggcacagcttgaaaaagtgcttgcaacgacggattacaaaaaattctttg
cgagcgtcgcaggataa

cmyr

atgctcactgcttttaagatgaatacggtaagtcttgctgttctggtctgccttccgctgtcggtgtctgccagtattcagag
tgctcagcaacccgagctgggttacgataccgtcccgttacttcatttatcaggcctgtcttttaaagatctgaaccgtgac
ggcaaactgaatccctatgaggactggcgactttcaccgcaaacacgggccgccgatttggttaaacgtatgagcgtcg
ctgaaaaagcaggcgtgatgatgcacggtaccgctcctgccgaaggcagcactttcggcaacggtagcgtctatgacag
cgaggctacccgcaaaatggttgtcgacgcgcatgttaacagtctgatcacccgcctgaacggtgaagaaccggccag
actggctgagcagaacaacatgatacaaaagaccgctgagacaacccggctgggtatccccgtcaccatcagcacgga
tccacgtaactcctatcaggctctggtgggcattagcaaccctgccggaaaattcacgcagtggccggaagctattggcc
tgggtgcggccgggagcgaggcgctggcgcaggaatatgcggatcatgtccggcgcgaataccgggcggtggggatca
ccgaagcactttcaccacaagctgacatcacaacggagcctcgctgggcaaggatcagtggaaccttcggagaagacc
ctgagctggctaaaaagcttgtcagaggatatatcactggcatgcaaaaaggggggcagggccttaatccccagagcg
ttgcggcagtcgtcaaacactgggtcggatatggcgcagcggaagacggctgggatggccataatgcctacggtaagc
ataccgtcctcagtaacgagagcctgcagaaacacattatccctttccgcggcgcctttgaggctaacgtggctgcggtc
atgccgacgtattccgtgatgaaaggcgtaacctggaatggcagggagactgagcaggttgctgctggcttcagtcattt
tctgctgaccgaccttctgagaaagcagaacaacttctccggagtcatcatcagcgactggctcatcactaatgactgtg
atgacgaatgcgttaatggtagtgcgccgggaaaaaaaccggttgccgggggtatgccctgggggttgagtctctttct
aaagagcggagatttgttaaggccgttaacgccggtatcgatcagtttggcggcgtgacggattccgctgtgctggttac
tgccgtagaaaaaggacttatcactcaggcacgtctcgacgcctcggttgaacgtattctgcagcagaaatttgagctgg
gtttattcgaacaaccgtacgtggatgcaaagcttgccgaaaaaattgtcggggctcctgacacgaaaaaagcggcgg
atgacgctcagttccgtacgcttgtattgctacagaataagaacattctgccgctaaaacccggcacaaaagtctggcttt
acggagcggataaaagtgccgccgaaaaagcgggtctggaggtggtttctgagccggaagatgctgatgtcgcgttga
tgcgaaccagtgccccccttcgaacaaccgcattacaactacttttttggtcggcgccatcatgaaggttcccttgaatacc
gggaagacaataaagacttcgcggttctgaagcgggtcagtaaacacacacctgtcatcatgacaatgtacatggagc
gacccgccgtcctgaccaacgtgacggataaaacgagcggttttattgccaactttggactcagtgacgaggttttcttca
gcagactgacgtcagatacccccctacacagcgcgccttccattcgcactaccttcctcaatggcgtccgtgcttaaacaa
aaatctgatgagcctgacgatctggacacaccgttattccagcgcggattcggactgacacgctga

APPENDIX 3 : RDP Database Results

short ID, orientation, similarity score, S_ab score, unique common oligomers and sequence full name

Isolate23

domain Bacteria (20)
phylum "Proteobacteria" (20)
class Gammaproteobacteria (20)
order "Enterobacteriales" (20)
family Enterobacteriaceae (20)
genus Escherichia/Shigella (19)

S000727873 not_calculated 1.000 1205 Escherichia coli; IF4; AB272358
S001602706 not_calculated 1.000 1254 uncultured Escherichia sp.; 50L-A1; GQ423062
S002043841 not_calculated 1.000 1239 uncultured bacterium; 16slp92-2h09.w2k; GQ159504

Isolate32

domain Bacteria (20)
phylum "Proteobacteria" (20)
class Gammaproteobacteria (20)
order "Enterobacteriales" (20)
family Enterobacteriaceae (20)
genus Escherichia/Shigella (19)

S000727873 not_calculated 1.000 1205 Escherichia coli; IF4; AB272358
S001602706 not_calculated 1.000 1254 uncultured Escherichia sp.; 50L-A1; GQ423062
S002043841 not_calculated 1.000 1239 uncultured bacterium; 16slp92-2h09.w2k; GQ159504

S002234579 not_calculated 0.999 1228 Escherichia fergusonii; KLU01; HQ214033

Isolate33

domain Bacteria (20)
phylum "Proteobacteria" (20)
class Gammaproteobacteria (20)
order "Enterobacteriales" (20)
family Enterobacteriaceae (20)
genus Escherichia/Shigella (20)

S000013935 not_calculated 1.000 1405 Escherichia/Shigella flexneri (T); X96963
S000021941 not_calculated 1.000 1405 Shigella sonnei; X96964
S000139289 not_calculated 1.000 1384 Escherichia/Shigella fergusonii (T); ATCC 35469; AF530475
S000389727 not_calculated 1.000 1418 Escherichia coli; AF2334513

Isolate35

domain Bacteria (20)
phylum "Proteobacteria" (20)
class Gammaproteobacteria (20)
order "Enterobacteriales" (20)
family Enterobacteriaceae (20)
genus Escherichia/Shigella (19)

S000727873 not_calculated 1.000 1205 Escherichia coli; IF4; AB272358
S001602706 not_calculated 1.000 1254 uncultured Escherichia sp.; 50L-A1; GQ423062
S002043841 not_calculated 0.999 1239 uncultured bacterium; 16slp92-2h09.w2k; GQ159504
S002234579 not_calculated 0.999 1228 Escherichia fergusonii; KLU01; HQ214033

Isolate 36

domain Bacteria (20)
phylum "Proteobacteria" (20)
class Gammaproteobacteria (20)
order "Enterobacteriales" (20)
family Enterobacteriaceae (20)
genus Escherichia/Shigella (20)

S000013935 not_calculated 1.000 1405 Escherichia/Shigella flexneri (T); X96963
S000021941 not_calculated 1.000 1405 Shigella sonnei; X96964
S000139289 not_calculated 1.000 1384 Escherichia/Shigella fergusonii (T); ATCC 35469; AF530475

S000250140 not_calculated 1.000 1357 Photorhabdus luminescens; AY444555

Isolate 48

domain Bacteria (20)
phylum "Proteobacteria" (20)
class Gammaproteobacteria (20)
order "Enterobacteriales" (20)
family Enterobacteriaceae (20)
genus Escherichia/Shigella (20)

S000013935 not_calculated 1.000 1405 Escherichia/Shigella flexneri (T); X96963
S000021941 not_calculated 1.000 1405 Shigella sonnei; X96964
S000139289 not_calculated 1.000 1384 Escherichia/Shigella fergusonii (T); ATCC 35469; AF530475
S000344063 not_calculated 1.000 1376 uncultured bacterium; p-229-o5; AF371848

Isolate 49

domain Bacteria (20)
phylum "Proteobacteria" (20)
class Gammaproteobacteria (20)
order "Enterobacteriales" (20)
family Enterobacteriaceae (20)
genus Escherichia/Shigella (20)

S000727873 not_calculated 0.994 1205 Escherichia coli; IF4; AB272358
S001602706 not_calculated 0.994 1254 uncultured Escherichia sp.; 50L-A1; GQ423062
S002043841 not_calculated 1.000 1239 uncultured bacterium; 16slp92-2h09.w2k; GQ159504
S002043844 not_calculated 0.995 1314 uncultured bacterium; 16slp92-2d12.w2k; GQ159507

Isolate50

domain Bacteria (20)
phylum "Proteobacteria" (20)
class Gammaproteobacteria (20)
order "Enterobacteriales" (20)
family Enterobacteriaceae (20)
genus Citrobacter (18)

S002961729 not_calculated 0.946 1146 Citrobacter freundii; AP9; JN797483
S003287113 not_calculated 0.944 1243 Citrobacter freundii; GTC 07890; AB741667
S003713527 not_calculated 0.936 1211 Citrobacter sp. DL4.2; JQ912532
S003715763 not_calculated 0.952 1205 Citrobacter sp. P005; KC252743

Isolate 57

domain Bacteria (20)
phylum "Proteobacteria" (20)
class Gammaproteobacteria (20)
order "Enterobacteriales" (20)
family Enterobacteriaceae (20)
genus Escherichia/Shigella (19)

S000727873 not_calculated 0.994 1205 Escherichia coli; IF4; AB272358
S001602706 not_calculated 0.994 1254 uncultured Escherichia sp.; 50L-A1; GQ423062
S002043841 not_calculated 0.994 1239 uncultured bacterium; 16slp92-2h09.w2k; GQ159504
S002234579 not_calculated 0.991 1228 Escherichia fergusonii; KLU01; HQ214033

Isolate58

domain Bacteria (20)
phylum "Proteobacteria" (20)
class Gammaproteobacteria (20)
order "Enterobacteriales" (20)
family Enterobacteriaceae (20)
genus Escherichia/Shigella (20)

S000727873 not_calculated 1.000 1205 Escherichia coli; IF4; AB272358
S001265276 not_calculated 1.000 1284 Shigella sp. 5; FJ594947
S001602706 not_calculated 1.000 1254 uncultured Escherichia sp.; 50L-A1; GQ423062
S002015199 not_calculated 1.000 1283 uncultured bacterium; HERMI16; GQ160461

Isolate59

domain Bacteria (20)
phylum Firmicutes (20)
class Bacilli (20)
order Lactobacillales (20)
family Enterococcaceae (20)
genus Enterococcus (20)

S000012626 not_calculated 1.000 1429 Enterococcus gallinarum (T); AF039900
S000387905 not_calculated 1.000 1432 Eubacterium sp. 1275b; AF135452
S000615647 not_calculated 1.000 1411 uncultured Enterococcus sp.; F25; DQ232852
S000709713 not_calculated 1.000 1411 uncultured bacterium; aaa53b05; DQ818102

Isolate 60

domain Bacteria (20)
phylum "Proteobacteria" (20)
class Gammaproteobacteria (20)
order "Enterobacteriales" (20)
family Enterobacteriaceae (20)
genus Escherichia/Shigella (20)

S003284942 not_calculated 0.995 1292 Escherichia coli; b4; JQ661111
S003284967 not_calculated 0.995 1295 Escherichia coli; p22; JQ661136

S003320650 not_calculated 0.995 1297 uncultured bacterium; OTU023_Xeno_Clone03; AB694228
S003320651 not_calculated 0.995 1294 uncultured bacterium; OTU023_Xeno_Clone04; AB694229

Isolate 61

domain Bacteria (20)
phylum Firmicutes (20)
class Bacilli (20)
order Lactobacillales (20)
family Enterococcaceae (20)
genus Enterococcus (20)

S000012626 not_calculated 0.997 1429 Enterococcus gallinarum (T); AF039900
S000387905 not_calculated 0.997 1432 Eubacterium sp. 1275b; AF135452
S000390547 not_calculated 0.997 1435 Enterococcus gallinarum; MG25; AF277567
S000615647 not_calculated 0.997 1411 uncultured Enterococcus sp.; F25; DQ232852

Isolate 66

domain Bacteria (20)
phylum Firmicutes (20)
class Bacilli (20)
order Lactobacillales (20)
family Enterococcaceae (20)
genus Enterococcus (20)

S000005780 not_calculated 0.999 1429 Enterococcus casseliflavus (T); AF039903
S000650499 not_calculated 0.999 1404 Enterococcus casseliflavus; F00240; DQ395282
S000950598 not_calculated 0.999 1403 uncultured bacterium; 3; EU124822

Isolate 71

domain Bacteria (20)
phylum Firmicutes (20)
class Bacilli (20)
order Lactobacillales (20)
family Enterococcaceae (20)
genus Enterococcus (20)

S000005780 not_calculated 0.989 1429 Enterococcus casseliflavus (T); AF039903
S000650499 not_calculated 0.989 1404 Enterococcus casseliflavus; F00240; DQ395282solate71
S001187491 not_calculated 0.989 1427 Enterococcus casseliflavus; F32; EU151766
S001418145 not_calculated 0.989 1397 Enterococcus gallinarum; IMAU10084; WZ30-2; FJ915740

APPENDIX 4 : Mascot results for FPLC

The Score stands for an overall protein score. This number reflects the combined scores of all observed mass spectra that can be attached to aminoacid sequences within a protein. The mass means the expected mass of the protein given. Number of matches indicates the number of MS/MS spectra that were matched to that protein (The significance threshold < 5%). The EmPAI indicates exponentially modified protein abundance index.

Score	Mass	Number of matches	Number of significant matches	Number of sequences	Number of significant sequences	emPAI	Description
688	41264	93	45	18	15	4.44	Phosphoglycerate kinase OS=Escherichia coli (strain K12) GN=pgk PE=1 SV=2
430	60483	84	31	23	16	2.04	Periplasmic dipeptide transport protein OS=Escherichia coli (strain K12) GN=dppA PE=1 SV=1
379	35398	49	21	19	11	1.93	Transaldolase OS=Escherichia coli O104:H4 (strain 2009EL-2071) GN=tal PE=1 SV=1
316	35865	62	26	26	16	3.5	Transaldolase A OS=Escherichia coli (strain K12) GN=talA PE=3 SV=1
265	38590	38	19	15	10	2.16	Protein tas OS=Escherichia coli (strain K12) GN=tas PE=1 SV=1
258	60975	57	20	22	13	1.09	Periplasmic oligopeptide-binding protein OS=Escherichia coli (strain K12) GN=oppA PE=1 SV=2
251	48532	39	17	19	10	1.06	sn-glycerol-3-phosphate-binding periplasmic protein UgpB OS=Escherichia coli (strain K12) GN=ugpB PE=1 SV=1
249	27126	51	21	16	11	3.03	Triosephosphate isomerase OS=Escherichia coli (strain K12) GN=tpiA PE=1 SV=1
204	51563	53	20	21	14	1.53	6-phosphogluconate dehydrogenase, decarboxylating OS=Escherichia coli (strain K12) GN=gnd PE=1 SV=2
199	42928	31	14	14	9	1.1	3-oxoacyl-[acyl-carrier-protein] synthase 1 OS=Escherichia coli (strain K12) GN=fabB PE=1 SV=1

Score	Mass	Number of matches	Number of significant matches	Number of sequences	Number of significant sequences	emPAI	Description
178	45996	28	17	12	10	1.14	Glucose-1-phosphatase OS=Escherichia coli (strain K12) GN=agp PE=1 SV=1
164	31639	27	14	13	8	1.46	Uncharacterized oxidoreductase YghA OS=Escherichia coli (strain K12) GN=yghA PE=1 SV=1
150	59863	40	13	21	9	0.71	Phosphoenolpyruvate carboxykinase [ATP] OS=Escherichia coli O6:K15:H31 (strain 536 / UPEC) GN=pckA PE=3 SV=1
139	57003	20	9	6	5	0.4	Cytoplasmic alpha-amylase OS=Escherichia coli (strain K12) GN=amyA PE=3 SV=3
107	23947	19	9	11	5	1.19	Probable phospholipid-binding protein MlaC OS=Escherichia coli (strain K12) GN=mlaC PE=1 SV=1
104	37238	27	11	12	8	1.16	Putative amino-acid ABC transporter-binding protein YhdW OS=Escherichia coli O157:H7 GN=yhdW PE=3 SV=1
104	60900	29	7	13	6	0.37	Protein UshA OS=Escherichia coli (strain K12) GN=ushA PE=1 SV=2
53	37292	22	5	9	4	0.53	Outer membrane protein A OS=Escherichia coli (strain K12) GN=ompA PE=1 SV=1
102	61605	37	10	18	6	0.37	Glucose-6-phosphate isomerase OS=Escherichia coli (strain K12) GN=pgi PE=1 SV=1
101	23083	18	7	7	4	0.72	Superoxide dismutase [Mn] OS=Escherichia coli (strain K12) GN=sodA PE=1 SV=2
101	13660	8	3	4	2	0.56	Enamine/imine deaminase OS=Escherichia coli (strain K12) GN=ridA PE=1 SV=2
97	21061	25	8	10	4	1.09	Osmotically-inducible protein Y OS=Escherichia coli (strain K12) GN=osmY PE=1 SV=1

Score	Mass	Number of matches	Number of significant matches	Number of sequences	Number of significant sequences	emPAI	Description
91	53189	17	8	10	4	0.27	Glutamate decarboxylase beta OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=gadB PE=3 SV=2
82	53235	12	8	8	5	0.35	Glutamate decarboxylase alpha OS=Escherichia coli O157:H7 GN=gadA PE=3 SV=1
91	57605	26	6	15	5	0.32	Probable D,D-dipeptide-binding periplasmic protein DdpA OS=Escherichia coli (strain K12) GN=ddpA PE=2 SV=1
80	53357	22	7	10	6	0.43	N-succinylglutamate 5-semialdehyde dehydrogenase OS=Escherichia coli (strain SE11) GN=astD PE=3 SV=1
77	43980	19	7	9	6	0.54	Succinylornithine transaminase OS=Escherichia coli (strain ATCC 8739 / DSM 1576 / Crooks) GN=astC PE=3 SV=1
75	22441	16	7	6	4	1.01	KHG/KDPG aldolase OS=Escherichia coli (strain K12) GN=eda PE=1 SV=1
73	56462	24	7	11	6	0.4	Glutathione-binding protein GsiB OS=Escherichia coli O6:K15:H31 (strain 536 / UPEC) GN=gsiB PE=3 SV=2
69	98015	25	5	15	5	0.18	Aconitate hydratase 1 OS=Escherichia coli (strain K12) GN=acnA PE=1 SV=3
62	34525	9	4	5	2	0.2	Cysteine synthase A OS=Escherichia coli (strain K12) GN=cysK PE=1 SV=2
61	31300	10	2	5	1	0.11	Molecular chaperone Hsp31 and glyoxalase 3 OS=Escherichia coli O1:K1 / APEC GN=hchA PE=3 SV=1
61	63825	19	7	10	4	0.29	Periplasmic trehalase OS=Escherichia coli (strain ATCC 8739 / DSM 1576 / Crooks) GN=treA PE=3 SV=1
55	21310	14	4	5	1	0.16	Superoxide dismutase [Fe] OS=Escherichia coli (strain K12) GN=sodB PE=1 SV=2

Score	Mass	Number of matches	Number of significant matches	Number of sequences	Number of significant sequences	emPAI	Description
54	8375	9	4	4	3	3.09	Major outer membrane lipoprotein Lpp OS=Escherichia coli (strain K12) GN=lpp PE=1 SV=1
50	26990	5	3	3	2	0.26	7-alpha-hydroxysteroid dehydrogenase OS=Escherichia coli (strain K12) GN=hdhA PE=1 SV=1
49	46070	33	5	13	4	0.32	Isocitrate dehydrogenase [NADP] OS=Escherichia coli (strain K12) GN=icd PE=1 SV=1
47	47484	12	3	7	3	0.22	Periplasmic AppA protein OS=Escherichia coli (strain K12) GN=appA PE=1 SV=2
47	40928	20	8	11	5	0.59	Putrescine-binding periplasmic protein OS=Escherichia coli (strain K12) GN=potF PE=1 SV=3
45	38395	13	4	7	3	0.28	Aldose 1-epimerase OS=Escherichia coli (strain K12) GN=galM PE=1 SV=1
41	42553	12	3	6	1	0.16	Putative ABC transporter periplasmic-binding protein YdcS OS=Escherichia coli (strain K12) GN=ydcS PE=3 SV=1
40	33096	4	1	2	1	0.1	Glutaminase OS=Escherichia coli O81 (strain ED1a) GN=glsA PE=3 SV=1
40	70902	23	4	14	3	0.2	2',3'-cyclic-nucleotide 2'-phosphodiesterase/3'-nucleotidase OS=Escherichia coli (strain K12) GN=cpdB PE=1 SV=2
39	13785	3	2	2	1	0.25	Protein GlcG OS=Escherichia coli (strain K12) GN=glcG PE=3 SV=1
38	36571	9	1	5	1	0.09	6-phosphogluconolactonase OS=Escherichia coli O6:K15:H31 (strain 536 / UPEC) GN=pgl PE=3 SV=1

Score	Mass	Number of matches	Number of significant matches	Number of sequences	Number of significant sequences	emPAI	Description
35	20900	2	1	1	1	0.16	Protein YceI OS=Escherichia coli (strain K12) GN=yceI PE=1 SV=1
33	49392	10	1	5	1	0.07	N-succinylarginine dihydrolase OS=Escherichia coli (strain SE11) GN=astB PE=3 SV=1
32	14678	5	1	3	1	0.23	Protein YhfA OS=Escherichia coli (strain K12) GN=yhfA PE=1 SV=1
28	40874	18	1	8	1	0.08	Glycerophosphoryl diester phosphodiesterase OS=Escherichia coli (strain K12) GN=glpQ PE=1 SV=2
28	33513	8	1	5	1	0.1	Glutamate/aspartate periplasmic-binding protein OS=Escherichia coli (strain K12) GN=gltI PE=1 SV=2
28	9529	2	1	1	1	0.36	DNA-binding protein HU-alpha OS=Escherichia coli (strain ATCC 9637 / CCM 2024 / DSM 1116 / NCIMB 8666 / NRRL B-766 / W) GN=hupA PE=1 SV=1
27	39223	7	3	4	2	0.18	Leu/Ile/Val-binding protein OS=Escherichia coli (strain K12) GN=livJ PE=1 SV=1
27	43831	11	3	7	3	0.24	Aspartate aminotransferase OS=Escherichia coli (strain K12) GN=aspC PE=1 SV=1
27	13939	2	1	1	1	0.24	Aspartate 1-decarboxylase OS=Escherichia coli (strain K12) GN=panD PE=1 SV=1
23	149076	6	2	2	2	0.04	DNA translocase FtsK OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=ftsK PE=3 SV=1
23	23867	3	1	3	1	0.14	Uncharacterized protein YcgM OS=Escherichia coli (strain K12) GN=ycgM PE=1 SV=1
23	49084	13	4	5	3	0.22	Glutathione reductase OS=Escherichia coli (strain K12) GN=gor PE=1 SV=1

Score	Mass	Number of matches	Number of significant matches	Number of sequences	Number of significant sequences	emPAI	Description
23	26161	11	1	7	1	0.13	Purine nucleoside phosphorylase DeoD-type OS=Escherichia coli (strain K12) GN=deoD PE=1 SV=2
23	34437	5	1	3	1	0.1	ABC transporter periplasmic-binding protein YtfQ OS=Escherichia coli (strain K12) GN=ytfQ PE=1 SV=1
23	41172	2	1	1	1	0.08	Mannitol-1-phosphate 5-dehydrogenase OS=Escherichia coli (strain ATCC 8739 / DSM 1576 / Crooks) GN=mtlD PE=3 SV=1
23	35681	5	1	3	1	0.09	Glyceraldehyde-3-phosphate dehydrogenase A OS=Escherichia coli (strain K12) GN=gapA PE=1 SV=2
22	69130	2	1	1	1	0.05	Chaperone protein DnaK OS=Escherichia coli (strain K12) GN=dnaK PE=1 SV=2
22	74349	2	1	1	1	0.04	Glycogen debranching enzyme OS=Escherichia coli O6:K15:H31 (strain 536 / UPEC) GN=glgX PE=3 SV=1
21	20418	1	1	1	1	0.16	Peptidyl-prolyl cis-trans isomerase A OS=Escherichia coli (strain K12) GN=ppiA PE=1 SV=1
21	4864	2	1	1	1	0.75	Entericidin B OS=Escherichia coli (strain K12) GN=ecnB PE=3 SV=1
21	32502	8	1	4	1	0.1	Malate dehydrogenase OS=Escherichia coli O6:K15:H31 (strain 536 / UPEC) GN=mdh PE=3 SV=1
20	12799	2	2	1	1	0.27	Protein YchN OS=Escherichia coli (strain K12) GN=ychN PE=1 SV=1
20	40614	2	1	1	1	0.08	Macrolide export protein MacA OS=Escherichia coli O157:H7 GN=macA PE=3 SV=1
20	102958	3	1	3	1	0.03	Signal transduction histidine-protein kinase BarA OS=Escherichia coli (strain K12) GN=barA PE=1 SV=1

Score	Mass	Number of matches	Number of significant matches	Number of sequences	Number of significant sequences	emPAI	Description
20	20554	2	1	1	1	0.16	Isochorismatase family protein YecD OS=Escherichia coli (strain K12) GN=yecD PE=1 SV=2
18	33695	6	1	4	1	0.1	Uncharacterized protein YcjY OS=Escherichia coli (strain K12) GN=ycjY PE=4 SV=2
18	23947	2	2	2	2	0.3	Oxygen-insensitive NAD(P)H nitroreductase OS=Escherichia coli (strain K12) GN=nfnB PE=1 SV=1
18	28539	2	1	2	1	0.12	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase OS=Escherichia coli O6:K15:H31 (strain 536 / UPEC) GN=gpmA PE=3 SV=1
17	49361	2	1	1	1	0.07	Xanthine permease XanQ OS=Escherichia coli (strain K12) GN=xanQ PE=1 SV=2
17	31535	5	1	3	1	0.11	4-hydroxy-tetrahydrodipicolinate synthase OS=Escherichia coli (strain K12) GN=dapA PE=1 SV=1
17	45410	2	2	1	1	0.07	Putative prophage CPZ-55 integrase OS=Escherichia coli (strain K12) GN=intZ PE=3 SV=2
17	35575	4	1	2	1	0.09	L-arabinose-binding periplasmic protein OS=Escherichia coli (strain K12) GN=araF PE=1 SV=2
16	21219	1	1	1	1	0.16	Thermosensitive gluconokinase OS=Escherichia coli (strain K12) GN=idnK PE=3 SV=1
16	32822	2	1	2	1	0.1	NADPH-dependent 7-cyano-7-deazaguanine reductase OS=Escherichia coli O1:K1 / APEC GN=queF PE=3 SV=1
15	37049	1	1	1	1	0.09	Selenide, water dikinase OS=Escherichia coli O6:K15:H31 (strain 536 / UPEC) GN=selD PE=3 SV=1
15	53110	1	1	1	1	0.06	Cytosol non-specific dipeptidase OS=Escherichia coli (strain K12) GN=pepD PE=1 SV=3

Score	Mass	Number of matches	Number of significant matches	Number of sequences	Number of significant sequences	emPAI	Description
14	46353	3	2	2	1	0.07	High-affinity branched-chain amino acid transport system permease protein LivM OS=Escherichia coli (strain K12) GN=livM PE=3 SV=2
14	55242	2	1	1	1	0.06	Probable glutamate/gamma-aminobutyrate antiporter OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=gadC PE=3 SV=1
14	25564	2	1	1	1	0.13	Ribonuclease PH OS=Escherichia coli (strain K12 / MC4100 / BW2952) GN=rph PE=1 SV=1
14	36997	1	1	1	1	0.09	Starvation-sensing protein RspB OS=Escherichia coli (strain K12) GN=rspB PE=3 SV=1
14	36570	5	1	1	1	0.09	HTH-type transcriptional regulator GntR OS=Escherichia coli (strain K12) GN=gntR PE=1 SV=1
14	35690	6	1	4	1	0.09	D-galactose-binding periplasmic protein OS=Escherichia coli (strain K12) GN=mglB PE=1 SV=1
14	37621	1	1	1	1	0.09	Regulator of RpoS OS=Escherichia coli (strain K12) GN=rssB PE=1 SV=1
13	21141	2	1	1	1	0.16	Murein DD-endopeptidase MepS/Murein LD-carboxypeptidase OS=Escherichia coli (strain K12) GN=mepS PE=1 SV=1