Proteomic investigation of an *Escherichia coli* terpene production factory: prospects for metabolic engineering

Faith Owabhel Robert, MBBS, MSc

Thesis submitted for the degree of Doctor of Philosophy

To the University of Sheffield, Sheffield, UK

On completion of research in the Biological and Environmental Systems Group within the Department of Chemical and Biological Engineering

July, 2013

Dedication

To my mum, Irene Yepayeye Uriah-Dieah, who promised to await my return with my PhD but passed away just before.

Thanks Mum. Forever in my heart

Acknowledgements

This PhD was a nerve racking up-hill experience that words may not describe accurately enough. This tumultuous yet illuminating journey would not have gone to a successful end but for the help and encouragement of some wonderful people.

Firstly, my supervisor, Professor Phillip Wright, who as the forge master-in-chief, gave me all the necessary guidance and encouragement. Thanks Phil, for the patience and understanding and for giving me the opportunity to be mentored by you. Cannot really thank you enough.

Very special thanks to my dearest friend, Esther. For your support, generosity and for all the kindness that you showed me. Thanks for sharing your knowledge and for the technical help. You are a friend indeed.

Thanks to Narciso who was always willing and ready to help. Thanks for your encouragement and friendly disposition.

Thanks to Jags Pandhal who took care of me and showed me the ropes and for being very approachable.

To all the other post doctoral fellows who helped me along the line. Special thanks to Khoa, Caroline, Yen and Joy. Thanks also to Ana, Joss and Wan.

To all of my colleagues and friends in ChELSI, especially to Raul, Mahendra, Stephen, Bharathi and Yimin. Thanks also to Joe, Nathan and Wen.

Thanks to Luana for being my friend in Sheffield, to Amabebe and to Owei as always.

Thanks to Ebi who has done a wonderful job with my boys all by herself and for enduring the 'stress'

To Azibapu, Azibatarari and Nadhioni for enduring my long absence. I am finally a certified scientist!-hope I have made you guys proud.

To my brothers, Jephthah and Azibaola and sister, Joyce for the spiritual, emotional and financial support.

And finally, to the almighty God, for everything

Abstract

Natural products, produced as secondary metabolites by plants and microbes, have been used for thousands of years to treat human ailments and still form the basis of drug discovery efforts. Terpenes form the largest category of natural products. Terpenes such as the antimalarial compound artemisinin and the anticancer compound taxol are potent medicines. Furthermore, terpenes are active ingredients in perfumes, nutraceuticals and biofuels. The fact that these products are produced in very small quantities in their natural host, from sources that are potentially exhaustible and prone to environmental fluctuations, have fuelled interest in sustainable, affordable and environmentally friendly means of production. Genetic engineering has proved an invaluable tool in producing considerable amounts of terpenes in heterologous hosts such as *Escherichia coli* and yeast. In this thesis, genetic/metabolic engineering and –omics strategies have been applied for the more efficient production of the anti-tumour compound β-elemene. Heterologous expression of sesquiterpene synthase from *Nostoc* sp. PCC 7120 in *E. coli* resulted in the successful production of β-elemene. Final β-elemene titres of up to 178 mg/L were quantified using targeted metabolomics strategies. Simultaneous monitoring of regulation of metabolic flux in recombinant *E. coli* producing β-elemene was carried out by applying isobaric tags for relative and absolute quantification (iTRAQ), a proteomics tool for relative quantification. A systematic evaluation of recombinant *E. coli* strains identified pathways that can be further optimised using metabolic engineering for commercially viable synthesis of β-elemene. This is the first report of an iTRAQ aided metabolic engineering study of terpene production in *E. coli*. This study demonstrates the importance of systems level –omic analyses in shedding light on bottlenecks that prevent commercially viable synthesis of highly beneficial natural products like terpene with *E. coli* as a biological chassis.

List of abbreviations

2DE	-	Two dimensional gel electrophoresis
ABC	-	ATP binding cassette
ATP	-	Adenosine triphosphate
ATPase	-	Adenosine triphosphatase
BSA	-	Bovine serum albumin
C	-	Carbon
cAMP	-	Cyclic adenosine monophosphate
CCM	-	Central carbon metabolism
CID	-	Collision induced dissociation
CoA	-	Coenzyme A
COG	-	Cluster of orthologous genes
CRP	-	cAMP receptor protein
CYP	-	Cytochrome P450
Da	-	Dalton
DAPA	-	Diamino pimelic acid
DHAP	-	Dihydroxy acetone phosphate
DMAPP	-	Dimethyl allyl diphosphate
DNA	-	Deoxyribo nucleic acid
dNTP	-	deoxy nucleoside triphosphate
DTT	-	Dithiothreitol
DXP	-	1 deoxy D xylulose 5 phophate pathway/ mevalonate independent pathway
E4P	-	Erythrose 4 phosphate
EDTA	-	Ethylenediaminetetraacetic acid
EI	-	Electron ionisation
ESI	-	Electrospray ionisation
eV	-	Electron volts
FDA	-	Food and Drug Adminstration
FDR	-	False discovery rate
FPP	-	Farnesyl diphosphate
G3P	-	Glyceraldehyde 3 phosphate
GC-MS	-	Gas chromatography - Mass spectrometry
GlcNAc	-	N acetyl glucosamine
GPP	-	Geranyl diphosphate
HILIC	-	Hydrophilc interaction liquid chromatography
His Tag	-	Histidine tag
HMG CoA	-	3 hydroxymethyl 3-1 glutaryl CoA
HMGR	-	3 hydroxymethyl 3-1 glutaryl CoA reductase
HMGS	-	3 hydroxymethyl 3-1 glutaryl CoA synthase
HPLC	-	High pressure liquid chromatography
ICAT	-	Isotope coded affinity tagging
ID	-	Identification
IDA	-	Iminodiacetic acid
IHF	-	Integration host factor

IPP	-	Isopentenyl diphosphate
IPTG	-	Isopropyl beta D 1 thiogalactopyranoside
iTRAQ	-	Isobaric tags for relative and absolute quantification
Kb	-	Kilo base
kDa	-	Kilo dalton
KEGG	-	Kyoto Encyclopedia of genes and genomes
LB	-	Luria Broth
LC	-	Liquid chromatography
m/z	-	mass/charge
MALDI	-	Matrix assisted laser desorption ionisation
MCS	-	multiple cloning site
MEP	-	2C methyl D erythritol 4 phosphate pathway
Min	-	Minute
MK	-	Mevalonate kinase
mRNA	-	Messenger ribonucleic acid
MS	-	Mass spectrometry
MS/MS	-	Tandem mass spectrometry
MVA	-	Mevalonate dependant pathway
NADH	-	Nicotinamide adenine dinucleotide (reduced)
NADPH	-	Nicotinamide adenine dinucleotide phosphate (reduced)
NCBI	-	National centre for biotechnology information
NHS	-	N hyroxy succinimide ester
NIST	-	National Institute for standards and technology
NMR	-	Nuclear magnetic resonance
NRPS	-	Non ribosomal peptides
OD	-	Optical density
OPG	-	Osmoregulated periplasmic glucan
PCC	-	Pasteur culture collection
PCR	-	Polymerase chain reaction
PEP	-	Phosphoenol pyruvate
PKS	-	Polyketides
PMK	-	Phosphomevalonate kinase
PTS	-	Phosphotransferase system
QToF	-	Quadrupole time of flight
RNA	-	Ribonucleic acid
Rnase	-	Ribonuclease
ROS	-	Reactive oxygen species
RP	-	Reverse phase
SCX	-	Strong cation exchange
SDS	-	Sodium dodecyl sulphate
SDS-PAGE	-	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SILAC	-	Stable isotope labelling of amino acids in cell culture
SPME	-	Solid phase micro extraction
SRS	-	Substrate recognition site
TB	-	Terrific Broth
TCA	-	Tricarboxylic acid cycle

TEAB	-	Triethyl ammonium bicarbonate
T_m	-	Melting temperature
TMAO	-	Trimethylamine N oxide
TMT	-	Tandem mass tags
tRNA	-	Transfer ribonucleic acid
UV	-	Ultraviolet
WHO	-	World Health Organisation

Table of contents

CHAPTER 1

Introduction

Natural products have had significant impact on human development. The hunt for the Earth's pool of natural resources have in turn impacted significantly on the natural environment, with sustainable environmental issues topping current discussions(Goudie, 2013). Many natural products such as the potent antitumor compound taxol from the Pacific Yew tree produced by living organisms are known to be produced in very small quantities. As such would require sacrificing huge portions of the natural host for any meaningful usage. Unfortunately, many of the products are have unique chemistries and as such are very difficult and not very cost effective for total chemical synthesis.

One such host of natural products that has received significant interest are the cyanobacteria. These products are synthesized through secondary metabolism giving rise to polyketides, non-ribosomal peptides and terpenes amongst others (Barrios-Llerena et al., 2007; A. M. Burja et al., 2001; Neilan et al., 1999; Tan, 2007; Welker and Von Döhren, 2006). Of all the classes of terpenes, the carotenoids are known to be widely produced by cyanobacteria (Hirschberg and Chamovitz, 2004), but not significantly much is known about other classes of terpene in this phylum. Efforts are currently ongoing to discover and characterize the non-carotenoid terpenes in cyanobacteria, many of which have demonstrated pharmacological properties (Agger et al., 2008; Gutiérrez et al., 2008; Jaki et al., 2000a, 2000b; Liu et al., 2007; Mo et al., 2009; Nagarajan et al., 2012). A sesquiterpene synthase from *Nostoc* sp.PCC7120 has been cloned and characterized by Agger et al. (2008). This sesquiterpene produces the anticancerous compound β-elemene.

Unfortunately, harvesting these useful cyanobacterial natural products for commercially viable and renewable manner is a herculean task and considering the slow growth of these organisms and specialized media requirement. Fortunately, with the characterization of many of these metabolites, biosynthesis in a cheap and tractable secondary host such as *Escherichia coli*represents the most commercially viable, sustainable and renewable option.

However, using a secondary host as a production factory for a foreign protein or metabolite produces a significant amount of stress on the host metabolism with system wide effects

which may affect the final product titres. Understanding the metabolic impact of these products in its production factory will provide a tool box for engineering better strains.

1.2 Aims of this study

The aim of this study is to gain an understanding of the global metabolic changes and systemic effects of a heterologous sesquiterpene synthetic pathway in an *E. coli* host

This thesis is laid out in 7 chapters as follows:

Chapter 2 sets the tone for the rest of the chapters by giving the necessary background information and rationale for the thesis. It discusses natural products with emphasis on terpene natural products and mechanisms of synthesis. It also discusses the cytochrome P450 which are known to act as downstream modifiers in terpene synthetic pathway and many other synthetic pathways as well as metabolic engineering and analytical tools used in the thesis.

Chapter 3 deals with the amplification and cloning of the sesquiterpene and cytochrome P450 from *Nostoc* sp.PCC7120 and verification of the insert.

In **Chapter 4,** the sesquiterpene synthase is expressed in an expression strain of *E. coli* and the terpene product is qualitatively and quantitatively assayed. The effect of an exogenous precursor supply to the terpene titres is also assessed and quantified using gas chromatography- mass spectrometry (GC-MS).

Chapter 5 uses an isobaric tag for relative and absolute quantification (iTRAQ)-based proteomics technique to study the effect of the sesquiterpene synthase on *E. coli.*

Chapter 6 also uses iTRAQ-based proteomics to evaluate the systemic changes that are produced in *E. coli* when the sesquiterpene synthase is expressed with a cytochrome P450 as well as an added source of terpene precursors.

Chapter 7 gives a summary of the major findings in this thesis and plans for future work arising from this work.

This thesis presents the production and quantitative assay ofβ-elemene,the antitumor sesquiterpene in an *E. coli* host and went further to increase the product titre by more than 34 fold by adding an exogenous source of biosynthetic precursors. The gobal metabolic changes in these strains were elucidated and could provide a template for rational metabolic engineering and further strain improvements.

CHAPTER 2

Literature review

2.1 Introduction

This chapter reviews the literature on natural products with emphasis on terpenes. The chapter is organized into 4 sections. The review starts with a brief introduction of cyanobacteria followed by (section 2.3) an introduction into natural products with emphasis on the terpenes while section 2.4 deals with the cytochrome P450s that also play significant roles in natural product synthesis. The evolution of metabolic engineering terpene production as well as some analytical tools used for engineering terpenes is discussed in section 2.5 and section 2.6 concludes with an outlook for this PhD thesis.

2.2 Cyanobacteria

Cyanobacteria are arguably one of the oldest known organisms on earth and are found in diverse habitats ranging from the oceans to fresh water bodies and to desert rocks (Harel et al., 2004). They are also found in such extreme environments as ice shelves, hot springs, and in the Arctic and Antarctic lakes as well as endosymbionts in lichens amongst others (Laybourn-Parry and Pearce, 2007; Steunou et al., 2006; Ward et al., 1998). Cyanobacteria are the only group of organisms that are able to reduce nitrogen and carbon in aerobic conditions and thus play a significant role in the nitrogen and carbon cycles as well as a major role in the evolution of life, by contributing to the oxygenation of the Earth (Herrero and Flores, 2008). They are photoautotrophic prokaryotes with a Gram-negative cell wall and contain chlorophyll a, phycobiliproteins, as well as photosystem I and II molecules (Stanier and Bazine, 1977; Woese, 1987).Cyanobacteria are a rich source of natural products comprising both primary and secondary metabolites including nonribosomal proteins, polyketides, terpenes and alkaloids and several of these are known to have anticancer, antiviral, UV protective activities as well as hepatotoxicity and neurotoxicity (Herrero and Flores, 2008).For example, *Nostoc* species have been employed therapeutically for pathologies such as renal stones, anal fistulas and many malignancies (Pietra, 1990). However, terpene synthetic pathways elucidation has been concentrated in

plants and fungi, but recent efforts in bacteria have shown that cyanobacterial strains just like the *Streptomyces* are capable of producing terpenes(Agger et al., 2008)

2.3 Natural Products

Natural products have found use in health and disease for thousands of years. Healers have known the usefulness of these natural products since the dawn of civilization, and still form a major corner stone of modern therapeutics (Butler, 2005; Newman et al., 2000). These products fall into several therapeutic classes such as anti-neoplastic, antioxidant, antibiotic, cytotoxic, photo-protective and anti-parasitic, and anti-viral (Adam M. Burja et al., 2001; Schwarzer et al., 2003). These natural products have not only served as drugs in their own right, but have also been as primers for the discovery and synthesis of new drugs (Newman et al., 2000). It was reported that between 1981 and 2010, 4% of the 1355 new chemical entities approved by the FDA were natural products and another 22% natural–product–derived, molecules, 15% as biologics with only 29% as totally synthetic molecules (Newman and Cragg, 2012). It is worthy of note that an estimated 80% of the world population according to the World Health Organisation (WHO) depend on traditional remedies for their primary health care needs (Newman et al., 2000). These natural products are classified structurally into polyketides (PKS), nonribosomal peptides (NRPS), alkaloids, lipoproteins, phenylpropanoids and terpenes (Adam M. Burja et al., 2001; Clardy and Walsh, 2004; Hanson, 2003). Most of these natural products are secondary metabolites that are produced by organisms from conditional pathways that are activated in certain situations such as stress, developmental needs, signaling, starvation and defense amongst others (Edwards and Ericsson, 1999). These products which are mainly sourced from plants, fungi and microbes have served therapeutics purposes for several millennia for the treatment of diverse and wide ranging conditions from infectious diseases to cancer to pain (Butler, 2005). Indigenous knowledge has played a significant role in many modern drug discovery efforts. A few examples include the antimalarial compound artemisinin, which was discovered as a result of collaborative efforts between modern scientists and traditional Chinese healers (Hsu, 2006). Vinblastines (Madagascar periwinkle), codeine (opium poppy), digitalis (purple foxglove) have been used for the treatment of diabetes, pain and cardiovascular conditions respectively for thousands of years (Newman et al., 2000).

2.3.1 Polyketides and non-ribosomal peptides (PKS/NRPS)

The polyketides and the non-ribosomal peptides are a large group of natural products that are produced through ubiquitous multi-domain synthases referred to as non-ribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) (Ansari et al., 2004; Walsh, 2004). These diverse groups of compounds with many important pharmacological properties are produced mainly from fungi and bacteria. Cyanobacterial species in particular have been shown to produce a large and wide variety of these compounds with varied activities ranging from antibiotics, immunosuppressant to anticancerous (Wase and Wright, 2008). Many of the known cyanobacterial bioactive compounds that have been characterised are known to be produced via the NRPS/PKS route (Hoffmann et al., 2003; Trimurtulu et al., 1994a). Some of these compounds include the antitumor compound GSV224 from *Nostoc sp*. (Smith et al., 1994; Trimurtulu et al., 1994a) ; the antiretroviral Cyanovirin-N which is active against HIV-1 and HIV-2 from *N. ellipsosporum* (Tsai et al., 2004), the boron containing anticancer Borophycin-11 useful in drug resistant prostate, breast, cervical and lung cancer from *Nostoc linkia* (Scorei and Popa, 2010). However, several other non-NRPS/PKS synthetic pathways leading to the production of secondary metabolites with the active participation of cytochrome P450 (CYP) are being appreciated, with the discovery of many CYP coding genes in their genome (Agger et al., 2008; Ikawa et al., 2001; Ke et al., 2005; Robert et al., 2010).

2.3.2 Alkaloids

The alkaloids are a diverse group of nitrogen-containing molecules that have been exploited by man for medicinal purposes, hunting, and poison as well as in warfare for thousands of years (De Luca and St Pierre, 2000).

These compounds are found mostly in plants and cyanobacteria but are also known to be sequestered by amphibians on their skin to deter predators (Daly, 1998). The amphibians, unlike plants and cyanobacteria, do not have the biosynthetic machinery for alkaloids synthesis but are sequestered from their diet and more than 800 bioactive alkaloid compounds have been extracted from them (Daly et al., 2005).

Several alkaloidal biosynthetic pathways have been elucidated with CYPs playing active role in many of them. The pathway for the L-tyrosine derived biosynthesis of the

medicinally important tetrahydrobenylisoquinoline group of alkaloids such as the analgesic morphine, the anti-microbial berberine, the anti-tussive codeine, the muscle relaxant tubocurarine, marcapine, and berbamurine require many cytochrome P450 especially CYP80B family members (Bird and Facchini, 2001; Chou and Kutchan, 1998; Pauli and Kutchan, 1998a).
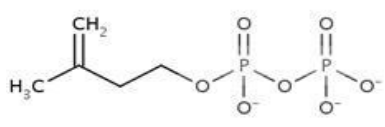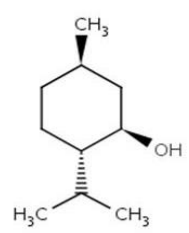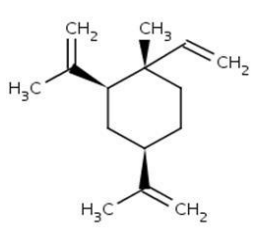
2.3.3 Terpenes

Terpenes, terpenoids or isoprenoids (Table 2.1) are one of the major structural classes of natural products and are produced by all organisms either as primary or secondary metabolites. They are reputed to be the most structural diverse group of natural products, with an estimated 30, 000 to 50,000 compounds already discovered to date (Keasling, 2010). Terpenes are mainly produced by plants, fungi and microbes and are known to have a wide range of biological functions (Kirby and Keasling, 2009) for which many have been exploited for mainly medical and agricultural purposes. Medically, they have been applied to treat cancers such as the diterpene taxol (paclitaxel) from *Taxus brevifola* (Jennewein and Croteau, 2001), and malaria via the antimalarial drug artemisinin a sesquiterpene from *Artemisia annua*(Liu et al., 2006). These molecules have also found use as insect antifeedants (azadirachtin from *Azadirachta indica*) (Brahmachari, 2004), nutraceuticals (carotenoids, lycopene, astaxanthin), as well as flavours and fragrances (limonene, geraniol). These molecules are mostly produced in very small quantities in the natural host. For example, 1 kg of taxol would need at least 2,000 pacific yew trees (Hartzell, 1991) and at a recommended dose of $135mg/m^2$every three weeks may only be sufficient for about two hundred patients with an average weight of 70kg in one year.
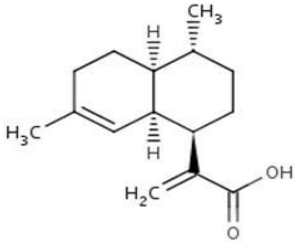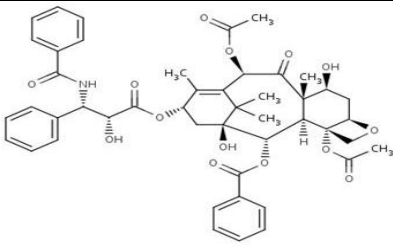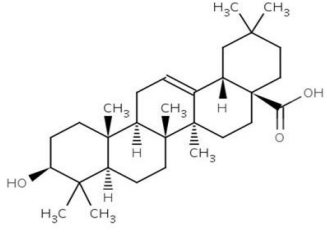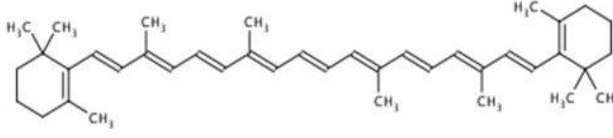
Additionally, terpenes are also produced as primary metabolites in all organisms such as gibberellic acid, steroids which serve to maintain membrane fluidity and as precursors of such steroid hormones as progesterone, oestrogen and testosterone, which serve very essential functions in these organisms (Hanson, 2003; Harrison, 1990). Other primary terpenes include the electron transport chain proteins, ubiquinone, plastoquinone and menaquinone as well as dolichol which is involved in glycosylation of membrane proteins and carotenoids in photosynthetic plants (Barkovich and Liao, 2001a). They also serve as pollinator attractants molecules and as signaling molecules (Edwards and Ericsson, 1999; Isman et al., 1990). All terpenes are synthesized from two basic linear building blocks

namely Isopentenyl diphosphate (IPP) and its structural isomer dimethylallyldiphosphate (DMAPP) (Withers and Keasling, 2007a).

These monomers are synthesized from simple carbon molecules through two independent biosynthetic pathways namely the mevalonate-dependent pathway (MVA) and the non-mevalonate pathway or the 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate pathway (MEP/DXP). DMAPP is combined with one, two or three molecules of IPP using prenyl transferases as catalysts to produce different carbon chain lengths of linear isoprene diphosphate backbones referred to as geranyl diphosphate (C_{10}), farnesyl diphosphate (C_{15}) and geranylgeranyl diphosphate (C_{20}) and geranylfarnesyl diphosphate (C_{25}), respectively (Lesburg et al., 1997). The chain length of the precursor isoprene backbone determines the class of isoprenoids produced based on the isoprene rule which states that terpenes are formed by the repetitive joining of isoprene units linked head to tail (Roberts, 2007; Ruzicka, 1953).

Table 2.1 some important terpenes classified according to chain length

Name	Carbon chain length	Class	Structure
Isopentenyl diphosphate	5	Isoprene monomer	
(−)-Menthol	10	Monoterpene	
(−)-β-elemene	15	Sesquiterpene	

Name	Carbon chain length	Class	Structure
(+)-Artemisinic acid	15	Sesquiterpene	
Taxol	20	Diterpene	
Oleanolic acid	30	Triterpene	
β-carotene	40	Carotenoid	

Source: Chemical Entities of Biological Interest (ChEBI)

These further undergo cyclization, rearrangement and modification to produce the various classes of isoprenoids. The cyclization reaction is carried out by a group of enzymes referred to as terpene synthases or cyclases (Rynkiewicz et al., 2001). The terpenes are classified based on the number of carbon atoms in their backbone into monoterpenes (C_{10}), sesquiterpenes (C_{15}), diterpenes (C_{20}), triterpenes (C_{30}) and carotenoids (C_{40}) (Trapp and Croteau, 2001). Additionally, there is a separate composite terpene derived group referred to as the meroterpenes. These are compounds made up of a terpene-derived unit bound to a non terpene unit such as chlorophyll, plastoquinone, some indole alkaloids and prenylated proteins (Chappell, 1995). The terpenes produced from the cyclization reactions could be subsequently tailored by modifying enzymes such as P450, aryl transferases and several

others to produce the final terpenoid in most cases (Walker and Croteau, 2001). The basic pathway for the final synthesis of most of these terpenoid requires the active participation of cytochrome P450s. The monoterpenes peppermint, perilla and spearmint are produced from the regiospecific hydroxylation of (-)-limonene at the C3, C7 and C6 respectively by CYP71D13/CYP71D15, CYP71D15 (Ito, 2008; C. J. Mau and Croteau, 2006). The highly effective antimalarial drug artemisinin is produced in the leaves of *Artemisia annua* by series of oxidative reaction catalysed by CYP71AV1, on the precursor amorpha-4,11-diene (Bertea et al., 2005; Kirby and Keasling, 2009). A host of other very important isoprenoids with CYP playing significant roles in their biosynthetic pathway include the anticancer drug, taxol, gibberellic acid, and oleanolic acid (Aharoni et al., 2005; Montellano, 2005).

2.3.3.1 Biosynthetic pathways

2.3.3.1.1 Mevalonate-dependent pathway (MVA)

The MVA pathway (Figure 2.1)uses two molecules of acetyl-CoA as starting substrates to synthesize IPP and DMAPP in the cytosol of cells. Eukaryotic cells as well as archaea and a few bacterial species (Table 2.2)such as *Borrelia burgdorferi, Staphylococcus aureus, Streptococcus pyogenes*, use the MVA pathway almost exclusively to make IPP and DMAPP from acetyl-CoA (Boucher and Doolittle, 2000; Chappell, 1995; Lichtenthaler et al., 1997). The pathway (Fig 2.1) is a seven enzyme pathway with an upstream 3 enzyme segment leading to mevalonate and a downstream 4 enzyme component leading to DMAPP. The pathway starts with sequential condensation of acetyl-CoA molecules catalyzed successively by acetoacetyl-CoA thiolase and 3-hydroxymethyl-3-glutaryl-CoA synthase (HMGS) to form 3-hydroxymethyl-3-l glutaryl-CoA (HMG-CoA).
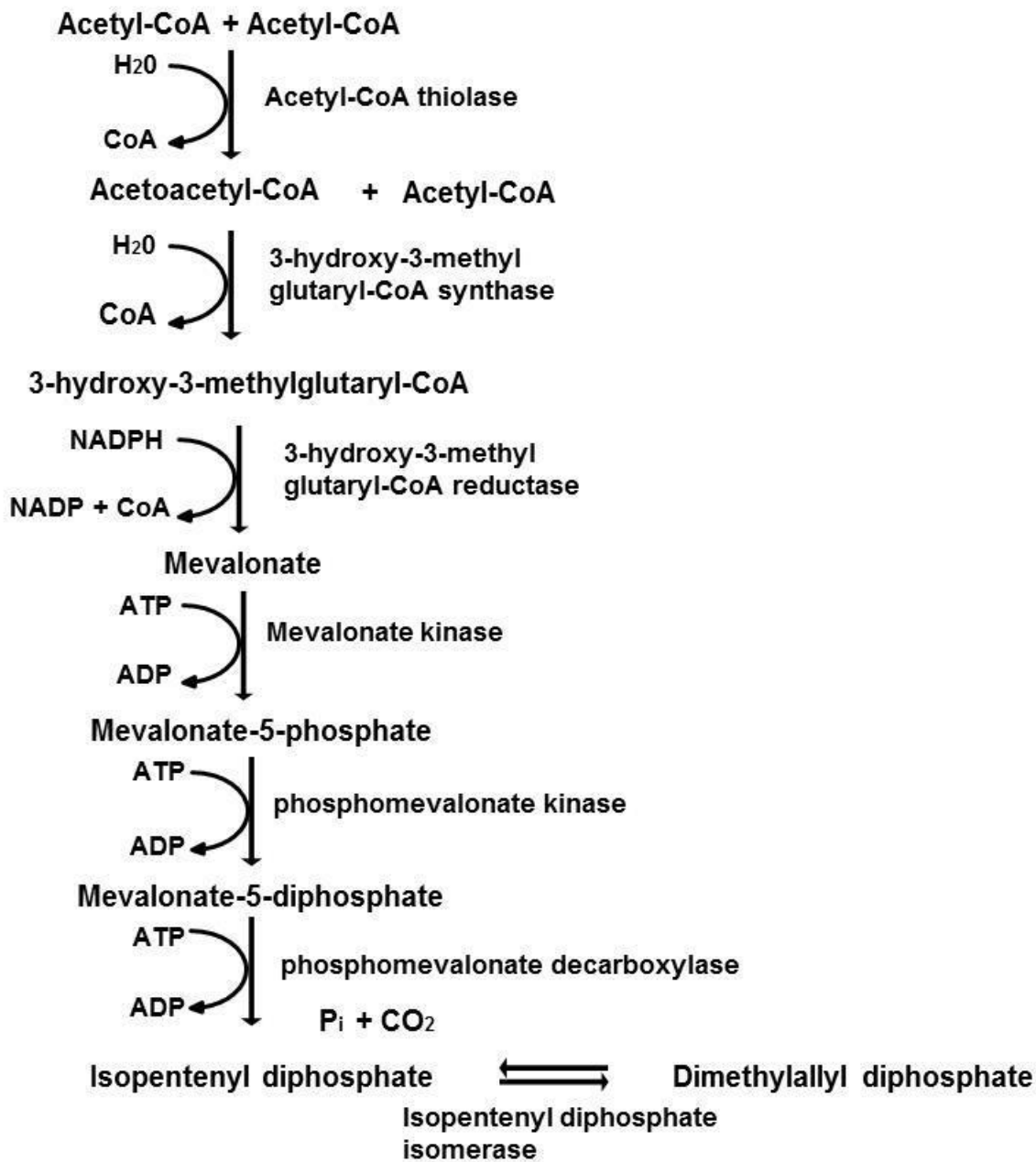
Figure 2.1 The mevalonate (MVA) pathway leading to the synthesis of IPP and DMAPP

This is then converted to mevalonic acid by HMG-CoA reductase (HMGR). This enzyme which catalyzes the rate limiting step in the synthesis of IPP/DMAPP is inhibited by specific inhibitors which forms the basis for the cholesterol lowering statins (Istvan and Deisenhofer, 2001; Lichtenthaler et al., 1997). The subsequent steps involve two sequential phosphorylation by mevalonate kinase (MK) and phosphomevalonate kinase (PMK) to yield mevalonate-5-phosphate and mevalonate-5-pyrophosphate. A final decarboxylation

reaction by mevalonate-5-pyrophosphate decarboxylase to yield IPP, which is then converted to the more reactive isomer DMAPP by the eukaryotic isomerase isopentenyl pyrophosphate isomerase (IdiI). This is an essential step in the biosynthetic pathway of isoprenoids and its derivatives(Bach, 1995).

2.3.3.2 Mevalonate independent pathway

It was assumed that all terpenes were made through the MVA pathway for decades until it was shown in the early 1990s through isotopic labeling using glyceraldehyde 3-phosphate (G3P) and pyruvate as precursors that there was an alternative pathway for IPP and DMAPP (Rohmer et al., 1993).

The non-mevalonate or 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate pathway (MEP/DXP) (Fig 2.3) is used by most bacteria and plants plastids, algae and some protozoa such as the *Plasmodium falciparum* (Table 2.2). Even in plants with both pathways, the cytosolic MVA pathway and the plastidial DXP are physically segregated but with some limited communication and mostly responsible for different synthetic processes (Bouvier et al., 2005). While the cytosolic MVA pathway is responsible for sterol biosynthesis, the plastidial DXP pathway is responsible for biosynthesis of the both primary metabolites such as carotenoids and chlorophyll components but also secondary metabolites such as terpenes while a few metabolites share precursors from both pathways (Eisenreich et al., 2004; Mandel et al., 1996; Nagata et al., 2002).

Unlike the MVA pathway however, the DXP pathway uses pyruvate and glyceraldehyde-3-phosphate as the starting substrates for IPP and DMAPP synthesis (Rohmer et al., 1996) . The synthetic pathway starts with the Mg^{2+}-ion dependent 1-deoxy-D-xylulose 5-phosphate synthase *(dxs)* as the starting and rate-limiting enzyme in a transketolase-like reaction that transfers a pyruvate derived acetaldehyde group to glyceraldehyde-3-phosphate to generate 1-deoxy-D-xylulose 5-phosphate (Eisenreich et al., 2004; Xiang et al., 2013). The subsequent step is a reductoisomerase reaction catalyzed by 1-deoxy-D-xylulose 5-phosphate reductase (*dsr* or ispC) which also requires a divalent cation in the form of Mg^{2+} or Mn^{2+} to form a branched aldose that is reduced with NADPH to 2-C-methyl-D-erythritol 4-phosphate (Koppisch et al., 2002; Takahashi et al., 1998).
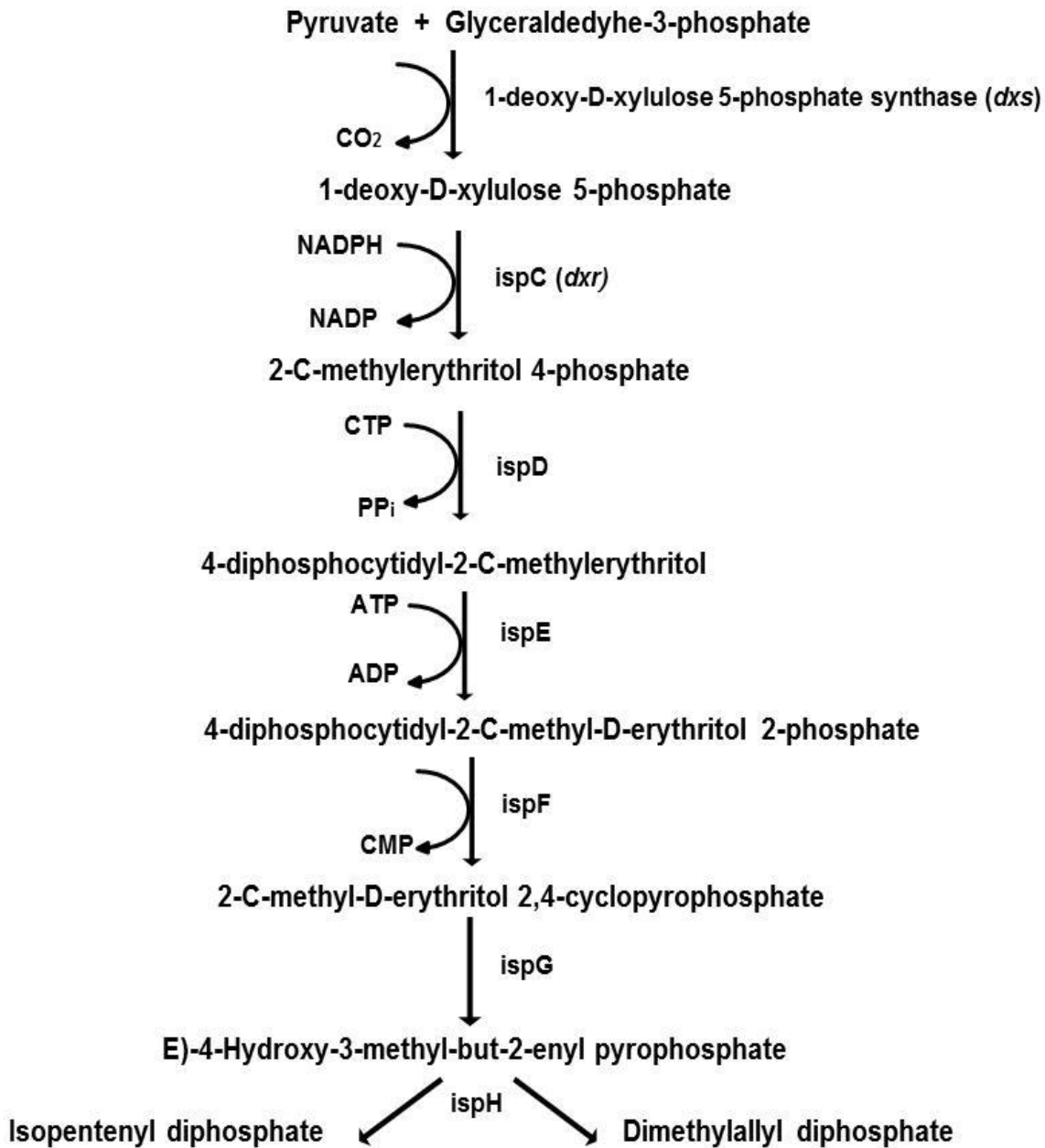
Figure 2.3 Representation of the mevalonate-independent pathway (DXP) for IPP and DMAPP synthesis

This enzyme (IspC) is inhibited by the antibiotic fosmidomycin and shows one example of the potential of this pathway as a drug target (Kuzuyama et al., 1998). The subsequent three steps require IspD (4-diphosphocytidyl-2-C-methyl-D-erythritol synthase) , IspE (4-diphosphocytidyl-2-C-methyl-D-erythritol kinase) and IspF (2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase) proteins that

transform 2-C-methyl-D-erythritol 4-phosphate to form 2-C-methyl-D-erythritol 2, 4-cyclodiphosphate by consecutive transfer of a cytidyl phosphate, ATP-dependent phosphorylation and molecular rearrangement with loss of the cytidyl monophosphate CMP (Herz et al., 2000; Lüttgen et al., 2000). The last two steps of this pathway are catalyzed by the iron–sulfur proteins IspG (4-hydroxy-3-methyl-but-2-enyl pyrophosphate synthase) and IspH (E)-4-hydroxy-3-methyl-but-2-enyl pyrophosphate reductase) to produce IPP and DMAPP at a molar ratio of 6 :1 through 4-hydroxy-3-methyl-but-2-enyl pyrophosphate (Gräwert et al., 2011a; Rohdich et al., 2002).

Unlike the eukaryotic IdiI, the IPP isomerase (idiII) is not found in most DXP utilizing bacteria and not essential in *Escherichia coli* (Cunningham et al., 2000; Hahn et al., 1999), however, it is known to be important for the biosynthesis of plants carotenoids and chlorophyll (Page et al., 2004). Importantly, some major disease causing protozoas such as *Plasmodium falciparum* as well as the tuberculosis causing mycobacterium use the DXP pathway to synthesis IPP which can serve as a useful drug target (Eisenreich et al., 2004b). In *E. coli* these precursors are used for the prenylation of tRNAs and synthesis of FPP which is then used to synthesize quinone and for cell wall biosynthesis (Connolly and Winkler, 1989; Martin et al., 2003a).

Table 2.2 IPP/DMAPP synthetic pathway used by different organisms

Organism	Non-mevalonate pathway	Mevalonate pathway	Idi I	idiII
Aquifales	Yes	No	No	No
Chlamydia group	Yes	No	No	No
Cyanobacteria	Yes	No	No	Yes
B. subtilis	Yes	No	No	Yes
M. genitalium	No	No	No	No
S. aureus	No	Yes	No	Yes
S. carnosus	No	Yes	No	Yes
S. coelicolor	Yes	No	Yes	No
L. monozytogenes	Yes	Yes	No	Yes
E. coli	Yes	No	Yes	No
R. prowazeckii	No	No	No	Yes
T. pallidum	Yes	No	No	No
B. burgdorferi	No	Yes	No	Yes
Archaea	No	Yes	No	Yes
Animals	No	Yes	Yes	No
Plants	Yes	Yes	Yes	No
P. falciparum	Yes	No	No	No
Yeast	No	Yes	Yes	No

Source (Eguchi et al., 1998; Eisenreich et al., 2004; Gräwert et al., 2011b; Horbach et al., 1993; Kakinuma et al., 1988; Moldoveanu and Kates, 1988)

2.3.3.1.3 Terpene synthases and prenyl transferases

Terpene synthases or cyclases are responsible for converting the acyclic terpene olefin derived from the action of prenyl synthases (geranyl pyrophosphate, farnesyl pyrophosphate, geranylgeranyl pyrophosphate) into the respective mono-, sesqui- and diterpenes (Starks et al., 1997). Physically, they are divalent metal ions (Mg^{2+} or

Mn^{2+})-dependent, soluble and acidic enzymes with native size ranging from 35-80 kDa (Leeper and Vederas, 2000). The primary structure of these enzymes do not show any significant sequence homology across plant, fungal and bacterial species despite their catalytic similarity even though the plant enzymes do show significant similarities (Starks et al., 1997). However, there are conserved features in the secondary structure that cuts across the terpene synthases and the prenyl synthases. These enzymes contain characteristic class I terpenoid cyclase fold with an aspartate rich motif DDXXDD on the C-terminus of helix D and a conserved NSE/DTE on the C-terminus of the H helix of the cyclases and are both implicated in magnesium ion binding for the initiation of catalysis (Aaron et al., 2010; Cane et al., 1996; Chappell, 1995; Christianson, 2006; Shishova et al., 2007; Starks et al., 1997). The prenyl synthases and the terpene cyclases both have similar catalytic activities but while the prenyl synthases produce chain elongation, the cyclases convert the acyclic prenyl olefin to mono, di or tricyclic compounds with some level of product degeneracy (Davis and Croteau, 2000).

The prenyl synthases produce a sequential head to tail condensation of IPP and DMAPP and subsequent addition of IPP to the products to generate longer chain olefins polymers of IPP and DMAPP. They do this through an intermolecular electrophilic attack of the allylic carbocation on the terminal double bond of the IPP molecule with subsequent coupling and stabilisation by deprotonation (Ogura and Koyama, 1998; Withers and Keasling, 2007b). The cyclization reaction proceeds in a similar manner. In this case, the reaction is an intramolecular electrophilic attack on a double bond down the terpene chain to cause a ring closure with subsequent involvement of hydride shift, methyl migration and final deprotonation (Chappell, 1995; Starks et al., 1997). The diversity of these compounds is primarily due to the terpene synthases which are capable of producing multiple products from a single substrate which could be further elaborated by downstream processing enzymes (Misawa, 2011). In this regard, even though the terpene synthases are substrate specific, they have the potential to produce many different products from a single substrate such as γ-humulene synthase which can produce up to 52 different products (Chappell, 1995; Christianson, 2006).

2.3.3.2 Classes of terpenes

2.3.3.2.1Monoterpenes

Monoterpenes are 10 carbon containing terpenes are synthesized by terpene synthases using geranyldiphosphate as universal precursors with some exceptions (Schilmiller et al., 2009; Yu and Utsumi, 2009). The geranyldiphosphate is derived from the condensation of one molecule of IPP with one molecule of DMAPP in a head to tail manner that are mainly produced through the plastidial DXP pathway (Yu and Utsumi, 2009).These molecules are, colorless, volatile organics with characteristic smell that make up the major constituents of plant essential oils (Geron et al., 2000; Gershenzon and Croteau, 1990). Structurally, they mostly have cyclic carbon skeleton such as the thujanes, pinanes, p-menthanes, bornanes, caranes, camphanes and fenchanes, while some are acyclic and both classes can be further modified by downstream enzymes such as cytochrome P450 which serve to increase the diversity (Mau and Croteau, 2006). The monoterpene synthases fall into two categories; those produced by gymnosperm synthases require a monovalent cation such as K^+ in addition to a divalent cation Mn^{2+} or Fe^{2+} and an alkaline pH for optimal activity while the angiosperm require only the divalent cation Mg^{2+} or Mn^{2+} and a neutral pH (Leeper and Vederas, 2000).

The essential oils are a complex mix of terpenes and non-terpenes with one or two monoterpenes making up a major component which also forms the basis for the naming of these oils (Hanson, 2007). More than 3400 of these compounds are known and they form the cornerstone of the perfumery and flavour industry and widely used in aromatherapy (Mau and Croteau, 2006). Menthol is the one of the most important monoterpenes and is produced mainly by plants of *mentha* species were it is the predominant constituent of peppermint oil in the mature plant (Croteau et al., 2005; Gómez-Galera et al., 2007). Menthol has a cooling-anaesthetic effect and pleasant aroma and is used extensively in toothpaste, cosmetics, teas and pharmaceuticals for which more than 7,000 metric tons are consumed with an annual market value of more than $300 million (Croteau et al., 2005).

The biosynthetic pathway for menthol biosynthesis comprises of 8 enzymes with limonene synthase catalysing the committed step of the pathway by cyclization of GPP and several

engineering strategies to increase yield of menthol and reduce undesirable side products such as menthofuran in transgenics have been reported (Croteau et al., 2005; Gómez-Galera et al., 2007; Mahmoud and Croteau, 2002).

Other important monoterpenes include myrcene, α-pinene, α-terpinolene, camphene, 2-methylisoborneol and camphor.

2.3.3.2.2 Sesquiterpenes

The sesquiterpenes are 15- carbon hydrocarbon compounds that that are widely distributed in plants as constituents of essential oils and some fungi as well some bacteria as well as marine organisms (Capon, 1995; Pham et al., 1991). These compounds are synthesized by sesquiterpene synthases using farnesyl diphosphate as substrate into more than 300 known hydrocarbon skeletons and also exist as oxygenated derivatives of their respective sesquiterpene olefin as alcohols, acids, ketones or as lactones (Agger et al., 2008b; Martin et al., 2001; Merfort, 2002). The sesquiterpenes unlike the monoterpenes are mostly produced from the mevalonate pathway in plants and the sesquiterpene synthases do not require a monovalent cation as seen in some monoterpene synthases (Davis and Croteau, 2000).

Many sesquiterpenes with wide ranging applications are known such as patchoulol in perfumery, the antibiotic pentalenolactone, the plant growth hormone abscisic acid, the antimalarial artemisinin, the biodiesel bisabolene as well as the brewery ingredient α-humulene (Cane et al., 1990; Connolly and Hill, 1991; Davis and Croteau, 2000; Leung and Giraudat, 1998; Peralta-Yahya et al., 2011a).

Artemisinin from the A*rtemisia annua* plant and several semisynthetic derivatives have been adopted as frontline drug to treat drug resistant malaria and is unique in the sense that it kills young ring stage parasite and therefore a more rapid reduction in parasitemia (Meshnick et al., 1996). The reliance on cultivation of the *Artemisia annua* plant made global prices of artemisinin to fluctuate between US$350 and US$1700 per kilogram but the advent of microbial production, prices are expected to reduce significantly to US$100 per kilogram (White, 2008). Another sesquiterpene with great pharmaceutical potential is

the β-elemene. This compound has been shown to have anti-cancer activities across several cancers such as brain, lung, prostate , cervical, ovarian and breast cancers as well as melanoma and leukemia (Chen et al., 2011; Li et al., 2010, 2005; Tao et al., 2006; Wang et al., 2005; Yu et al., 2011). It is naturally foundalongside germacrene-A as the main constituent of the oils from the leave of the cedar *Cedrela odorata* and also produced by the cheese associated fungi *Penicillium roqueforti*and has also been extracted from over 50 chinese herbs and plants(Larsen, 1997; Maia et al., 2000; Tan et al., 2011).

2.3.3.2.3 Diterpenes

The diterpenes are 20-carbon compounds that are made from the 20-carbon geranylgeranyl diphosphate precursor but unlike the sesquiterpenes are made through the plastidial DXP pathway in plants (Gómez-Galera et al., 2007). There are several important diterpenes including the phytohormone gibberellic acid, the antimicrobial casbene and the anticancer taxol (Davis and Croteau, 2000). Gibberellic acid produced from the fungus *Gibberella fujikuroi* is useful as a plant growth hormone and also used in the malting step in the brewing of beer (Hanson, 2007). Taxol from the bark of *Taxus brevifola* is perhaps the best known diterpene. Taxol marketed as paclitaxel and its several structural derivatives are touted as the most potent and commercially successful anticancer drugs (Kingston and Newman, 2007). It is active against a wide array of cancers such as breast, ovarian, cervical, lung and head and neck cancers (Jennewein and Croteau, 2001).

2.4 Cytochrome P450 superfamily

2.4.1 Discovery and spectral characteristics

The cytochrome P450s (CYPs) were discovered as a result of the pioneering work of several researchers in the late 1940s to the early 1960s who greatly benefitted from the development at that time of three key scientific instruments namely, the mass spectrometer, the double beam rapid scanning spectrophotometer and the oxygen electrode (Estabrook, 2003). It was Hayaishi et al (1951) in 1950, who discovered the oxidation of catechol to muconic acid by the incorporation of one atom of oxygen by an enzyme contained in soil,

which he named pyrocatechase. He subsequently went on to show that the oxygen atom that was incorporated into the substrate was from molecular oxygen and not water as previously believed and this group of enzymes were then named dioxygenases (Hayaishi and Stanier, 1951; Hayaishi, 2005). In 1958, Klingenberg first reported the presence of an unusual pigment in rat liver microsomes which was capable of binding carbon monoxide when reduced with a characteristic absorption peak at 450 nm (Garfinkel, 1958; Klingenberg, 2003). This pigment was later shown to be a heme protein, containing protoheme IX and was given the name cytochrome P450 (Omura, 1999). Buoyed by this discovery, Estabrook and colleagues confirmed the hydroxylation capacity of cytochrome P450 (Estabrook et al., 1963). They realised that the adrenal cortex microsomes which catalysed the C-21 hydroxylation of 17-hydroxyprogesterone had exactly the same carbon monoxide binding properties of the cytochrome P450 hemoprotein. Further confirmation that the cytochrome P450 served as a terminal oxidase of several other reactions came soon after, when it was shown that carbon monoxide also inhibited the liver microsome catalysed oxidative demethylation of aminopyrine (Orrenius et al., 1964). Indication that the enzyme was indeed ubiquitous with multiplicity of forms was shown by Alvares et al (1967). They observed in their experiments that there was a shift of the absorption maximum from 448 nm to 450 nm of the CO-complex of reduced P450 present in liver microsomes from animals treated with methylcholanthrene as compared to those from animal liver microsomes treated with phenobarbital (Alvares et al., 1967). Since the first paper on cytochrome P450 was published in 1962, the volume of knowledge and publications have increased dramatically with more than a thousand papers per annum since the late 80s (Estabrook, 2003). A survey of publications containing "cytochrome P450" in their title in the Web of knowledge database (accessed: 18/07/13) reveals that a least five hundred publications per annum has been maintained since the mid 1990s. This is seen in Figure 2.4
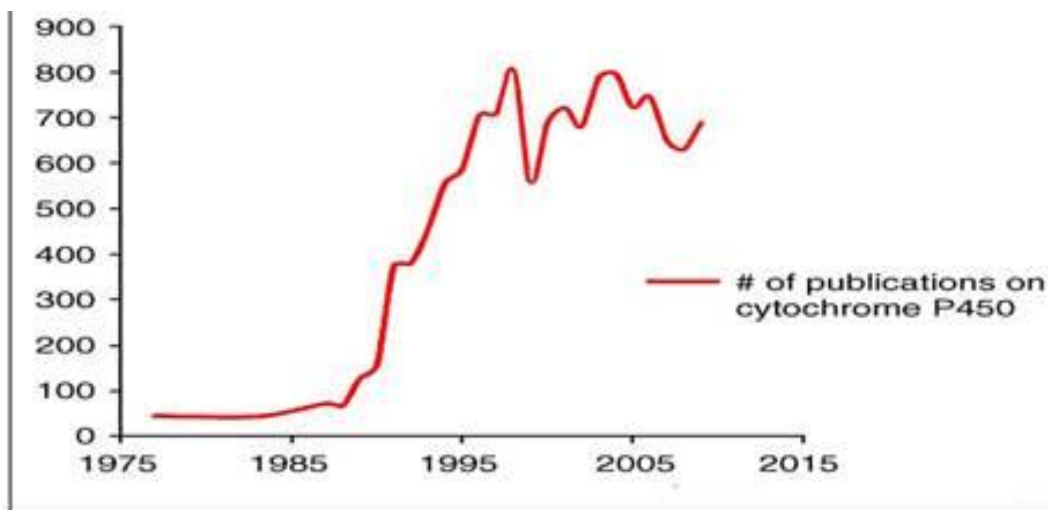
Figure 2.4 Number of cytochrome P450 publications since the 1980s (Source: Web of Knowledge)

Cytochrome-P450s are characterised by their vivid orange-yellow pigment on complexation with carbon monoxide. An intense soret absorption maximum band at 450 nm in the UV spectrum which is reversible by light, and hypothesised to be due to the trans configuration between the cysteinate and carbon monoxide ligands of the prosthetic heme group (Dawson and Sono, 1987; Klingenberg, 2003). This characteristic soret peak is a feature that is shared only with the chloroperoxidases, and nitric oxide synthases (Griffin, 1991; Stuehr and Ikeda-Saito, 1992).

2.4.2 Occurrence and distribution

The cytochrome P450s are a group of ubiquitous hemoproteinoxygenases that are found in all branches of the tree of life (Nebert et al., 1989). More than 11,000 genes encoding P450s have been identified in all kingdoms of life distributed in approximately 977 families to date and these are collected and updated by David Nelson (http://drnelson.uthsc.edu/P450.statistics.Aug2009.pdf). The cloning and characterisation of a cytochrome P450 in the archaeon, *Sulfolobus solfataricus*(McLean et al., 1998) and with at least 18 sequences discovered in the archaeabacteria to date, suggest that these enzymes are very ancient. The great diversity, occurrence and distribution of cytochrome P450 suggests that they could be involved in essential or critical metabolism such as

defense against environmental pollutants, drugs detoxification, synthesis of important molecules amongst others even though a few primitive forms of bacteria like *E. coli* do not have the enzyme (McLean et al., 1998). Cytochrome P450s are found as soluble proteins in prokaryotic cytosols, while in eukaryotes, they are mostly found in the endoplasmic reticulum as well as in the mitochondria, and all except the bacterial cytochrome P450s are membrane proteins anchored on the endoplasmic reticulum or the inner mitochondrial membrane by a transmembrane amino terminus(Nelson and Strobel, 1988).

In mammals, P450s are expressed in most tissues such as kidneys, prostate, brain and placenta (Waterman, 1992; Williams et al., 2000). However, it is in the liver that they are most abundant where they are found in the smooth endoplasmic reticulum, occupying more than 1% of the total weight of the liver hepatocytes (Williams et al., 2000).

2.4.3 Nomenclature and Structure

The intense interest in cytochrome P450 system as a result of its intriguing properties and characteristics has yielded significant results over the decades as fallout of the efforts of chemists, biophysicists, pharmacologist, biochemists and molecular biologists (Denisov et al., 2005). The cytochrome P450 superfamily nomenclature is based on sequence similarity in which sequences with a 40% amino acid sequence identity are collected into a single family and subfamilies are grouped based on more than 55% similarity with subfamily members designated by capital letters (Nelson et al., 1993). The cytochrome P450s are also classified based on their electron carrier protein and system of electron delivery that they use to supply the required electrons into ten classes (Hannemann et al., 2007). These are illustrated in Table 2.3

Table 2.3 Classes of P450 systems based on electron transport protein

Class/source	Electron transport chain	Localization
Class I		
Bacterial	NAD(P)H►[FdR]►[Fdx]►[P450	cystosolic, soluble
Mitochondrial	NADPH►[FdR] ►[Fdx]►[P450]	P450: inner mitochondrial membrane FdR: membrane associated Fdx: mitochondrial matrix, soluble
Class II		
Bacterial	NADH►[CPR]►[P450]	cystosolic, soluble; *Sulfolobuscarbophilus*
Microsomal A	NADPH►[CPR]►[P450]	membrane anchored, ER
Microsomal B	NADPH►[CPR]►[cytb5]►[P450]	membrane anchored, ER
Microsomal C	NADPH►[cytbRed]►[cytb5]►[P450]	membrane anchored, ER
Class III		
Bacterial	NA (P)H►[FdR]►[Fdx]►[P450]	cystosolic, soluble, *Citrobacterbraakii*
Class IV		
Bacterial	Pyruvate, CoA►[OFOR]►[Fdx]► P450]	cystosolic, soluble, *Sulfolobustokadaii*
Class V		
Bacterial	NADH►[FdR]►[Fdx-P450]	Cystosolic, soluble, *Methylococcuscapsulatus*
Class VI		
Bacterial	NAD(P)H►[FdR]►[Fldx-P450]	Cystosolic, soluble *R. rhodococcus*strain 11Y

Class/source	Electron transport chain	Localization
Class VII		
Bacterial	NADH► [PFOR-P450]	Cystosolic, soluble, *Rhodococcussp* strain NCIMB 9784, *Ralstoniametallidurans*
Class VIII		
Bacterial, fungi	NADPH►[CPR-P450]	Cystosolic, soluble *Bacillusmegaterium, Fusariumoxysporum*
Class IX		
Only NADH dependent, fungi	NADH►[P450]	Cystosolic, soluble, *Fusariumoxysporum*
Class X		
Independent in plants/mammals	[P450]	membrane bound, ER

Fdx (iron-sulfur-cluster); FdR (ferredoxinreductase (FAD)); CPR (cytochrome P450 reductase, (FAD, FMN)); Fldx (flavodoxin, (FMN)); OFOR (2-oxoacid:ferredoxinoxireductase (thiamine pyrophosphate, [4Fe-4S] cluster)); PFOR (phthatate-family oxygenasereductase (FMN, [2Fe-2S] cluster)). Reproduced from (Hannemann et al., 2007)

With an average mass of 50 kDa (Spatzenegger and Jaeger, 1995), the enzyme is made up of a protein component and a heme prosthetic group around which the catalytic properties are built. The conserved domain of cytochrome P450 is made up of a three parallel helices called D , L and I and an antiparallel helix, E and a prosthetic heme which is sandwiched between the proximal L and distal I helices respectively (Montellano, 2005, 1995; Poulos et al., 1987). The proximal L helix contains the absolutely conserved cysteine that serves as the proximal or fifth ligand of the heme iron (Nelson and Strobel, 1988). The cysteine-

heme ligand loop contains the signature sequence of the cytochrome P450s FxxGx(H/R)xCxG (Denisov et al., 2005). In addition, cytochrome P450s contain regions within the protein that are responsible for substrate recognition and channelling referred to as substrate recognition sites (SRS) (Gotoh, 1992; Roussel et al., 2000). These SRS which are six in number are responsible for the specificities and selectivity exhibited by these enzymes and mutations can drastically affect substrate specificity of any enzyme (Gotoh, 1992).

A greater insight into the uniqueness of cytochrome P450 was revealed when the first crystal structure from CYP101 (P450$_{cam}$) of the enzyme was published (Montellano, 1995). Subsequently, several other structures have been solved and deposited in the protein data bank. With over 51 unique structures either substrate bound or free enzyme crystallized (for example see 3G5F, 3E5J, 3B4X, 3DAM, 2Q9G in the PDB: http://www.pdb.org/pdb/home/home.do), this has given sufficient evidence of the unique fold of cytochrome P450 which is yet to be seen in any other enzyme system (Montellano, 2005). Despite less than 15% protein sequence homology across families, the structural core remains essentially conserved suggesting that the tertiary structure evolved to bring all the essential components together irrespective of source or location (Degtyarenko and Archakov, 1993) which perhaps explains the common mechanism of oxygen activation. Despite this conservation of tertiary structure across families, the observation that many cytochrome P450s especially the drug metabolising ones in humans have a very wide range of substrates specificities suggests some element of variability. This is indeed the case, as the most variable regions in the structure across families have been confirmed to be the F and G helices (SRS2 and SRS3) and their connecting loop which can change from open to close conformation (Poulos, 2005) favouring an induced-fit mechanism for substrate binding (Pylypenko and Schlichting, 2004).

2.4.4 Reactions catalyzed by cytochrome P450

The Earth's atmosphere changed from a reducing one to an oxidising one about 2 billion years ago, as evidenced by the dating of banded iron formations as a result of the photosynthetic activities of cyanobacteria (Lewis, 2001; Marguilis et al., 2000). This increase in atmospheric oxygen was essential for evolution from simple forms to more

complex multi-cellular organisms that invariably required oxygen for their metabolism. In order to utilize the oxidative properties of oxygen, which is normally chemically inert to organic compounds, nature had to evolve key enzymes with the capacity to activate oxygen from its triplet ground state to the active singlet excited state needed for essential metabolism such as biosynthesis, growth and respiration, of these organisms (Hamdane et al., 2008; Lewis, 2001). The cytochrome P450s are involved in three fundamental functional roles in organisms. Firstly, they are needed for the synthesis of essential primary metabolites such as synthesis of steroid hormones in mammals required for growth and development and homeostatic regulation (Ioannides, 1996). Secondly, they are responsible for metabolism and detoxification of exogenous molecules for removal from the organism as a protective measure such as the phase I biotransformation of drugs in humans by liver cytochrome P450 (Spatzenegger and Jaeger, 1995). Lastly, they are major players in the biosynthesis of many secondary metabolites whose role to the organisms may not yet be evident but which are highly sought for their useful roles as medicines amongst others (Montellano, 2005). The role of cytochrome P450 as detoxifiers in animals is thought to be a co-evolutionary response to the production of potentially toxic plant P450 metabolites (Gonzalez and Nebert, 1990) and in general, cytochrome P450s present in non-animal species have a role in the metabolism of endogenous substrates primarily (Lewis, 2001).

As with all other enzymes, cytochromes P450s are able to catalyse reactions at physiological conditions by reducing the activation energies of the reactions. Uniquely however, cytochrome P450s do this principally by gradually increasing the oxygen-oxygen bond length with resultant reduction of the oxygen-oxygen bond energy that eventually leads to splitting of the dioxygen bond for substrate monooxygenation (Lewis and Pratt, 1998). The interest in these enzymes is not only due to this, but also due to their unique spectral properties, wide range of substrate classes, selectivity and specificity, and significantly also by the diverse range of reactions that they catalyse (Coon et al., 1996; Groves, 1997).The one electron reduction of the heme moiety changes the oxidation state of iron to a high spin ferrous state which enables binding of molecular oxygen to produce a ferrous cytochrome P450-dioxygen complex. A second one electron reduction of the ferrous cytochrome P450-dioxygen complex and protonation produces a low spin ferric hydroperoxy state.

A second protonation and subsequent cleavage of the dioxygen bond gives the highly reactive compound I which is a Fe (IV)=O compound. The second oxygen atom is reduced to water. The oxygen atom from compound I is transferred to the bound substrate with subsequent dissociation to form the oxygenated products as seen in Figure 2.5
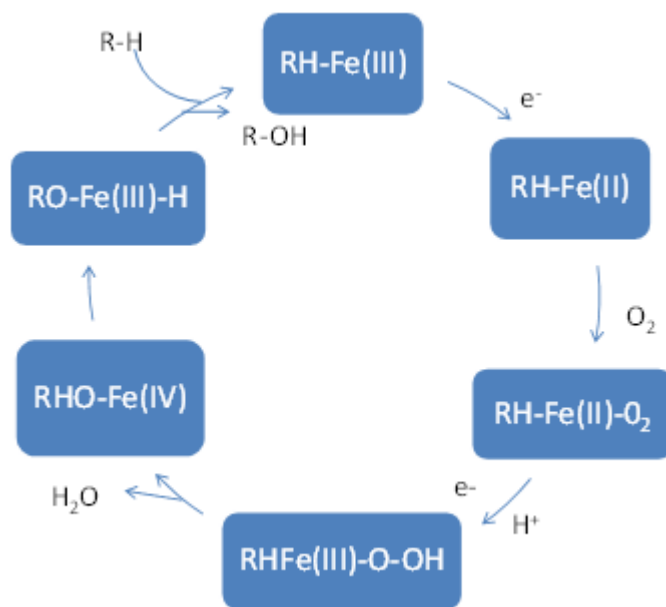
Figure 2.5 Cytochrome P450 catalytic cycle

2.4.5 Cytochrome P450 metabolic pathways

The cytochromes P450s have been employed by nature to catalyze numerous synthetic and catabolic processes in organisms for the overall wellbeing of these organisms. The metabolic pathways that utilize cytochrome P450s either for synthetic, catabolic or biotransformation are very numerous and diverse.

2.4.5.1 Primary metabolism of cytochrome P450

2.4.5.1.1 Endogenous metabolism of cytochrome P450

Cytochrome P450s are critically involved in many essential endogenous metabolic aspects of cells that are required for growth and development, homeostatic regulation, signalling, embryogenesis and defences (Lamb et al., 2006; Stoilov et al., 2001; Zerbe et al., 2002).

They are involved in the biosynthetic pathway of several molecules that are essential for the physiological functions of many organisms from simple prokaryotes to the complex multi-cellular mammals. In mammals, cytochrome P450 play critical roles in the synthetic pathway for the synthesis of prostaglandins, fatty acids, steroids, bile acids, fat soluble vitamins and sex hormones. These synthetic P450 are generally product specific i.e. they yield one product as opposed to the multipurpose xenobiotic metabolising cytochrome P450 (Schenkman, 1992; Stoilov et al., 2001). Many steroids and steroid hormones in mammals use cholesterol as the precursor which is produced by the action of CYP51 and the actions of several other cytochrome P450 such as CYP11A1, CYP17A1, CYP11B1, CYP19A1, CYP21A1 at various points leading to the production or interconversion of several steroid hormones such as progesterone, corticosteroids, oestrogen, and testosterone (Lewis, 2001; Payne and Hales, 2004). This has resulted in the suggestion of targeting some of these CYPs for cholesterol control (Pikuleva, 2006).

In plants, cytochrome P450 activities are even more essential as they inherently rely on CYPs for most of their biological functions (Bolwell et al., 1994; Morant et al., 2007; Nelson, 2006). The highest number of CYP genes are found in plants, and since the cloning of the first plant CYP gene in 1990 (Bozak et al., 1990), at least 4200 other CYP genes have been discovered to date (http://drnelson.uthsc.edu/P450.statistics.Aug2009.pdf)and these genes constitutes up to 1% of the total protein coding genes (Nelson et al., 2008). In addition to using CYPs for synthesis of several secondary metabolites, plants use CYPs for the synthesis of primary metabolites such as plant hormones/signalling compounds (gibberellins, absicic acid, and brassinosteroid), cytokinins, lignin, defence compounds, UV-protectants, fruit and flower colour (Bolwell et al., 1994; Choe et al., 1998; Clouse et al., 1996; Durst and Nelson, 1995; S.D Clouse and Sasse, 1998), these are classified into four basic classes as summarised in Table 2.4

The evolutionary phylogeny of plants CYPs shows that the earliest P450 families are needed for such primary function as biochemistries of sterol and carotenoid synthesis, followed by those CYPs families apparently needed for their roles in adaptations to a land environment, while the newest CYPS families are thought to be mainly responsible for synthesis of pigments, flavours and specie specific secondary metabolites (Nelson et al., 2008).

Table 2.4 substrate classes of plant cytochrome P450s. Reproduced from (Lewis, 2001).

Class	Example	P450 enzyme

Phenylpropanoids
   Lignin
   Pigments
  UV-protectant
	Cinnamic acid	CYP73

Terpenoids
   Steroids
   Hormones
   Defence molecules
   Aroma/odorant molecules
	Geraniol	CYP71

Fatty acids Cutin precursors Suberin precursors Defence molecules	Oleic acid	CYP94
Diverse substrates	Tyrosine	CYP79

Microbial species, unlike higher species of plants and animals are not yet known to require CYPs for essential metabolism except for CYP51 which is found in animals, plants, fungi, protists and some bacteria, and is required in the synthetic pathway for steroid biosynthesis (Lepesheva and Waterman, 2007). Their roles in secondary metabolism and natural product biosynthesis will be examined in later sections.

2.4.5.2 Secondary metabolism and P450 based natural products

Most of the reactions that are catalysed by cytochrome P450 that are of most use to man in plants and microbes are from the so called secondary metabolic pathways with cytochrome P450 involved in key products such as codeine, morphine and erythromycin (Cupp-Vickery and Poulos, 1995; Schwab, 2003a; Unterlinner et al., 1999). Secondary metabolites are usually not produced constitutively by organisms but may be produced during periods of

stress and may not be essential for normal physiological functions (Mann, 1987), as such many serve as natural products, harvested for their medicinal and other applications.

Cytochrome P450 in microbes as in plants, are numerous and very diverse. This is demonstrated by cataloguing of the abundance of CYP genes in their genomes, with 2570 fungal, 905 bacterial, 22 archaeal and 247 protists (http://drnelson.uthsc.edu/P450.statistics.Aug2009.pdf), collated to date (18/07/2013). There is, though as yet no delineated structural or physiological role reported to date apart from CYP51 (Kelly et al., 2001). It can be inferred from this that they are mostly employed in secondary metabolism. Despite this abundance of CYPs in microbial genomes, there are still several such organisms as *E. coli* without CYP genes (Montellano, 2005), which further buttresses the point that CYPs may not be necessary for primary metabolism. CYPs in plants, fungi and bacteria are responsible for the biosynthesis of important products (Table 2.5) and with more than 100,000 natural products produced in plants, CYPs are known to be active in most of the pathways (Montellano, 1995; Schwab, 2003b). The classes of secondary metabolites produced by plants and microbes that have significant CYP activity in their synthetic pathway include isoprenoids, alkaloids, phenylpropanoids, and fatty acids as well as in the polyketide synthetic pathway (Edwards et al., 2004; Lewis, 2001). In the synthesis of macrolide antibiotics such as methymycin, amphotericin and erythromycin, the cytochrome P450 is responsible for imparting functionality and specificity through their site- specific modifying reactions (Carmody et al., 2005; Donadio et al., 1993). Several other therapeutic agents produced in microbes such as the antibiotics vancomycin from *Streptomyces orientalis,* amphotericin B from *Streptomyces nodosus,* erythromycin from *Saccharopolysporaerythraea*, the anticancer, adriamycin have CYPs actively involved in their biosynthetic pathway (Hopwood, 2007; Levine, 2006). The antimitotic indole alkaloids vincristine and vinbalstine used in cancer chemotherapy is biosynthesised by *Catharanthus rosea* (Madagascar periwinkle) using CYP71D12 (Schröder et al., 1999).

Table 2.5 Examples of cytochrome P450 catalysed bioactive natural products

Source organism	Compound	Activity	References
Lyngbya majuscula	Hectochlorin	Antifungal	(Ramaswamy et al., 2007)
Saccharopolypspora erythraea	Erythromycin	Antibacterial	(Cupp-Vickery and Poulos, 1995)
Streptomyces nodosus	Amphotericin	Antifungal	(Lamb et al., 2006)
Streptomyces avermitilis	Avermectin	Antiparasitic	(Lamb et al., 2003)
Aspergillus fumigatus	Fumitremorgin C	Anticancer	(Grundmann et al., 2008)
Catharanthus rosea	Vincristine	Anticancer	(Schröder et al., 1999)
Taxus brevifola	Taxol	Anticancer	(Walker and Croteau, 2001)
Artemisia annua	Artemisinin	Antiparasitic/anti-malarial	(Bertea et al., 2005)
Papaver somniferum	Morphine	Opiate, analgesic	(Pauli and Kutchan, 1998b)
Streptomyces avermitilis	Filipin	Antifungal	(Xu et al., 2010)

2.4.6 Cyanobacterial P450s

Several cytochrome P450 and cytochrome P450-like genes have been identified in cyanobacterial genomes but little attention has been given to their functional characterisation. Cyanobacteria, especially those from the marine environment have been implicated as prolific producers of natural products (Adam M. Burja et al., 2001; Faulkner, 1988; Herrero and Flores, 2008; Hoffmann et al., 2003; Simmons et al., 2005; Trimurtulu et al., 1994b), and as evidence for the importance of cytochrome P450s in production of secondary metabolites in many organisms has been presented above, it would be beneficial to investigate any linkage between them. There is ample evidence that cytochrome P450 in cyanobacteria participate in specific metabolic processes and did not arise as a result of

isolated gene transfer and the crystal structure of CYP120A1 has proved that it indeed participate in retinoid metabolism (Alder et al., 2009; Ke et al., 2005; Kühnel et al., 2008). However, most of the presently known natural products from cyanobacteria are produced through the NPRS, PKS or NRPS/PKS pathway (Jones et al., 2009; Wase and Wright, 2008). It is a little surprising that the first cloning, heterologous expression and functional characterisation of a cytochrome P450 in a cyanobacterial specie was only done as recently as 2005(Ke et al., 2005) despite the importance of both the organism and the molecule. It is emerging that cyanobacteria have several CYPs in their genomes. A BLASTp search in the NCBI cyanobacterial genome database using the characterised protein NP_488726 from *Nostoc sp*. strain PCC 7120 (Agger et al., 2008b) or CYP120A1 from *Synechocystis* sp. PCC 6803 (Kühnel et al., 2008) gives 100 putative cytochrome P450s sequences distributed in most of the known cyanobacterial species as seen in Figure 2.6, 63 of these proteins have been characterised and named with most (30) belonging to the CYP110 family, 12 fall into CYP213 family while CYP120 family currently has 10 members. CYP110 from *Nostoc sp*. strain PCC 7120 unlike most bacterial CYPs is not a cytosolic enzyme but membrane bound just like eukaryotic CYPs; not induced by alkanes or participates in alkane biodegradation but is involved in ω-hydroxylation of long chain fatty acid and play a role in nitrogen fixation (Wase and Wright, 2008).
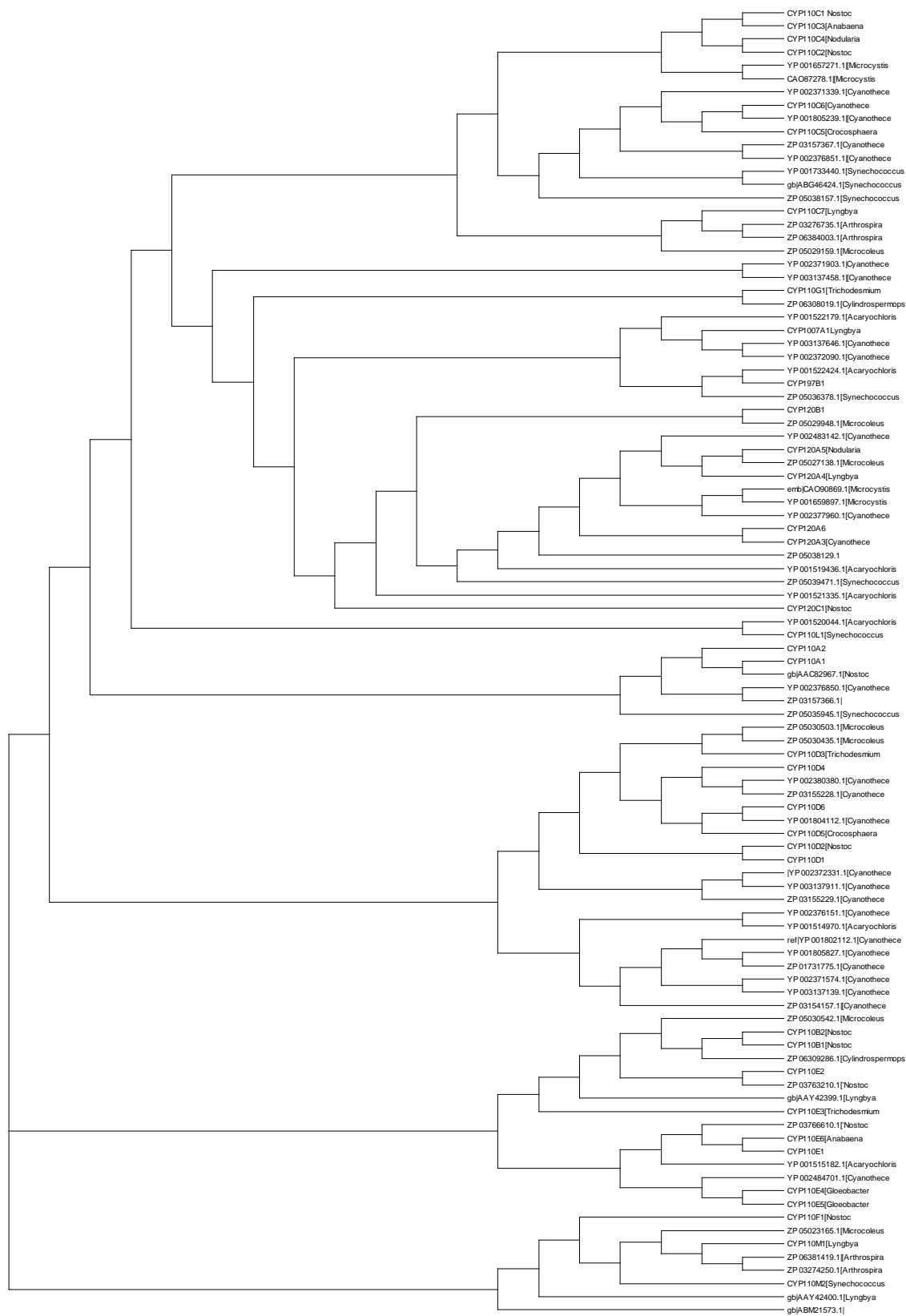
Figure 2.6 Phylogenetic tree of cytochrome P450 in cyanobacteria generated through BLASTP search in NCBI cyanobacterial database using NP_488726.

2.5 Metabolic Engineering

The availability of natural products for potential therapeutic and other uses is often limited by the ability of the primary producers to yield enough products and in renewable quantities (Chang and Keasling, 2006). Natural products are endowed with complex structures with many chiral centers that are essential for their unique biological roles (Koehn and Carter, 2005). However, total chemical synthesis of these complex structures have proved quite difficult and when successful, the yield may be poor or may not be as effective (Chemler and Koffas, 2008; Koehn and Carter, 2005). Extraction of these very important natural products from their natural host is not without problems. Firstly, they are produced at unappreciable levels (Kirby and Keasling, 2009). Secondly, they may be produced in a complex mixture with other less useful products, making downstream separation quite difficult (Koehn, 2008; Martin et al., 2003a). Thirdly, many of these products are mined from rare and exotic species that raise very serious issues of conservation in addition to the prohibitive cost of production in their natural host and unpredictable impact on the environmental. It has been shown that it requires a minimum of six 100-year Pacific yew trees to yield adequate quantities of taxol for the treatment of a single patient and although total chemical synthesis is now possible, there are associated issues of product yield and cost to consider (Horwitz, 1994; Wessjohann, 1994).

The development of metabolic engineering has been spurred not only by the major challenges of the 21st century such as climate change, renewable energy, food security, pollution and health but also by development of new technologies (Betenbaugh and Bentley, 2008; Lindberg et al., 2010). Metabolic engineering was first defined by James Bailey as "*the improvement of cellular activities by manipulation of enzymatic, transport, and regulatory functions of the cell with the use of recombinant DNA technology*"(Bailey, 1991). James Bailey who was a pioneer in the field saw the cell as a production line only, and he proposed the recruitment of heterologous activities for strain improvement and the redirecting of metabolic flow at branch points as the essential strategies for metabolic engineering. Stephanopoulos (1999) went a bit further in 1994 when he defined metabolic engineering as "*the design of biochemical reaction network to accomplish a certain objective*". His definition, though less constrained, was still limited. However, with the advent of the post genomic era, availability of new tools and technologies and development

of system biology/synthetic biology, there has been a paradigm shift from seeing the cell as a production line to that of a production factory (Cortassa, 2002; Kholodenko and Westerhoff, 2004; Smolke, 2009). This shift in focus from reductionist view to a system wide view of the cell as a system of elements with complex regulatory network operating at different levels, has given impetus to the development of synthetic biology which is a subject that aims to design and construct novel biological entities or modify already existing ones for enhanced productivity (Fu and Panke, 2009). Hence metabolic engineering has been recently defined as a systemic approach to the problem of remodelling and reconfiguring the many pieces of the cell through careful understanding and measurement of how small changes affect the network as a whole (Tyo et al., 2007). The ultimate aim is to achieve strain improvement and cell engineering using transcriptomics, proteomics, metabolomics, computational systems biology, protein engineering and synthetic biology. Metabolic engineering as depicted in Figure 2.7 can be seen as a product pathway that is fed by other stand-alone pathways with a common general theme.
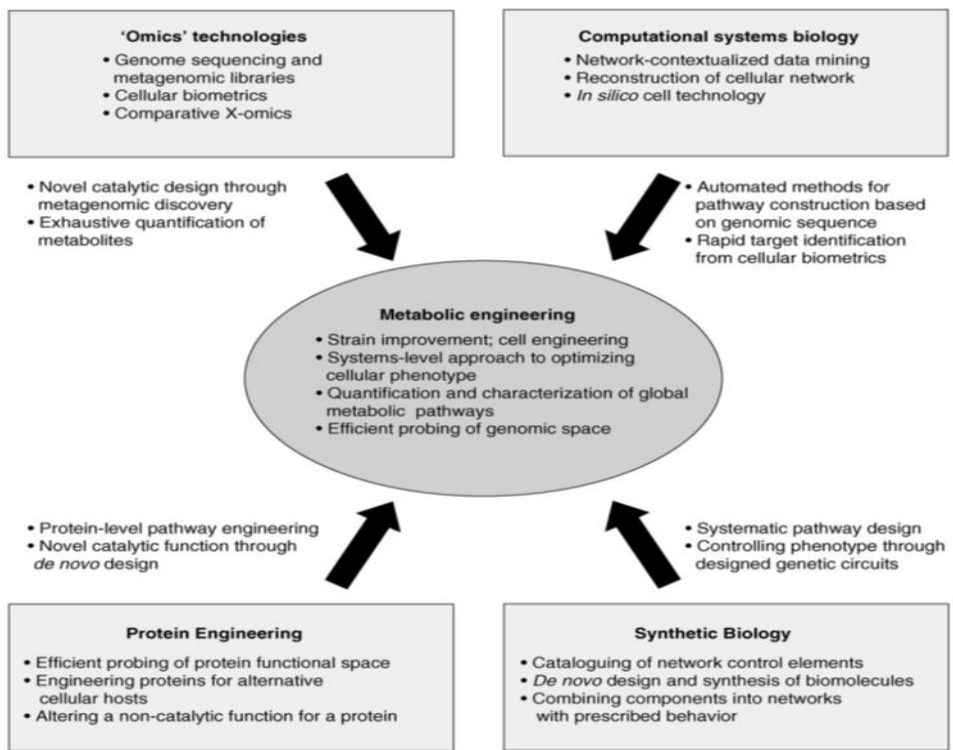

Figure 2.7 Metabolic engineering in relation to other complementary fields. Reproduced from (Tyo et al., 2007)

The success of metabolic engineering has been astounding so far with many useful compounds already produced through metabolic engineering in the two decades of the discipline. In therapeutics, several drugs or drug leads such as the antimalarial compound artemesinin and the anticancerous drug taxol which hitherto were produced exclusively by cultivation have either been wholly heterologously synthesized or a biosynthesis of their direct precursor achieved (DeJong et al., 2006a; Dietrich et al., 2009a; Ro et al., 2006a). In addition several other molecules such as isoprenes, flavonoids, dihydroxyacetone and ethanol have been produced in high titre (20.8 mg/l of naringenin, 700 mg/l dihydroxyacetone, 0.46 g/g of ethanol) through metabolic engineering (Nguyen and Nevoigt, 2009; Trantas et al., 2009; Watts et al., 2004). Many of these pathways have multiple cytochrome P450 enzymes at critical points of the pathway. In this project a proteomics based approach is proposed to evaluate the systemic changes in a terpene producing *E. coli* mutant. This could provide a rational basis for engineering an improved strain.

2.5.1 Proteomics

The proteome of a cell is the total protein complement expressed by the genome of that cell (Wilkins et al., 1996); and the proteome unlike the genome is dynamic and subject to environmental influences and conditions. As opposed to the gene expression focus of transcriptomics which is 'blind' to post translational modifications, proteomics gives a more biological representation of cellular activities of the cell at any time point (Han and Lee, 2006). Proteomics looks at the proteins not as single separate entities but as part of dynamic interacting system that change with time and conditions (Liebler, 2002). To this end, proteomics has shown significant rapid development from the traditional 2-dimensional electrophoresis 2DE gel based methods to the present day mass spectrometry based gel–free (shotgun) proteomics (Evans et al., 2012; Ow and Wright, 2009; Peng and Gygi, 2001; Wilkins et al., 1996). While 2DE gel saw to the birth of proteomics, it soon became obvious that it was very labor intensive, low abundance proteins were mostly missed and also excluded very acidic/basic proteins, very large and small proteins as well as membrane proteins (Aebersold and Mann, 2003; Gygi et al., 2000; Peng and Gygi, 2001). For these reasons other methods of protein separation without the use of 2DE gels

were developed and optimized. These methods aim to reduce the complexity of the protein/peptide mixture typically through liquid chromatography before MS analysis. MS-based proteomics can be used to identify and quantify relative and absolute protein expression profile of a cell, the post translational modifications that these proteins have undergone as well as the protein interactions (Cox and Mann, 2011). Not only is MS-based proteomics useful for 'shotgun' system wide investigations and application but as seen in the work of Redding-Johanson et al. (2011a) can be applied to specific small-scale projects as 'targeted proteomics' with significant outcome. Targeted proteomics uses the same basic proteomic pipeline as shotgun proteomics, but in this case focus is placed on a small set of known protein targets for fast analysis and with a significantly greater dynamic range and sensitivity (Picotti et al., 2009).

There have been significant improvements in proteomics and as such the proteome coverage has also shown significant leap from about 50% in unicellular organism (Bantscheff et al., 2007)to almost 100% for the yeast proteome as reported by Mann et al. (2013). These improvements have been achieved through a combination of liquid chromatographic improvements through high pressure-high performance liquid chromatography with long column and gradient and high resolution mass spectrometry (Köcher et al., 2011; Mann et al., 2013).

The modern quantitative proteomic tool box has two basic paradigms: the stable isotope label and label-free methods. Several stable isotope labeling methodologies from *in vivo* metabolic labels such as stable isotope label of amino acid in cell culture (SILAC) to in vitro labeling techniques such as enzymatic labelling, isotope-coded affinity tagging (ICAT, tandem mass tag (TMT) and isobaric tag for relative and absolute quantification (iTRAQ) each of which has its advantages and disadvantages (Evans et al., 2012; Gygi et al., 1999; Miyagi and Rao, 2007; Ong et al., 2002; Ross et al., 2004; Schmidt et al., 2005; Thompson et al., 2003). This high-throughput method of looking at the cell would give a pictorial view of the functional dynamics of the cell and the consequences of any external or internal stimuli or stresses. This would also enable a better understanding of the metabolic and regulatory adjustments that a cell makes in order to continue to survive in the face of a stressful input. Understanding the metabolic and regulatory perturbation of a cell (through iTRAQ-based proteomics) as a result of a single or multiple stimuli or stresses could potentially provide a tool box to redirect or tailor the organism towards a desired productive end.

2.5.1.1 iTRAQ-bases proteomics

iTRAQ was first described in 2004 as a multiplexed quantitation strategy for relative and absolute quantification of proteins in a complex mixture (Ross et al., 2004). This method as the name implies uses isobaric chemical labels to derivatize the peptides or proteins for subsequent MS analysis. The label is a molecule of three parts; a methylpiperazine reporter group and an N-hydroxysuccinimide ester (NHS) primary amine reactive group linked by a balancing carbonyl group (Fig 2.4). It forms covalent bonds by attaching to the N-terminus and to the epsilon-amine group of lysine side-chain residues of proteins or peptides. During the labelling reaction, the NHS-ester acts as a leaving group, whilst the reporter and the normalizer groups are attached to the peptide or protein. By distributing strategically stable isotopes of carbon or nitrogen or oxygen (light and heavy $^{12}C/^{13}C$, $^{14}N/^{15}N$ and $^{16}O/^{18}O$) across the reporter and normalizer groups, reporter ions of different masses are generated without altering the total mass of reagents. The mass of the reporter and balance groups are maintained at 144 Da for the 4-plex (Figure 2.8) and 304 Da for the 8-plex with the piperazine group ranging in mass from 114.1, 115.1, 116.1 and 117.1m/z for a 4-plex and 113.1, 114.1, 115.1, 116.1,117.1 118.1, 119.1 and 121.1 m/z for an 8-plex (Pichler et al., 2010).The isobaric nature of the labels ensures that they are indistinguishable when bound to peptides from different samples such as mutant or wild type and can then be used to quantify peptides when they undergo collision-induced dissociation to release the reporter ion which populate the low mass region of the MS/MS spectrum. The multiplex capabilities allows for analysis of up to 8 samples derived from different conditions in one experiment. The relative peak intensities of these reporter ions directly correlate to the relative abundance of the peptide and hence the protein from which the peptide is from. Since quantitative information in iTRAQ is obtained from tandem MS, the same peptide from each sample appears as a single peak in the MS spectrum with reduced complexity with significant reduction in the chemical noise and therefore increase in sensitivity compared to ICAT(Aggarwal et al., 2006; Evans et al., 2012). When compared to 2DE, iTRAQ permits multiplexing and biological/technical replication of samples for better statistical analysis, gives more reproducible results and is versatile enough to quantify proteins across functional categories cellular location as well as abundance level (Aggarwal et al., 2006; Leitner and Lindner, 2006). iTRAQ-based quantification have been shown to have good

precision due to the equal distribution of errors across the low mass quantification region. The detection of regulated proteins (up and down) is usually not compromised (Li et al., 2012; Pichler et al., 2010; Udeshi et al., 2012).

Accuracy is reduced by the existence of isotopic contamination. However, the isotopic impurities in the iTRAQ reagents can be corrected when the manufacturer provides correction coefficients (this is usually the case for iTRAQ 4-plex). Non-reporter ions, such as immonium ions, can also be a source of interference with the reporter ions in the low-mass region. For example, interference of M+1 peak of the phenylalanine (F) immonium ion (with m/z 120.03) interferes with the m/z 121.1 iTRAQ reporter ion. The specific interference of the F immonium can be eliminated using correction coefficients (S. Y. Ow et al., 2009). Usually, ion contamination is weaker than the reporter ions but this can still have a dramatic effect in reporter ion intensities affecting the accuracy of the ratios measurements and quantification. Note, that if high resolving power MS platforms are used, it becomes possible to confidently differentiate between them and the reporter ions (Evans et al., 2012)
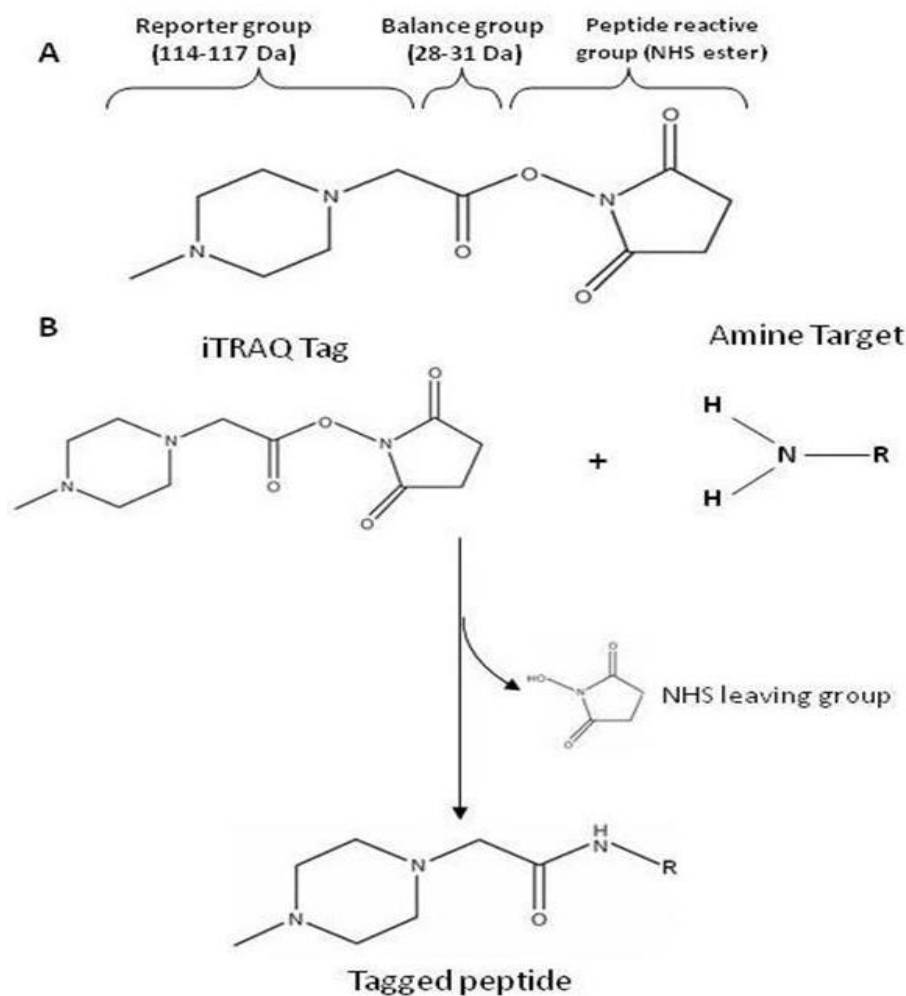
Figure 2.8 A 4-plex iTRAQ label showing the configuration of the 3 components. Modified from (Ross et al., 2004)

Since its discovery, iTRAQ has been employed to study a wide range of systemic perturbations in many different cell types such as *E. coli,* yeast, cyanobacteria, human cell lines (Castrillo et al., 2007; Fuszard et al., 2013; Glen et al., 2008; Mukherjee et al., 2011; Ow et al., 2008a).

2.5.2 Analytical methods

2.5.1.1 SCX/HILIC Chromatography

The obvious advantages of shotgun proteomics over 2DE based methods have principally been due to the power of the separation ability of liquid chromatography. In addition to the

preferred reasons for using the LC-MS based proteomics outlined above, the efficiency of this system is significantly greater and more high-throughput in view of the fact that it can be directly coupled to the mass spectrometer either as an online or offline with offline offering more advantages (Peng et al., 2003). Multi-dimensional liquid chromatography applications employing orthogonal separation significantly reduces the complexities of peptide mixture, enhance the resolution with attendant increase in both the proteome coverage and dynamic range (Motoyama and Yates III, 2008; Washburn et al., 2001). In our lab, a 2-dimensional separation comprising of an offline separation employing either strong cation exchange (SCX) or hydrophilic interaction liquid chromatography (HILIC) and an online reverse phase (RP) is mostly used in the proteomics workflow. Both these methods take advantage of different physico-chemical properties of the peptides for separation.

SCX chromatography has been widely employed as a first dimensional offline fraction method for quantitative proteomics. The separation of peptides takes advantage of the electrostatic interaction between the negatively charged polysulfonate stationary phase and the positively charged amino acid (caused by low pH of binding buffer). This interaction is then modulated by a graded change in ionic strength of the mobile phase with elution of peptides across the salt gradient depending on the strength of the ionic interactions (itself a function of net charge of peptide, charge distribution/density, peptide length and conformation) (Lorne Burke et al., 1989). Some organic solvent is usually added to the mobile phase in order to reduce the effect of hydrophobic interaction between the stationary phase and the peptides. Ross et al. (2004) have demonstrated that an SCX-HPLC is well suited for iTRAQ based quantitative proteomics without any undue influence on the labeled peptides. However, the high salt content of eluted peptide fractions may require an additional clean up step before RP-MS analysis.

HILIC unlike SCX uses a hydrophilic stationary phase (polysulfoethyl) with a high percentage of hydrophobic organic mobile phase (usually more than 70%) (Alpert and Andrews, 1988). The high percentage of organic in the binding buffer causes an enrichment of an aqueous layer on the hydrophilic stationary phase in which the hydrophilic peptides can partition between the hydrophobic mobile phase and hydrophilic stationary phase (Alpert, 1990). The high organic content of HILIC and minimal salt (unlike SCX) means that it is more compatible with RP-MS. In HILIC, the gradient is based on gradual reduction of the organic solvent and a proportionate increase in the water causing the

elution of more hydrophobic (less hydrophilic) first while the more hydrophilic (less hydrophobic) are eluted later as the organic solvent is reduced and water increased (Naidong, 2003; Zhu et al., 1991).

HILIC has been reported to have superior separation performance to SCX and is more orthogonal to RP (Gilar et al., 2005). HILIC has been successfully applied to iTRAQ based proteomics and as shown by Ow et al., (2011a) and Longworth et al., (2012), is superior to SCX in reducing under-estimation well as proteome coverage.

2.5.2.2 Tandem Mass spectrometry

Analysis of molecules by mass spectrometric techniques has seen tremendous development spanning several decades but this powerful analytical tool was not available to answer biological questions until the early 1990s when both electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) were developed (Hoffmann and Stroobant, 2013) and brought with it the era of MS-based quantitative proteomics. This development was so significant that the 2002 Nobel Prize in chemistry was awarded in part to John B. Fenn and Koichi Tanaka for 'their development of soft desorption ionization methods for mass spectrometric analyses of biological macromolecules' In proteomics, the ESI is usually coupled to a Quadrupole Time-of-flight (Q-ToF) and measures the mass/charge (m/z) of eluting peptides fromthe coupled LC. The eluting positively charged (using a very low pH carrier buffer of formic acid) peptides from the electrospray are 'dried' using a counter current high temperature gas in a process of desolvation before analysis (Fenn et al., 1989). A survey mass spectrum (MS) is acquired and selected peptides are then subjected to further fragmentation (MS^2) typically under CID conditions to produce either N-terminal fragments (b series) or the C-terminal fragments (y series) upon which peptide identification is based.

In recent years, quantitative MS-based proteomics has benefitted from development of very high resolution instrument enabling near total proteome coverage (Mann et al., 2013). The resolution of a mass spectrometer defined as the peak width at half maximum height divided by mass is a dimensionless quantity with great significance. Instruments with high resolutions have made it possible for increased mass accuracies and with the ability to differentiate peptides with similar masses and significantly increase the confidence level of peptide identification(Cox and Mann, 2011, 2008). Another factor in mass spectrometer

performance is the significant increase in speed of acquiring a survey mass spectrum and fragmentation of the selected peptides (Andrews et al., 2011; Michalski et al., 2011)

2.5.2.3 Gas chromatography-mass spectrometry (GC-MS) analysis

Gas phase chromatography connected to a mass spectrometer (quadrupole analyser) has been applied to study low molecular weight compounds which have high thermal stability and high volatility. Thus GC-MS provides a method for both qualitative and quantitative assay of the metabolome of a cell or of specific metabolites and is applied to measure volatile, semi-volatile or non-volatile compounds that can be derivatized (Mashego et al., 2007).

Metabolites are low-molecular weight organic compounds of less than 1000 Da that are produced by the cell and participate in cell metabolism or product cell metabolism (ter Kuile and Westerhoff, 2001). Some these metabolites as stated earlier are not necessarily involved in primary metabolism but are secondary metabolites without defined functions in the cell (van der Werf et al., 2005).

Since one of the goals of metabolic engineering is to achieve reliable, sustainable and cheap cell factories for fine chemicals, a robust and reliable method of qualitative and quantitative assay of these chemicals would be a necessity.

The GC-MS (Figure 2.9) is made up of a gas chromatograph attached to a mass spectrometer which serves as a detector. The analyte is injected into the injector port either in the gas phase or as a liquid which is then vaporized by the high temperature of the injection port and carried by an inert carrier gas through to the capillary column in the oven. The column provides a platform for interaction between the gaseous mobile phase (analyte) and the stationary phase which can be modulated by changes in temperature. The chemical properties of the stationary phase as well as the length, diameter and thickness determines the interactions between the analyte and the column as well as the binding capacity of the column (Hoffmann and Stroobant, 2013). The operation of a temperature gradient in the oven enables the constituent to be eluted through the column and reach the mass spectrometer.

In the source of the mass spectrometer, these compounds are then ionised upon bombardment with high energy electrons released from a tungsten filament, typically 70 eV, a Process known as electron ionization (EI). At the same time they are also fragmented

and both molecular ion and fragments are separated by the mass analyser. Once they reach the detector electrical signals are generated which are proportional to the abundance of the ions in m/z values (Fiehn et al., 2000).
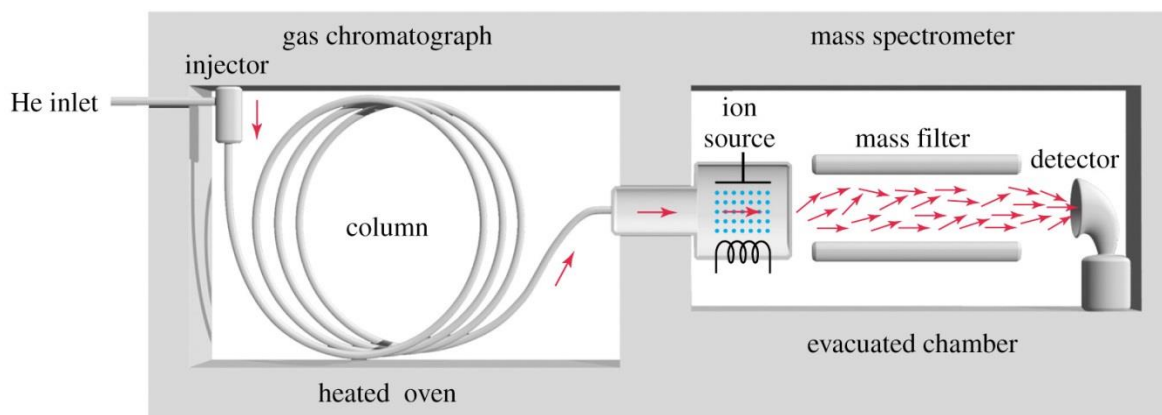
Figure 2.9 Schematic representation of a GC-MS showing the oven and the mass analyser.

However a fast and convenient method of qualitative assay of volatile or semi-volatile organic compounds employs the head space sampling and direct injection into the GCMS (Zambonin et al., 2004). This solid phase micro extraction (SPME)-GCMS method uses an SPME fibre to selectively adsorb and concentrate specific groups of volatile organics from environmental samples as well as culture head space (Grimm et al., 2001; Larsen, 1997; Nilsson et al., 1996). This provides a convenient way of screening culture samples for the presence of an expected metabolite before subsequent extraction and quantitative assay. The combined information provided by the gas chromatographic retention time and the mass spectra is used to identify the compound(s) or the spectra can be matched through a (national institute for standard and technology (NIST) database search.

2.5.3 Engineering of mevalonate pathway

Due to the wide ranging applications of terpenes in medicine and industry and agriculture combined with the limitations imposed by sustainable and commercially viable natural sources, heterologous expression in tractable hosts such as *E. coli* and yeast have proved an excellent alternative. These organisms could provide a steady, reliable and cheaper source

of these terpenes. *E. coli* has been employed for heterologous expression of terpenes because it is a fast growing, cheap to cultivate and has been well-studied with a host of wide ranging molecular genetics and bioinformatics resources available (Han and Lee, 2006).

Harada et al. (2009) used the mevalonate pathway from yeast starting from HMG-CoA with *idi* 2 with mevalonolactone supplementation to give more than 13 fold (958 µg/mL) of α-humulene over a control strain (using native DXP) expressing humulene synthase gene using acetoacetate as the carbon source.

Martin et al., (2003a) engineered a strain of *E. coli* containing the seven the MVA pathway genes from yeast and the FPP synthase and showed a 40 fold increase yield of amorphadiene yield to 110 mg/L with a codon optimized amorphadiene synthase compared to synthesis with native DXP pathway. This increased to 400 mg/L in a biphasic reactor (Newman et al., 2006). Pitera et al. (2007) and Kizer et al. (2008) went further to show using a combination of microarray analysis and targeted metabolic profile that further strain improvement was compromised by the accumulation of HMG-CoAwhich causes inhibition of fatty acid synthesis and this could be relieved by feeding the cells with paltmitic acid. An imbalance of single or multiple enzymes could result in the accumulation of a precursor(s) and thus compromise pathway carbon flux with negative implications as shown by these authors. Further, using targeted proteomic methodologies, it was revealed that there was reduced expression of two of the mevalonate pathway enzymes namely mevalonate kinase (MK) and phosphomevalonate kinase (PMK) and on that basis, the two enzymes were codon optimized and expressed using a stronger promoter and this led to a more than 500 mg/L amorphadiene titres (Redding-Johanson et al., 2011a).

The mevalonate pathway template or modified forms of it has also been employed with other terpene synthases to increase productions of terpenes. Notable examples include a more than 900 mg/L production of the biodiesel bisabolene when this was coupled with bisabolene synthase (Peralta-Yahya et al., 2011a). Using this pathway, the acyclic sesquiterpenefarnesol was produced at a titre of 136 mg/L (Wang et al., 2010). Wang et al. (2011) were also able to increase the yield of the bio-jet fuel α-farnesene from a baseline production of 1.2 mg/L to 380 mg/L.

The taxol biosynthetic pathway starting from the Isopentenyl diphosphate/dimethylallyl diphosphate precursor is made up of twenty steps, with the first committed step catalysed by taxadiene synthase (a terpene synthase) converting gerynylgerynyl diphosphate to taxa-

4(5), 11(12)-diene. Taxa-4(5), 11(12)-diene subsequently undergoes a series of eight cytochrome P450 catalysed reactions and two acylation steps to produce a molecule of 10-deacetyl baccatin III, which is a direct precursor of taxol (Chau et al., 2004; Hefner et al., 1996; Jennewein and Croteau, 2001; Wildung and Croteau, 1996). DeJong and colleagues (2006b) have described the first metabolic expression for the synthesis of baccatin III in yeast cells with some limited success. They succeeded in functionally expressing five genes leading to the production of the committed taxadiene and also eight taxoid downstream genes, but baccatin III yield was speculatively hampered at the first hydroxylation step because they relied on endogenous cytochrome P450 reductase for electron transfer from NADPH. However, Engels et al have been able to increase taxadiene production by 40 fold to more than 8mg/l in yeast (Engels et al., 2008).This was done by increasing flux toward taxadiene synthesis by combining the expression of a truncated version of HMG-CoA reductase to increase supply of isopentenyldiphosphate with the expression of a regulatory mutant allele of *upc2.1* that helps to reduce flux towards competitive steroid biosynthesis.

On the other hand, Ajikumar et al. (2010) were able to use combinatorial methodologies (promoter strength and plasmid copy number) and the native DXP pathway of *E. coli* to increase titres of taxadiene to 1g/L with limited accumulation of the toxic indole.

2.6 Conclusions

Engineering a desired pathway either in the native host or in a heterologous organism is never an isolated event whether the process is hypothesis driven or not.

Every organism has inherent regulatory and metabolic properties which can potentially interfere with the expression of an exogenous protein which could cause a wholesale targeted degradation, mis-folding or formation of insoluble aggregates because the right chaperones are not available, poorly expressed.

Any pathway or network that serves as a template for in situ constructive metabolic engineering or heterologous inverse metabolic engineering would necessary be a part of an existing network or system within that organism or could potentially form part of or affect an existing network in the secondary hosts. These secondary networks are linked to other

networks or pathways which may not be known and which could potentially exert a significant control on the engineered pathway. In effect, the consequence of a single exogenous gene product may not be localized to a single pathway or an adjacent pathway but may actually be felt far afield which may be the cause of failure in such a project (Bailey et al., 1996).

Therefore, for a successful metabolic engineering of any synthetic pathway either heterologously, or in the native host, requires a thorough elucidation of the pathway, the enzymes involved, their mechanism of action, the structural genes that encode these enzymes, rate limiting steps and the regulatory networks controlling it. Knowledge of all the parameters involved in any synthetic pathway will enable the metabolic engineer to decide on the system perturbation, iterative/refinement steps and optimization protocol.

The use of –omics technologies to investigate and direct the rational engineering of terpenes is not widespread with the few examples using transcriptomics and targeted proteomics have given very encouraging results. A systemic study of a terpene producing system in *E. coli* using iTRAQ-based shotgun proteomics would therefore add a significant layer of understanding and thus facilitate further metabolic engineering. There is however no known application of iTRAQ-based proteomics exploration of terpene producing *E. coli* system to my knowledge. This project therefore aims to evaluate the metabolic changes in a terpene producing cell factory upon which further strain improvement and refinement can be explored.

CHAPTER 3

Cloning

3.1 Introduction

In *Nostocsp*. PCC 7120, a single terpene synthase gene has been identified alongside an adjacent cytochrome P450 (Agger et al., 2008). This gene lies in a gene cluster that comprise a downstream cytochrome P450 separated by a 21 bp intergenic segment and a 5.7 Kb sensor (serine/threonine kinase) gene downstream of the cytochrome P450. Advances in methodologies for the biosynthesis of important terpenoids from cyanobacteria require identification, cloning and expression of these genes a in suitable host that is cheap and cost effective. Expression of the terpene synthase and cytochrome P450 proteins in a genetically tractable organism such as *E. coli* provides the advantages of low cost, high level of heterologous protein expression, lack of post-translational protein modification, scalability and fast turnover (Edwards and Palsson, 2000; Peti and Page, 2007; Studier, 2005). This ensures that sufficient proteins for biochemical and biophysical characterization as well as in vivo production of the terpenoid product. Additionally, and more importantly, it provides a platform for evidence-based strain improvement through systemic biology analysis and metabolic engineering. Cloning of these genes into an appropriate plasmid with an inducible promoter for expression in *E. coli* therefore provides the necessary tools for this project. The Novagen duet vectors (pCDFDuet -1 and pETDuet -1) provide suitable vehicles for cloning by having two multiple cloning sites, compatible replicons and drug resistance cassettes for co-expression as well as an inducible T7*lac* promoter. The *lacI* gene they carry provides a means of regulating basal expression with an N-terminal His Tag sequence for purification of fusion protein. Additionally, they are also compatible with other co-expressed plasmids that provide for a complete and enhanced terpenoid biosynthetic pathway in *E. coli*.

Therefore, the aim of this chapter is to describe the genetic engineering strategies and optimizations carried out in order to successfully transfer the genes important for the biosynthesis of terpenoids from *Nostoc* sp. PCC 7120 into *E. coli*.

3.2 Materials and methods

3.2.1 Culture conditions

3.2.1.1 Nostoc sp. PCC 7120

Nostoc sp. PCC 7120 was grown 100 ml of BG11 (Table 3.1) media supplemented with trace metal solution (Table 3.2) in a 250 ml flask. The inoculated culture medium was grown in a plant growth chamber (Sanyo, Japan) at 25° C, 50 µE m^{-2} sec^{-1} illumination and 60% humidity in continuous light conditions. Culture was maintained for 6 weeks and checked for bacterial contamination using light microscopy at intervals prior to harvest.

3.2.1.2 Plasmid amplification and purification

E. coli DH5α(NEB, UK) used for plasmid amplification were grown in Luria-Bertani (LB) medium. 5 ml of freshly autoclaved media in 50 ml tube supplemented with the appropriate antibiotics i.e. ampicillin (100 µg/ml), spectinomycin (50 µg/ml), chloramphenicol (34 µg/ml), kanamycin (30 µg/ml) and tetracycline (12 µg/ml) was inoculated with a single colony from agar plate and incubated at 37^oC for 12-16 hours. The strains and plasmids used are listed in Table 3.3 and 3.4

Table 3.1 BG11 media recipe

Constituent	Concentration (g/L)
Ammonium ferric citrate	0.006
Calcium chloride dihydrate	0.036
Citric acid	0.006
EDTA	0.001
Magnesium sulphate heptahydrate	0.075
Potassium hydrogen phosphate (K_2HPO_4)	0.04
Sodium carbonate	0.02
Sodium nitrate	1.5
Trace metals solution	1ml

Table 3.2 Recipe for trace metal solution

Constituent	Concentration (g/L)
Boric acid (H_3BO_3)	2.86
Cobalt(II) nitrate ($Co(NO_3)_2.6H_2O$	0.049
Copper sulphate ($CuSO_4.5H_2O$)	0.079
Manganese chloride ($MnCl_2.4H_2O$	1.81
Sodium molybdate ($Na_2MoO_4.2H_20$)	0.390
Sodium orthovanadate (Na_3VO_4)	0.593
Zinc sulphate ($ZnSO_4.7H_2O$)	0.222

Trace metal solution was made by dissolving in distilled water, pH adjusted to 7.4 and made to 1 litre.

Table 3.3 List of strains used

Strains	Description/genotype	source
Nostocsp. PCC 7120	Wild-type strain	ATCC 27893
E. coli DH5a	*huA2 lac(del)U169 phoA glnV44 Φ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs
E.coli BL21 (DE3)	*F⁻ompT hsdS_B(r_B⁻m_B)gal dcm* (DE3)	New England Biolabs

Table 3.4 Plasmids

Plasmid	Description	source
pCDFDuet-1	Cloning vector, CloDF13, two T7lac promoters, two MCS regions, and a single T7 terminator. lacI gene Sm^r	Novagen
pETDuet-1	Cloning vector, ColE1 (pBR322),two T7lac promoters, two MCS regions, and a single T7 terminator. lacI gene Amp^R	Novagen
pCDFDuet-NS1	*NS1* amplified from *Nostoc sp*. and cloned into *NcoI* and *NotI* sites of pCDFDuet-1	This study
pCDFDuet-NStag	*NS1* amplified from *Nostoc sp*. and cloned into *BamHI* and *NotI* sites of pCDFDuet-1	This study
pCDFDuet-P450	*P450* amplified from *Nostoc sp*. and cloned into *NcoI* and *NotI* sites of pCDFDuet-1	This study
pCDFDuet-P450tag	*P450* amplified from *Nostoc sp*. and cloned into *BamHI* and *Not*I sites of pCDFDuet-1	This study
pETDuet-NS1	*NS1* amplified from *Nostoc sp*. and cloned into *NcoI* and *NotI* sites of pETDuet-1	This study
pETDuet-NStag	*NS1* amplified from *Nostoc sp*. and cloned into *BamHI* and *NotI* sites of pETDuet-1	This study
pETDuet-P450	*P450* amplified from *Nostoc sp*. and cloned into *NcoI* and *NotI* sites of pETDuet-1	This study
pETDuet-P450tag	*P450* amplified from *Nostoc sp*. and cloned into *BamHI* and *NotI* sites of pETDuet-1	This study
pCDFDuet-NS-P450	*NS-P450* amplified from *Nostoc sp*. and cloned into *NcoI* and *NotI* sites of pCDFDuet-1	This study
pETDuet-NS-P450	*NS-P450* amplified from *Nostoc sp*. and cloned into *NcoI* and *NotI* sites of pETDuet-1	This study
pET-NS1-P450NS	*NS1 and P450* inserted into *NcoI-PstI* and *NdeI-EcoRV* sites respectively of pETDuet-1	(H. Harada et al., 2011)
pBbA5c-RFP	RFP cloned into pBb plasmid using standard BglBrick cloning strategy.	(Lee et al., 2011)
pBbA5c-MevT(CO)-T1-MBIS(CO, ispA)	Mevalonate pathway genes from *S. cerevisiae* cloned into the pBb plasmid using standard BglBrick cloning strategy	(Peralta-Yahya et al., 2011b)
pBBR1-MCS2	Cloning vector, Kan^R	(Kovach et al., 1995)
pBb	Cloning vector, Cm^r	(Lee et al., 2011)

3.2.2 Genomic DNA extraction and purification

To extract genomic DNA from *Nostoc sp.* PCC 7120 culture, the 100 ml culture was divided into two 50 ml aliquots in sterile tubes and centrifuged to pellet the cells at 4000 rpm at 4°C for 10 minutes in a Multifuge 3 refrigerated centrifuge (Haraeus). Supernatant was decanted leaving approximately 200 µl of cell pellet. Cell pellet was washed 3 times by resuspending in 1 ml of secret wash buffer (50 Mm Tris-HCl, pH 7.4, 100 mM EDTA, pH 8.0, 25% sucrose) and transferred to a sterile 2 ml tube and vortexed for 30 seconds. Pellet was again resuspended in 0.5 ml of resuspension buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA) to which 0.6 g of 0.4-0.6 mm diameter glass beads (Sigma, UK), 25 µl of 10% sodium dodecyl sulphate (SDS), 250 ml of phenol and 250 ml of chloroform were added. This was then chilled on ice for 1 minute and vortexed for 30 seconds and chilled on ice again for another 1 minute. This was repeated for 4 cycles and then centrifuged at 14000 rpm for 10 minutes at 4°C. The top aqueous layer was carefully aspirated into a fresh sterile 2 ml tube and 500 µl of chloroform was added and vortexed to mix. The top aqueous layer was aspirated into a new sterile 2 ml tube and DNA precipitated by adding 1/10 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of 100% ice-cold ethanol and incubated at -20°C for 30 minutes. The mixture was then centrifuged at 14000 rpm for 10 minutes at 4°C to pellet the DNA. Supernatant was carefully aspirated and the pellet DNA was washed with 500 µl of ice cold 70% ethanol, centrifuged and carefully aspirated. The DNA was allowed to dry on the bench after which it was resuspended in 100 µl of nuclease free water (Sigma, UK). The concentration of DNA was measured using a spectrophotometer (Ultraspec 2100 pro UV/Visible spectrophotometer). Quantified DNA was diluted to make 100 ng/µl concentration and further purified using a MoBio Ultraclean microbial DNA isolation kit, (MoBio, Carlsbad, CA, USA) according to the manufacturer's protocol as follows: 200 µl of 100 ng/µl DNA solution was loaded to a sterile 2 ml tube and 900 µl of MD3 was added and mixed by vortexing for 5 seconds and then loaded onto a spin filter and centrifuged at 13000 rpm for 30 seconds. The spin filter was washed with 300 µl of solution MD4 and centrifuged at 13000 rpm for 30 seconds. DNA was eluted from the spin filter using 50 µl of MD5 into a sterile 1.5 ml tube by centrifuging at 13000 rpm for 30 seconds. DNA was again measured using spectrophotometer and diluted to100 ng/µl with nuclease free water. 1 µl of RNAse A was added and left at room temperature for 1 hour. DNA was verified by running in a 0.8% agarose gel.

3.2.3 Agarose gel electrophoresis

DNA was frequently analyzed by running on a 0.8% agarose gel containing 0.0015% (v/v) ethidium bromide solution (Sigma, UK). DNA was resuspended in 6x DNA loading buffer (NEB, UK) and run at 80V using 1x TAE buffer until the dye front had travelled ¾ of the gel. DNA was visualized under UV light in a gel dock.

3.2.4 PCR Primer design

The oligonucleotide primers for the Polymerase Chain Reaction (PCR) were designed using Clone manager software version 9 (http://www.scied.com/pr_cmbas.htm. The primers were designed in such a way as to avoid possible primer dimers, significant hairpin loop formation, secondary priming site in template and low specific binding at the 3' end. Also GC content was aimed at 45-60% where possible, melting temperature in the range of 55-65°C and temperature difference between primer pairs within 5°C of each other in most cases. These primers were synthesized by Integrated DNA Technologies (Leuven, Belgium) and supplied at 25 nmole with standard desalting. 100 µM stock solutions were made using nuclease free water (Sigma, UK) and stored at -20°C. Working concentrations of 10 µM were used for PCR mix. The primers used are listed in Table 3.5.

Table 3.5 List of primers used for PCR

Name of primer	Primer sequence	Restriction site	Tm of primer
NSTag_F	5'-GCGGGATCCTATGGAAAAAATTACTTTCCCAAATTTATATTGG-3'	*BamHI*	60.6^{o}C
NS_R	5'-CGC GGCGGCCGCATTCTGGTAACTTTTTCAAGAAG-3'	*NotI*	67.6^{o}C
P450Tag_F	5'-CGCGCGTCCTATGAAATATCAAATACAGAGACCTAATC-3'	*BamHI*	63.5^{o}C
P450_R	5'-GCGGCGGCCGCTACTATGCTGTGAATGTTGTTGAG-3'	*NotI*	68.5^{o}C
NS-P450_F	5'-CGCGCCATGGGGAGGCACAAATGACTTATGG-3'	*NcoI*	65.9^{o}C
ACYCDuetUP1	5'-GGA TCTCGA TCT CCC T-3'	Nil	58.5^{o}C
DuetDOWN1	5'-GATTATGCGGCCGTGTACAA-3'	Nil	55.5^{o}C
pETUpstream	5'-ATGCGTCCGGCGTAGA -3'	Nil	56.7^{o}C
DuetUP2	5'- TTGTACACGGCCGCATAATC -3'	Nil	55.5^{o}C

3.2.5 Polymerase Chain Reaction (PCR)

Genes from *Nostoc sp*. PCC 7120 were amplified from genomic DNA using PCR and cloned into the appropriate plasmid for expression in *E. coli* as seen in Figure 3.1

3.2.5.1 Amplification of Sesquiterpene synthase

Sesquiterpene synthase (NP_488725, alr4685) was amplified using primers NS-P450_F and NS_R. In this instance, the NS-P450_F primer had an added *NcoI* site for ligation into the *NcoI* site of the corresponding plasmid. All PCR components were bought from Sigma and PCR was set up on ice using *Taq* DNA polymerase. The PCR mix had final concentrations of 0.4 µM of each forward and reverse primer, 200 µM of each dNTP, 1x PCR buffer and 2.5 mM of magnesium chloride. To this was added 3 µl of 50 ng/µl DNA template, 0.5 µl of Taq DNA polymerase and nuclease free water to make 50 µl final volumes. The PCR was performed on a Veriti 96 Well Thermal Cycler (Applied Biosystems) after gentle mixing and pulse centrifugation. The PCR parameters were as follows: initial denaturation at $95^{o}C$ for 1 minute followed by denaturation at $95^{o}C$ for 15 seconds. Annealing was done at $63^{o}C$ for 15 seconds and extension at $72^{o}C$ for 30 seconds all for 30 cycles. A final one cycle extension step at $72^{o}C$ for 7 minutes was done and then held at $4^{o}C$.
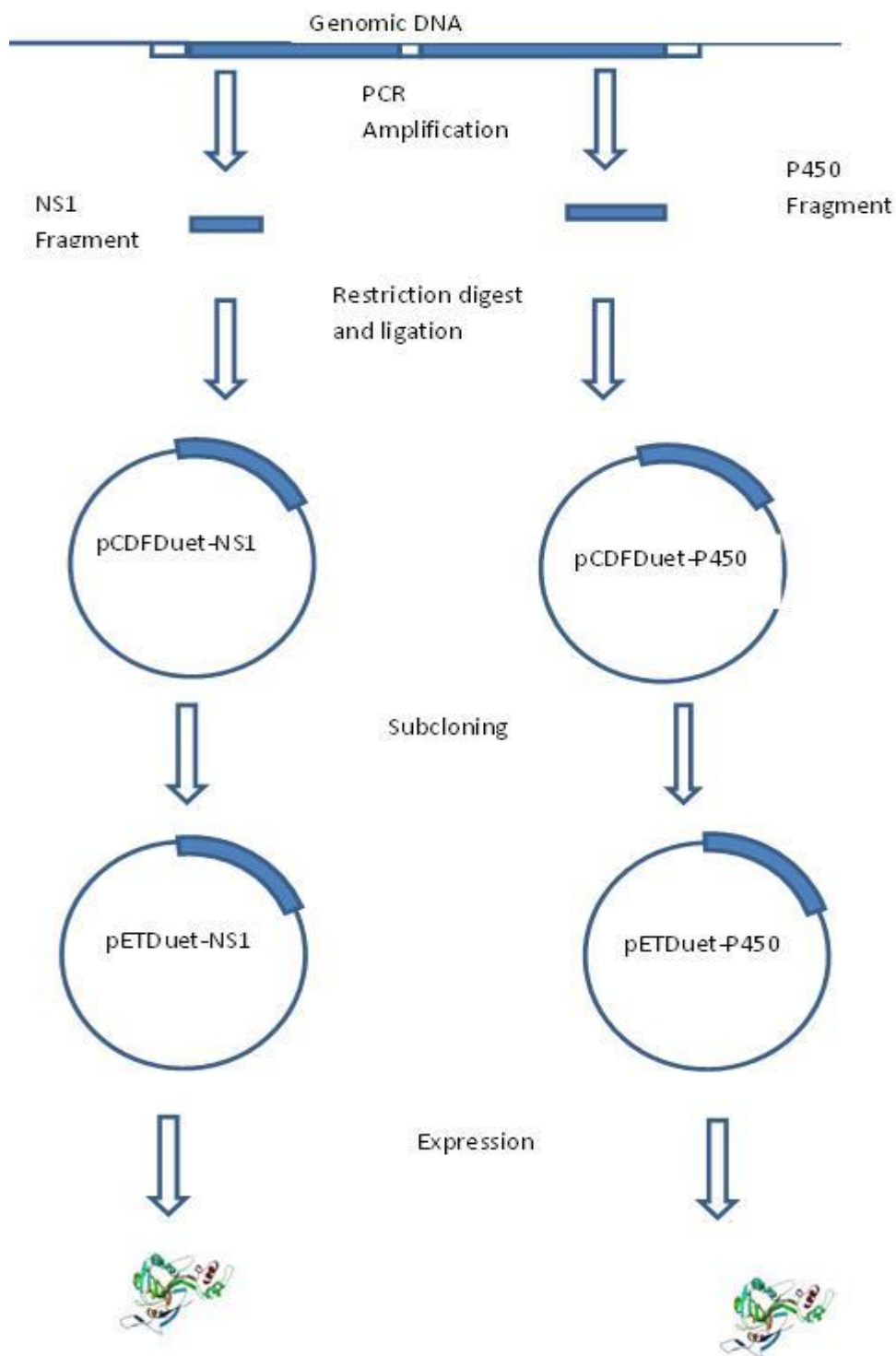
Fig 3.1 Schematic representation of the cloning of terpene synthase (NS1) and cytochrome P450 (P450) genes from *Nostoc* sp. PCC7120

3.2.5.2 Amplification of Sesquiterpene synthase using high fidelity DNA polymerase

PCR reaction was also performed using a Phusion high fidelity DNA polymerase (Finnzymes, UK). The PCR was set up as described in section *3*.2.3.1 above with the following modifications: there was no additional magnesium chloride as the 5X buffer provided 1.5 mM final concentration and 0.02 U/µl final concentration of DNA polymerase was used. The reaction parameters were as described above but denaturation was at 98°C and annealing temperature was at 68.9^{o}C (3°C above the lower Tm primer as recommended by the manufacturer). Following this, a temperature gradient PCR was done with same parameters as above using the following annealing temperatures 55°C, 57°C, 60°C, 65°C, 68°C and 70°C.

3.2.5.3 Amplification of Hexa-histidine tagged sesquiterpene synthase

Sesquiterpene synthase for protein expression and purification was amplified from *Nostoc sp*. PCC 7120 genomic DNA. In order to add an N-terminal six histidine tag to the expressed protein in pCDFDuet-1plasmid (Figure 3.2), a 5' *BamHI* sequence was added to the forward primer. This strategy would utilize and re-establish the *BamHI* restriction site and produce an expressed protein in-frame with the hexa-histidine producing sequence in the plasmid vector. Primers generated were named NSTag_F and NS_R for forward and reverse primers respectively. The PCR was set up as done in section 3.2.3.2 using phusion high fidelity DNA polymerase but with an annealing temperature of 60.6^{o}C.

3.2.5.4 Amplification of Hexa-histidine tagged Cytochrome P450

Cytochrome P450 (NP_488726, alr4686) was amplified using primers P450Tag_F with a *BamHI* restriction site and P450_R and set up as described in section 3.2.3.2 with an annealing temperature of 65.9^{o}C and extension at 72°C for 40 seconds.

3.2.5.5 Subcloning of genes from pCDFDuet-1 into pETDuet-1

Sesquiterpene synthase and cytochrome P450 genes in pCDFDuet-1 (Figure 3.2) parents were sucbloned into pETDuet-1 (Figure 3.3) for co-expression with other plasmids. This was because the terpene synthase (NS1) and P450 were cloned into a pETDuet-1 plasmid (305) the pCDFDuet-1 has a streptomycin resistance marker which was not compatible with some of the other co-expressed plasmids. Since these inserts were cloned into MCS1 of the plasmid (Figure 3.2), MCS1 plasmid primers, ACYCDuetUP1 and DuetDOWN1 were used as forward and reverse primers respectively. The PCR was set up as in section 3.2.3.2 but using 50ng of mutant plasmid with negative control containing wild type plasmid. Annealing temperature was set at 55.5°C for 30 seconds. 5 µl of the reaction was visualized on a 0.8% agarose gel as described in section 3.2.3. The rest was purified using SureClean (Bioline, UK) according to the manufacturer's protocol and quantified spectrophotometrically using an Ultrospec 2100 pro UV/Visible spectrophotometer (Amersham Bioscience).
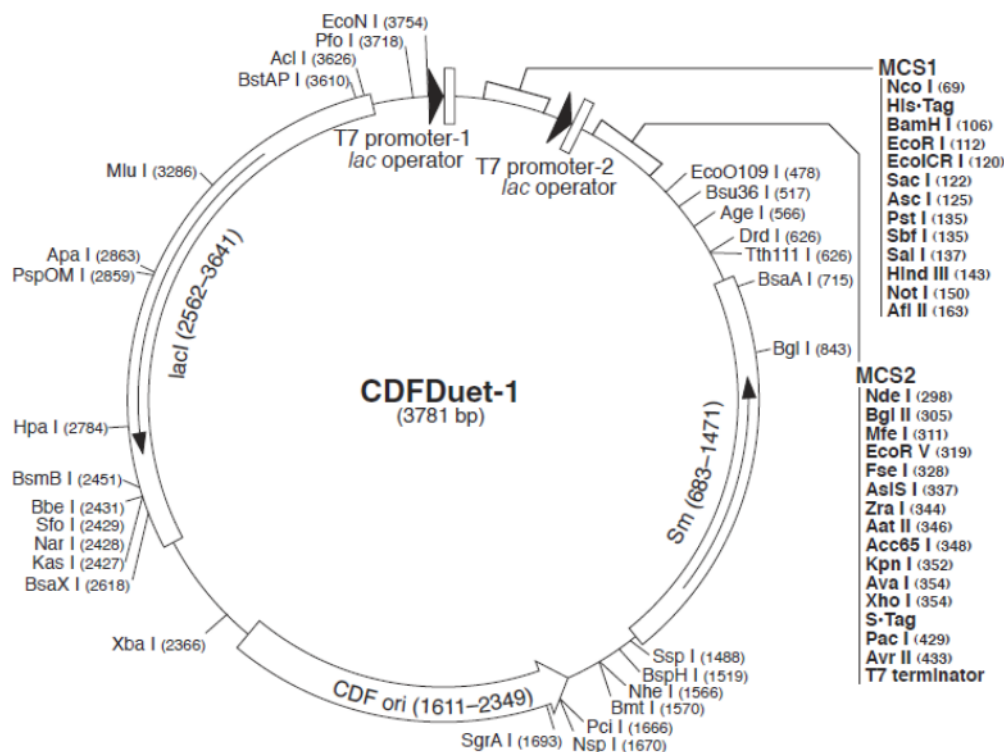
Figure 3.2 Plasmid map of pCDFDuet-1 showing with restriction sites and multiple cloning sites (MCS1 and MCS2). Reproduced from Novagen duet vector user manual.

3.2.5.6 Amplification of terpene synthase and cytochrome P450 (NS-P450)

To amplify the 2.5 kb long adjacent terpene synthase and cytochrome P450 genes from *Nostoc sp*. PCC 7120, magnesium chloride gradient PCR reactions was set up. The PCR mix was set up to give final magnesium chloride concentrations of 0.5mM, 1mM, 1.5mM, 2mM, 2.5mM, 3mM, 4mM and 5mM. Set up was as done as in section 3.2.3.1 with NS-P450_F and P450_R as forward and reverse primers respectively, using Taq DNA polymerase. Annealing temperature was set at 63°C for 30 seconds and extension was 72°C for 1 minute.
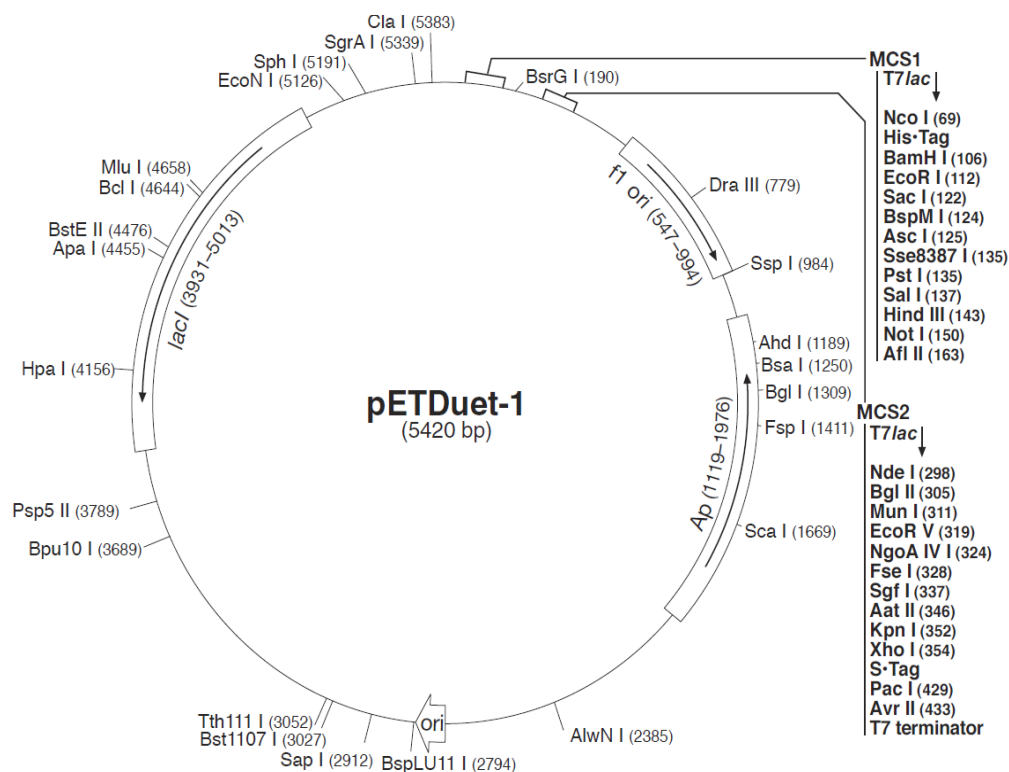
Figure 3.3 Plasmid map of pETDuet-1 showing with restriction sites and multiple cloning sites (MCS1 and MCS2). Reproduced from Novagen duet vector user manual.

3.2.5.7 High fidelity amplification of NS-P450

PCR was also done using high fidelity DNA polymerase as done in section 3.2.3.2 using NS-P450_F and P50_R as forward and reverse primers with annealing temperature set at 65.9^{o}C for 15 seconds and extension at 72°C for 1 minute 30 seconds.

3.2.6 Purification of PCR product

The PCR reactions were purified by either by agarose gel extraction or by using a column free purification method.

3.2.6.1 PCR gel extraction

The entire PCR reaction was separated on a 0.8% agarose gel as described in section 3.2.3. Purification was done by cutting out the correct sized bands using a scalpel blade under minimal low intensity UV light. The excised product was placed in pre-weighed 1.5 ml microcentrifuge tubes and extraction of amplified DNA was done using an Isolate DNA kit (Bioline, UK) according to the manufacturer's protocol. All tubes were then pooled and final concentration and purity measurements done with Ultrospec 2100 pro UV/Visible spectrophotometer (Amersham Bioscience).

3.2.6.2 Column free PCR product purification

For column free PCR product purification, 5 µl of the PCR reaction was removed for agarose gel visualization and the remaining 45 µl was purified using SureClean Plus reagent (Bioline, UK) according to the manufacturer's protocol. The recovered PCR products were dissolved in nuclease free water and concentration and purity measurements done using an Ultrospec 2100 pro UV/Visible spectrophotometer (Amersham Bioscience).

3.2.7 Digestion and purification of PCR product

Sticky ends for ligation into plasmid were produced by double restriction enzyme digest. The double digest in each case was done using high fidelity versions of the enzymes (New England Biolabs, UK). About 500 ng of purified PCR product was digested in a 50 µl final volume using 20 units of each enzyme and compatible buffer in each case. The buffer compatibility for the double enzyme digest was obtained through the New England Biolabs online buffer chart tool at http://www.neb.com/nebecomm/tech_reference/restriction_enzymes/buffer_activity_restriction_enzymes.asp#.UNtzWoXN5kg. Each digest was supplemented with 0.5 µl of BSA mixed and incubated at 37°C for 1 hour. The reaction was then purified to remove restriction enzymes and to concentrate the doubly digested PCR fragments using SureClean (Bioline, UK) according to the manufacturer's protocol and pellet dissolved in 20 µl of nuclease-free water (Sigma).

3.2.8 Digestion of Plasmid and purification

The double digestion was set up in PCR tubes on ice using 500 ng of plasmid. The set up was mixed and incubated at 37°C for 1 hour. A small digest using each of the enzymes separately were also set up to check for linearization of the plasmid The digest was purified by SureClean (Bioline, UK) using the manufacturer's protocol. 2.5 µl of the purified double digest and the linearization reactions were run on a 0.8% agarose gel as previously described.

3.2.9 Dephosphorylation of digested plasmid

The doubly digested plasmid was dephosphorylated prior to ligation reaction using Calf Intestinal Phosphatase (NEB, UK) according to the manufacturer's protocol and incubated at 37°C for 1 hour. The mixture was purified using SureClean and dissolved in 20 ul of nuclease-free water (Sigma, UK). The concentration of the doubly digested,

dephosphorylated and purified plasmid was measured using an Ultrospec 2100 pro UV/Visible spectrophotometer (Amersham Bioscience).

3.2.10 Ligation reactions

Ligation of purified PCR products into their respective plasmid was done using T4 DNA ligase (NEB, UK) by using different ratios of vector to insert ranging from 6:1 to 1:3. 50 ng of the plasmid vector was used in each instance and reactions were set up on ice in PCR tubes. The amount of insert needed in each ligation was worked out according to the following formula:

$$\frac{concentration\ of\ vector\ \times size\ of\ vector}{size\ of\ plasmid} \times ratio\ of\ vector\ to\ insert$$

The tubes were mixed and incubated at room temperature for 1 hour. The ligase enzyme was inactivated by heating at $65^{o}C$ for 10 minutes. To check the background of spontaneous re-ligation/uncut plasmid, a control using the digested and dephosphorylated plasmid with dH_2O in place of an insert was employed.

3.2.11 Ligation PCR

To ensure that transformation was done using only successfully ligated plasmid, a ligation PCR was done as a screening test by using 2 µl of the ligation reaction. A time saver MyTaq^{TM} Red DNA Polymerase and buffer (Bioline, UK) was used. The 5 x buffer already contained dNTPs, MgCl2 and loading dye for direct loading after the PCR. A master mix for eight reactions including negative and positive controls was set up and 48 µl of the master mix was aliquoted into PCR tubes and 2 µl of each ligation reaction was added, 2 µl of nuclease free water was used for negative control and 2 µl of uncut plasmid vector (pCDFDuet-1/pETDuet-1) was used as a positive control. The forward and reverse primers were the plasmid specific primers for MCS1 (Figure 3.2 and 3.3) which are ACYCUP1/pETUPStream and DUETDOWN1 respectively. The tubes were mixed and loaded into the PCR machine. Program was as follows: initial denaturation at $95^{o}C$ for 2 minutes and subsequent denaturation at $95^{o}C$ for 15 minutes for 30 cycles. Annealing temperature was set at $53^{o}C$ for 15 seconds and extension was $72^{o}C$ for 1 minute for 30

cycles. A final extension at 72°C for 4 minutes and held at 4°C. 10 µl of the reaction was separated on a 0.8% agarose gel as described in section 3.2.3

3.2.12 Transformation into cloning strain

Based on the result of the PCR screening test for the ligation reaction, transformation using positive ligation reactions were set up using 1 µl of the ligation reaction into 50 ul of chemically competent DH5α cells (NEB, UK) according to the manufacturer's instructions. 1 µl of double digested plasmid was also used to transform to check for background (plasmid re-ligation and uncut plasmid) and a negative control.

3.2.13 Colony screen PCR

Colony PCR was used to screen colonies for positive clones containing the gene insert prior to sequencing. This was done by making a master mix for several reactions (10-30) with the time saver MyTaq^{TM} Red DNA Polymerase and buffer (Bioline, UK). The master mix had 10 µl of 5 x buffer, 2 µl of each primer, 1 µl of polymerase and 33 µl of nuclease-free water. This was mixed and 48 µl aliquoted into each PCR tube. Colonies were picked with sterile tips, dipped in labelled PCR tube containing 20 µl of nuclease-free water and same tip was used to inoculate 5 ml LB broth in correspondingly labelled 50 ml falcon tubes. The inoculated PCR tubes were heated to 95°C for 10 min, spun at 5, 000 x g for 5 minutes and 2 µl of supernatant used as DNA template for PCR reaction. The 50 ml tubes were incubated at 37°C while the PCR tubes were set up. The PCR reactions had a negative control of nuclease free water in place of DNA and a positive control of wild type plasmid vector. The PCR program was run as follows: initial denaturation at 95°C for 2 minutes and subsequent denaturation at 95°C for 15 minutes for 30 cycles. Annealing temperature was set at 53°C for 15 seconds and extension was 72°C for 1 minute for 30 cycles. A final extension at 72°C for 4 minutes and held at 4°C. The PCR reactions were separated on a 0.8% agarose gel and visualized as described in section 3.2.3.

3.2.14 Plasmid mini-prep

The culture tubes corresponding to positive clones from the colony PCR screen were retained and the negative ones were discarded. After overnight growth, the cells were harvested by centrifuging at 5, 400 x g for 10 min at 4°C and supernatant discarded. Plasmid DNA was extracted from the cell pellet using a QIAGEN plasmid mini prep kit according to the manufacturer's protocol. DNA quantification was done directly by UV spectrophotometry and through agarose TAE gel estimation using hyperLadder I (Bioline, UK).

3.2.15 DNA sequencing

20 μl of purified plasmid vector containing the cloned inserts from all positive clones were sent for sequencing at the University of Sheffield Medical School Core Genetic facility. 20 μl of 10 μM forward and reverse primers (insert specific and plasmid specific flanking primers) were used and sequencing was done on an ABI 3730 capillary sequencer.

3.2.16 Analysis of DNA sequence results

The chromatogram of the DNA sequencing results were viewed and manually curated using FinchTV version 1.4 (http://www.geospiza.com/Products/finchtv.shtml). The sequences were analyzed on clone manager software version 9 (http://www.scied.com/pr_cmbas.htm). Inserts were checked for nucleotide deletion, insertion, mutation. The sequences were also checked for translational frame shifts and alignment with the N-terminal hexa-histidine tagging sequence. Insert sequences were then used as query in a BLASTX (Altschul et al., 1997) in both the non-redundant and *Nostoc sp.* 7120 database at NCBI.

3.2.17 Competent cell preparation

The expression of cloned genes was done using *E. coli* BL21 (DE3) and chemically competent Rosetta-gami 2 (Novagen). The *E. coli* BL21 (DE3) cells were made chemically competent using Z-competent^TM *E. coli* Transformation kit (Zymoresearch) according to the manufacturer's protocol.

3.2.18 Transformation

Plasmid containing correct gene insert as well as wild type plasmid vector were transformed into chemically competent *E. coli* BL21 (DE3) and Rosetta-gami 2 (Novagen) according to the manufacturers' protocol.

3.2.19 Glycerol stocks

Single colonies from transformants were picked and used to inoculate 5 ml of LB media containing the appropriate antibiotics in a 50 ml tube. Culture was incubated overnight and the next morning, 200 µl of culture in late exponential phase was placed in a fresh sterile 1.5 ml microcentrifuge tube. 800 µl of sterile 80% glycerol was added, vortexed and flash frozen in liquid nitrogen and stored at -70°C.

3.3 Results

3.3.1 Genomic DNA extraction

Genomic DNA extracted from *Nostoc sp.* PCC7120 biomass was purified using MoBio Ultraclean microbial DNA isolation kit (MoBio, Carlsbad, CA, USA) and the absorbance results of the initial extract compared to the purified DNA seen in Table 3.6. The genomic DNA extracted and purified with the MoBio kit had better purity as seen by the A_{260}/A_{230} and $A_{260/280}$ values despite the significant reduction in concentration.

Table 3.6 Measurements of Genomic DNA extracted from *Nostoc sp.* 7120

Sample	DNA conc (µg/ml)	A_{230}	A_{260}	A_{280}	$A_{260/230}$	A_{320}	$A_{260/280}$
Nostoc sp. 7120	1355	1.841	2.846	1.706	1.59	0.130	1.76
Nostoc sp. 7120 MoBio kit	451	0.394	0.891	0.464	2.23	-0.010	1.90

The significant improvement in A_{260}/A_{230} ratio may be attributed to residual carbohydrate contamination as well as possible phenolic/ethanol contamination used in the extraction process. In all, a significantly good concentration and level of purity was achieved as impurities such as phenol are known to affect PCR amplification negatively (Al-Soud and Rådström, 1998).

3.3.2 PCR Amplification

The required open reading frame from extracted genomic DNA from *Nostoc* sp. PCC 7120 was amplified by PCR. Successfully cloned genes were also amplified by PCR from their parent for subcloning into another host.

3.3.2.1 Sesquiterpene synthase amplification using Taq DNA polymerase

The open reading frame coding sesquiterpene synthase (NS1) from *Nostoc* sp. PCC 7120 was amplified by PCR. Several PCR reactions were done using genomic DNA extract. PCR optimization steps where instituted as a result of failure of initial PCR runs. These optimizations included temperature and $MgCl_2$ concentration gradients as well as BSA and DMSO supplementation as well as changes in dNTP and primer concentrations without success. It was thought that inhibiting factors may be present in the DNA extract as substances such as ethanol, phenol, salts, detergent, EDTA and proteinases are known to inhibit PCR (Davalieva and Efremov, 2010; Eilert and Foran, 2009). As seen in Figure 3.4,

subsequent amplification using genomic DNA from kit purification yielded good product. This confirmed the presence of inhibitory contaminants in the initial DNA extract as addition of the original extract and kit purified DNA did not give any amplification. Following this, all subsequent PCR amplification from genomic DNA was done using the kit purified DNA as template.
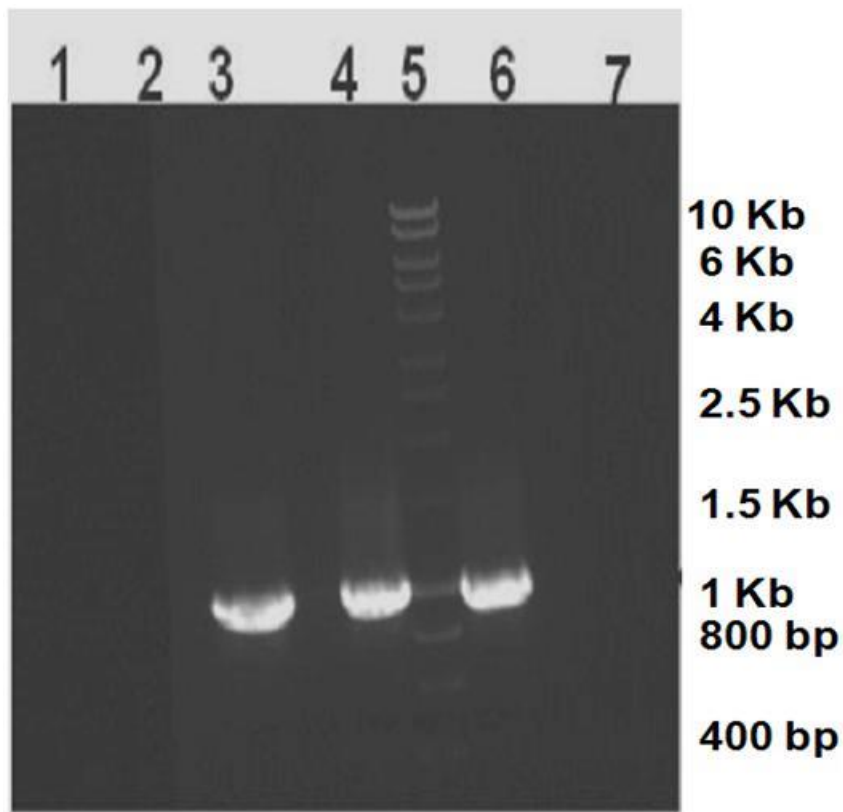
Figure 3.4 Agarose gel visualization of amplified sesquiterpene synthase from genomic DNA of *Nostoc* sp. PCC 7120 with kit purified DNA (3, 4 & 6). Lane 1 and 2 are from unpurified and lane 7 is mix of purified and unpurified DNA extract with no amplicon indicating the presence of PCR inhibitors in the unpurified DNA extract.Lane 5 is DNA size marker.

3.3.2.2 Sesquiterpene synthase amplification using high fidelity DNA polymerase

Sequencing results of amplified genes using Taq DNA polymerase showed significant number of mutations and deletions. For this reason, a high fidelity DNA polymerase was used subsequently in all PCR amplification for cloning purposes.   However, as

recommended by the manufacturer, an annealing temperature that was 3°C (68.9°C) higher than the lower T_m primer (65.9°C) was used for the PCR but yielded no product. A temperature gradient PCR run was then used to ascertain the optimal annealing temperature. As seen in Figure 3.5, amplification was achieved with temperatures as low as 57°C which was almost 9°C lower than the T_m of the lower T_m primer. The optimum annealing temperature was seen to be 65°C which was just about the T_m of the lower primer. The upper annealing temperature for amplification was seen at 68°C.
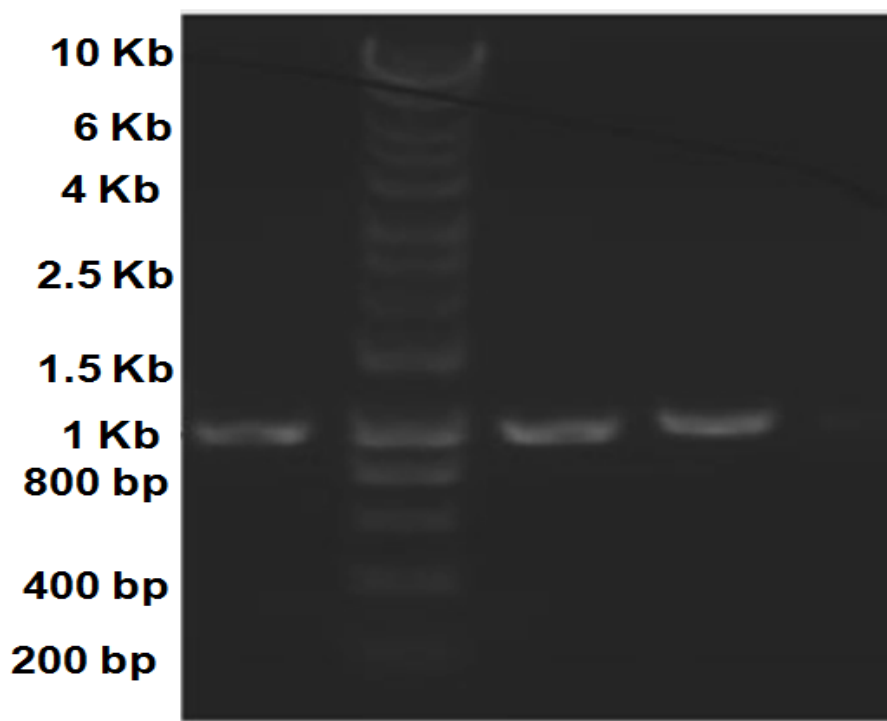
Figure 3.5 Agarose gel electrophoresis showing PCR amplification of NS1 at different annealing temperatures using Phusion high fidelity polymerase. Lane 1 was at 57°C, lane 2 is DNA size marker, lane 3 was at 60°C, lane 4 was at 65°C and lane 5 was at 68°C.

3.3.2.3 Sesquiterpene synthase amplification from pCDFDuet-NS1

Sesquiterpene synthase (NS1) was amplified from the pCDFDuet-NS1 parent for subcloning into pETDuet-1. PCR amplification using pCDFDuet-NS1 as template was therefore done. Since the insert was originally cloned into *NcoI* and *BamHI* restriction sites

in MCS1 of the parent plasmid (Fig 3.2), the PCR primers used were the plasmid specific primers for MCS1. As seen in Figure 3.6, the amplified product was larger than the expected 1 Kb size since the forward primer anneals at a site 168 bp upstream of the insert and the reverse primer anneals 58 bp downstream thereby adding more than 200 bp to the original amplicon. As shown in Figure 3.6, the genomic amplicon is run alongside to show the disparity in size.
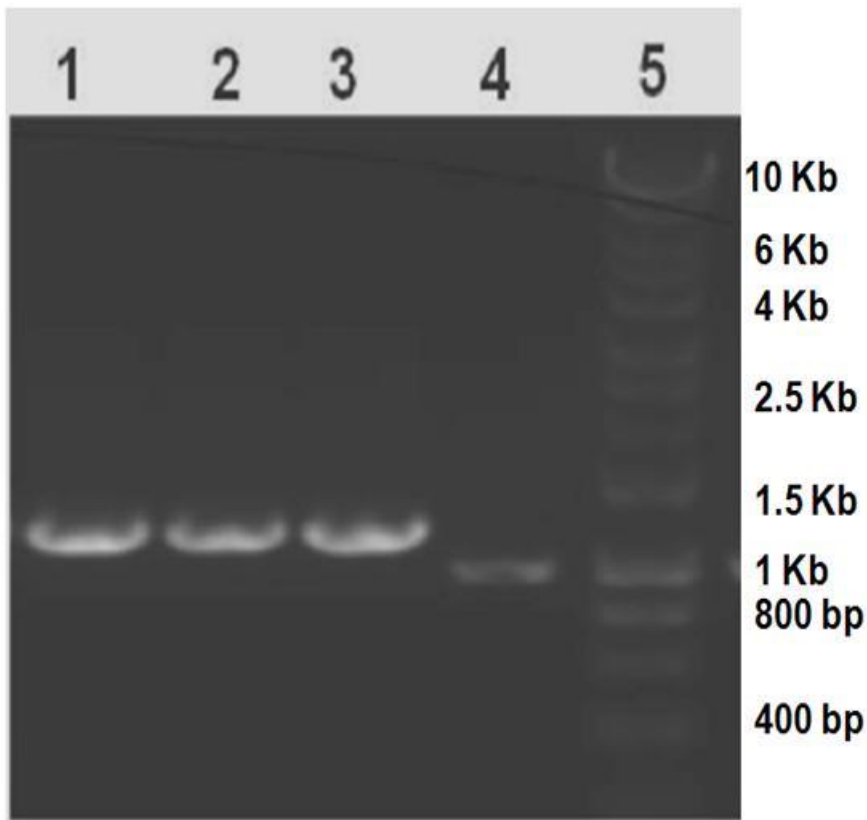
Figure 3.6 Agarose visualization of amplified sesquiterpene synthase from pCDFDuet-NS parent (lanes 1, 2, 3) using plasmid specific primers. The amplified product is slightly larger than 1 Kb due to primers annealing outside the insert and adding about 200 bp to the amplicon. Lane 4 amplified product from *Nostoc sp.* 7120 genomic DNA and lane 5 is the DNA size marker

3.3.2.4 Amplification of cytochrome P450

As discussed in the chapter introduction, the open reading-frame coding for a cytochrome P450 (P450NS1) is located immediately downstream of the sesquiterpene synthase gene. This gene was amplified from kit purified genomic DNA from *Nostoc* sp. PCC 7120 and visualized by running on a 0.8% agarose gel. Figure 3.7 demonstrates the amplification of the 1.4 Kb long genes.
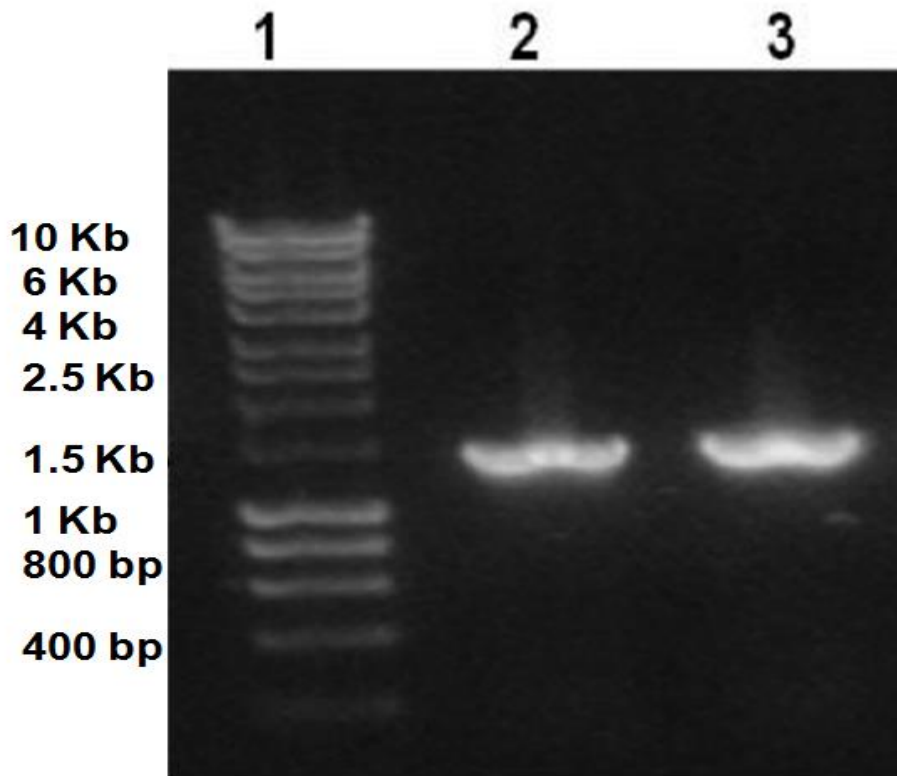
Figure 3.7 shows 1.5 Kb amplified PCR products seen in lanes 2 and 3 that correspond to the size of the cytochrome P450 from Nostoc sp. PCC 7120

3.3.2.5 Amplification of sesquiterpene synthase (NS1) and cytochrome P450 (P450)

The sesquiterpene synthase (NS1) and cytochrome P450 (P450) from *Nostoc sp*. PCC7120 are in an apparent gene cluster in series with each other separated by 21 nucleotide long segment. The two contiguous genes are 2.4 Kb long and Figure 3.8 shows a confirmation of successful amplification using high fidelity DNA polymerase.
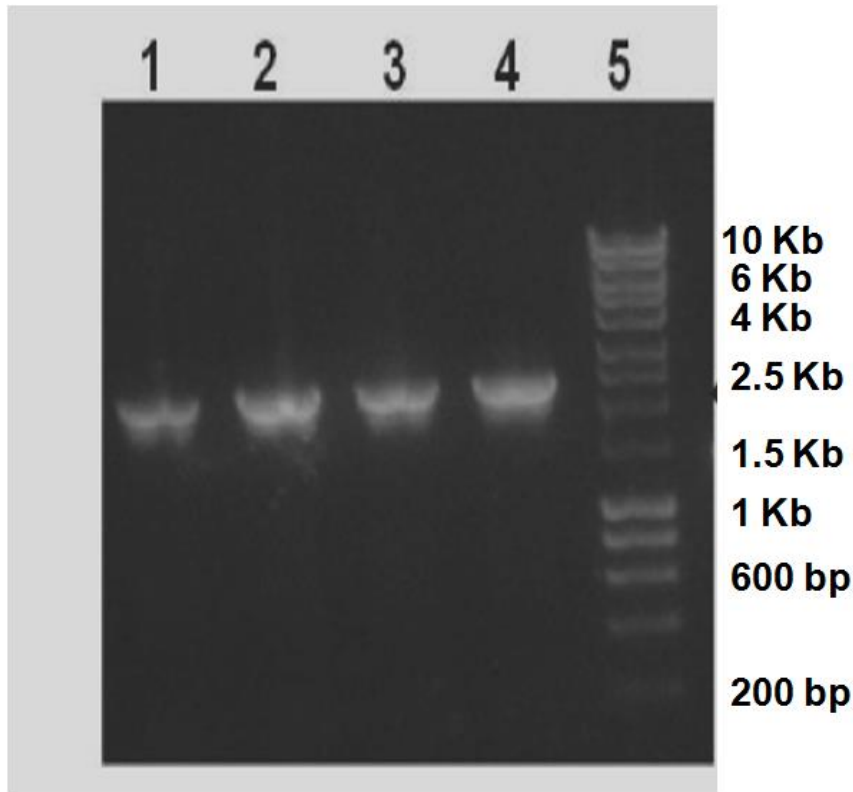
Figure 3.8 Visual confirmation of amplified sesquiterpene synthase plus cytochrome P450 (NS1+P450) from Nostoc sp. PCC 7120 are seen in lanes 1-4 corresponding to the expected 2.5 Kb. Lane 5 is DNA size marker.

3.3.3 Restriction enzyme digestion of plasmid

Plasmid digestion was done and the digest was visualized on an agarose gel and compared to the undigested plasmid as seen in Figure 3.9
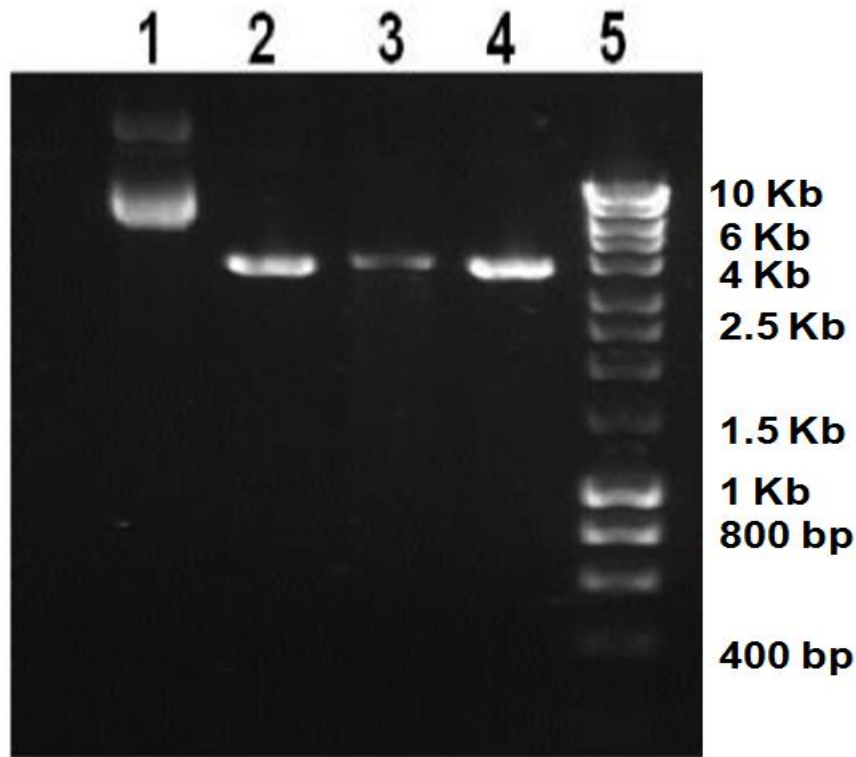
Fig 3.9 Agarose gel view of restriction enzyme digest of pCDFDuet-1 showing the uncut plasmid in lane 1 compared to the one linearised with *NotI* (lane 3) and double cut (lane 2 cut with *BamHI/NotI* and lane 4 cut with *NcoI/NotI*). There is a slight difference in size between the linearised and double cut plasmid

Evidence of successful double digestion of plasmid is revealed by the slight difference in size between the doubly digested plasmids as seen in lanes 2 and 4 compared to the linearized plasmid in lane 3 which was singly cut by *NcoI*. This was necessitated by the fact that the double plasmid digest did not produce a DNA segment long enough for accurate visualization on the gel used.

3.3.4 Restriction enzyme digestion PCR products

An indirect method of validating digestion of PCR products was by digesting the PCR products with plasmid. Figure 3.10 shows evidence of digestion of plasmid pCDFDuet-1

with *NcoI* and *PacI* as shown by the 400 bp fragment in the presence of the 2.4 Kb NS+P450 PCR fragment.
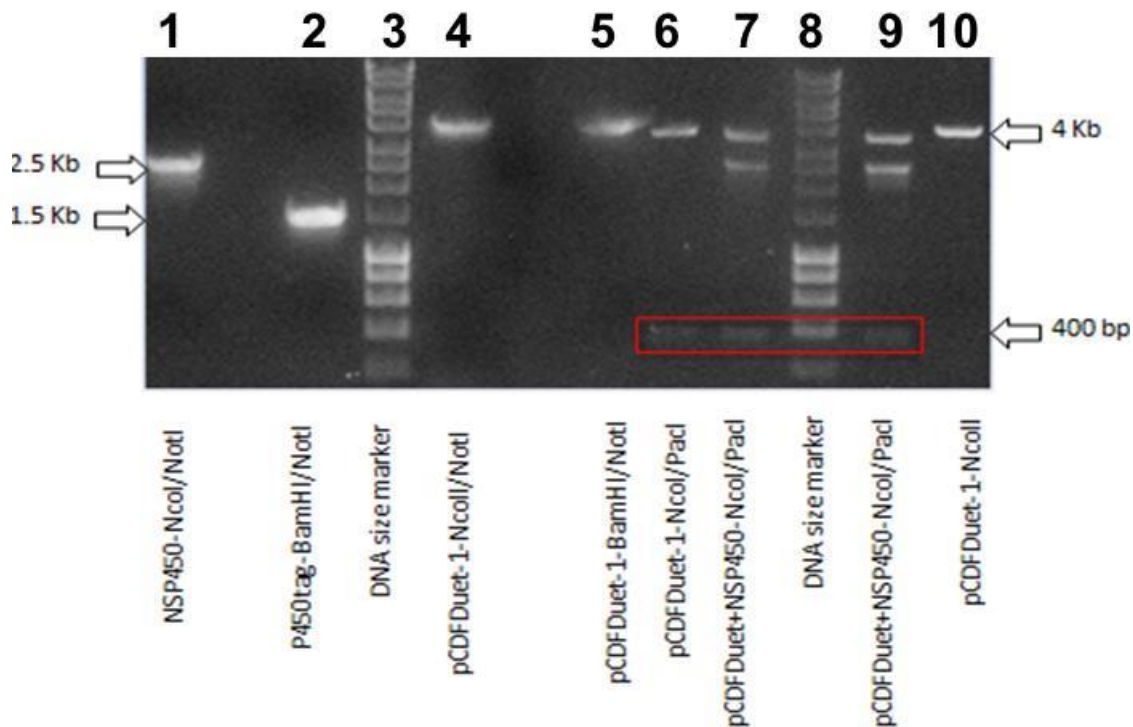
Fig 3.10 Agarose gel visualization of enzyme digestion of pCDFDuet-1 and PCR products (NS+P450 and P50tag). Lane 1 is NS + P450 digested with *NcoI* and *NotI*, lane 2 is P450tag digested with *BamHI* I and *NotI*. Lanes 6, 7 and 9 were used to test for efficacy of digestion of PCR products.

However, PCR products amplified from pCDFDuet-1 parent for subcloning into pETDuet-1 were easily confirmed on agarose gel electrophoresis. As shown in figure 3.11, the PCR product amplified from MCS1 using plasmid primers added more than 150 bp to the 5' end and is easily visualized. This digested product also showed size difference with the undigested product.
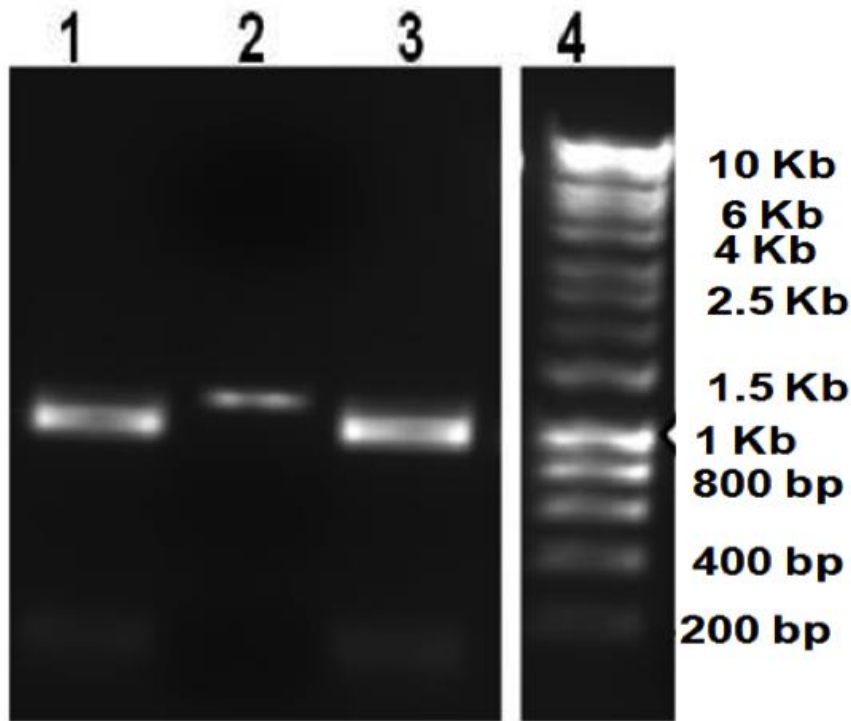
Fig 3.11 Agarose gel visualization of PCR products (NS1) amplified from pCDFDuet-NS parent using MCS1 plasmid primers and digested by *NcoI* and *NotI* (lanes 1 and 3) and undigested product (lane 2). The digested PCR product shows bands of about 150 bp confirming digestion

3.3.3 Ligation PCR screen

In order to ensure that only successful ligations were transformed and thus reduce the number of colonies to be screened as well as check the level of undigested plasmid background, ligation PCR was done as an additional screening tool. The result as seen in Figure 3.12 shows PCR amplification of insert from ligation mix containing insert to vector ratios of 6:1, 3:1, 1:1and 1:3. The figure shows that there was successful ligation in all cases as seen by the amplified PCR product corresponding to the desired insert. Significantly also, there was a 200 bp amplicon in each case which corresponds to amplification of the uncut or re-ligated plasmid, which represents the level of background in each case. There was no significant difference in intensity of the 200 bp amplified band

but there was obvious difference between NS +P450 and the P450tag. This possibly implies much less successful ligation between vector and insert for NS+P450. Additional evidence of differences in successful ligation is seen in the negative control (uncut wild-type plasmid) as seen in lanes 5 and 13 which shows the same level of intensity of the 200 bp band which are just as intense as that from NS+P450 ligation. The use of plasmid specific primers ensured that only successful insert of the right size will be at the 1.5 kb and 2.5 Kb marker lanes.
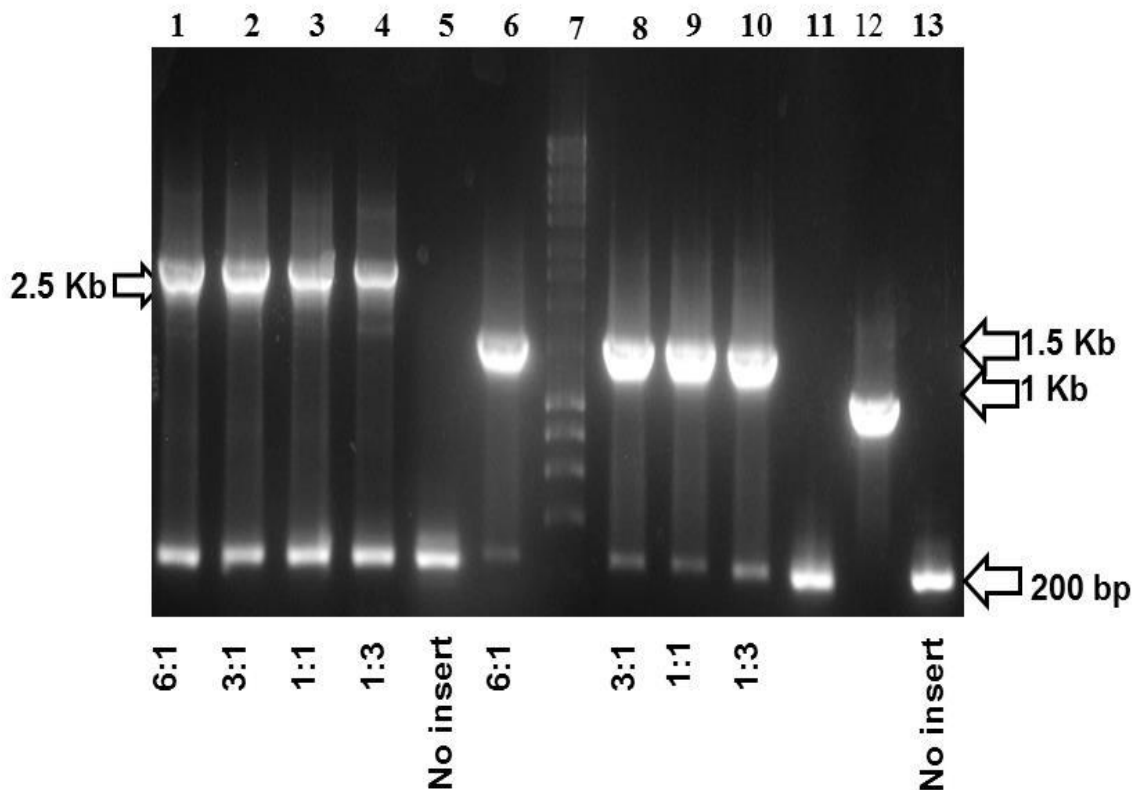
Fig 3.12 Agarose gel visualization of PCR of different ratios of ligation mix for pCDFDuet + P450tag and pCDFDuet + (NS + P450). Lanes 1 to 4 are for pCDFDuet + (NS+P450) and lanes 6 to 9 are for pCDFDuet + P450tag. Lanes 5, 11 and 13 are negative controls and lane 12 is a positive control using pCDFDuet-NStag

3.3.4 Colony PCR screen

Colony PCR was employed as a further screening tool to select clones of transformed cells that had the correct gene insert before subsequent sequencing. Several colonies had to be

screened before getting any positive clone. This indicated a very high level of background of cells harbouring plasmids that were not completely digested or have spontaneous re-ligated without the required insert. As seen in Figure 3.13, about 25% of the clones screened had the desired insert and this was significantly increased when the plasmid was dephosphorylated before ligation.
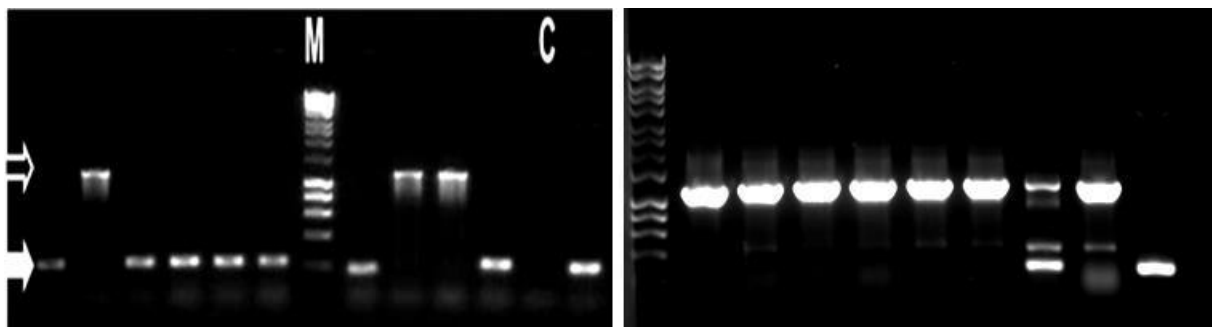
Fig 3.13 colony PCR screen for sesquiterpene synthase of bacteria transformed with a 3:1 ligation mix (insert: vector) comparing ligation with un-dephosphorylated to a dephosphorylated plasmid. M and C are DNA size marker and negative control respectively with top and bottom arrows showing positive and negative clones respectively.

Additionally, the ligation ratio was seen to have an effect on the nature of the insert. Figure 3.15 and 3.16 are colony screen PCR of bacteria transformed with 6:1 and 3:1 ratio of insert to vector. While Figure 3.15 shows a very high percentage of successful ligation, an unexplained band of about 700 bp was seen in all of the positive colonies except one. On the other hand, the 3:1 ligation ratio produced just as good successful ligation but the 700 bp band was seen in just one clone of those screened.
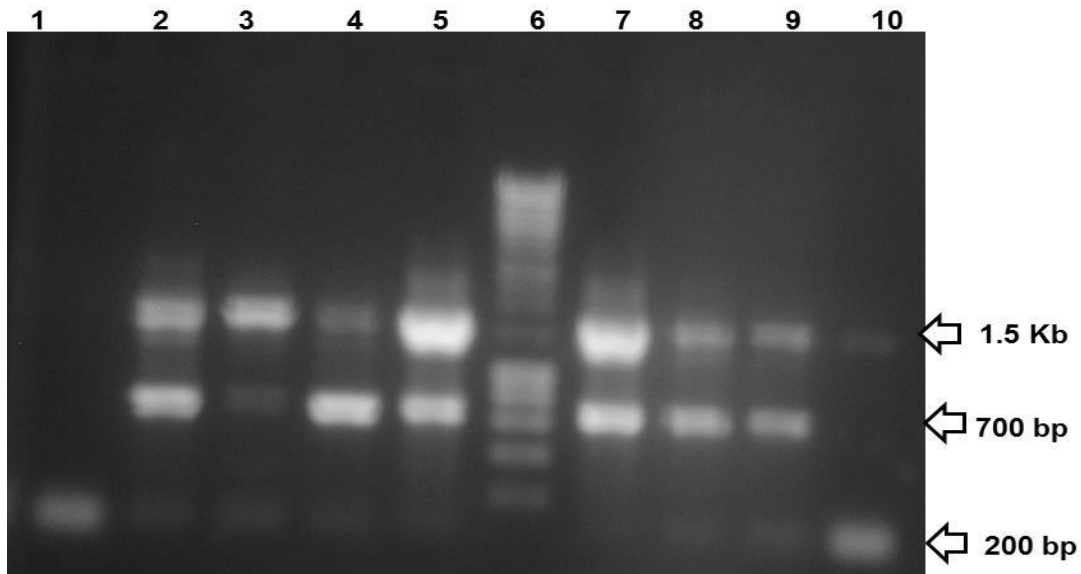
Fig 3.15 Colony PCR screen of bacteria transformed with 6:1 ligation mix (pCDFDuet + P450tag). Notice that all of the lanes except lane 1 had positive clones but in addition to the expected 1.5 Kb band they all had an unexplained band corresponding to 700 bp.Lane 10 did not contain the unexplained 700 bp band. Lane 6 is a DNA size marker
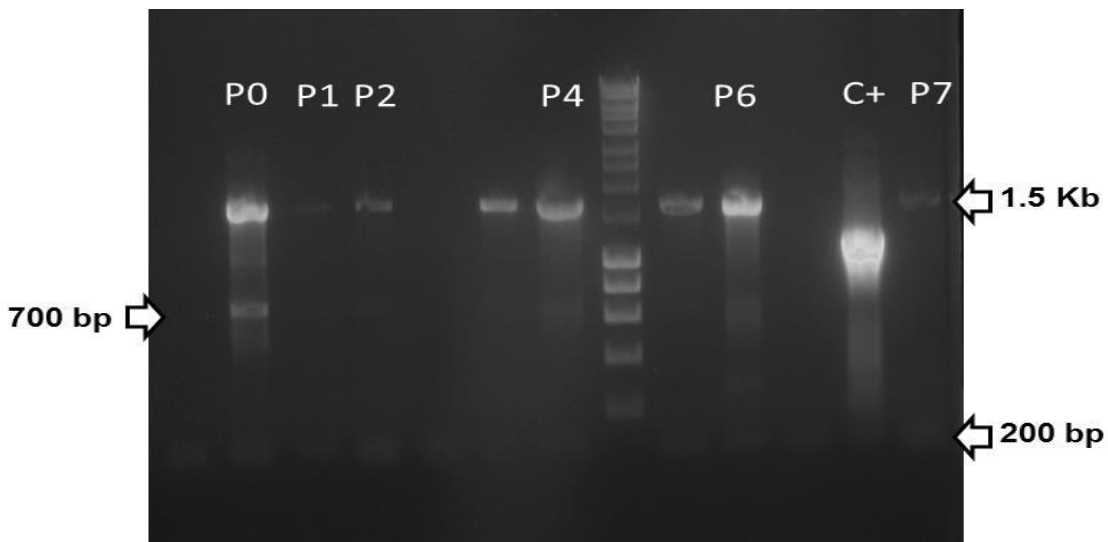
Fig 3.16 Colony PCR screen of bacteria transformed with a 3:1 ligation mix (pCDFDuet + P450tag) showing 7 positive clones (P1 to P7) and one positive clone (P0) with an unexplained 700 bp amplicon. C+ is a positive control using verified clones of pCDFDuet-NStag

The colony PCR screen of the bacteria transformed with 3:1 ratio of ligation mix for 2.5 Kb insert showed a greatly reduced number of positive clones. As seen in Figure 3.17, just two positive clones were identified in a screen of twenty colonies which is just about 10%. This result is in keeping with the ligation screen PCR result seen earlier in Figure 3.12.
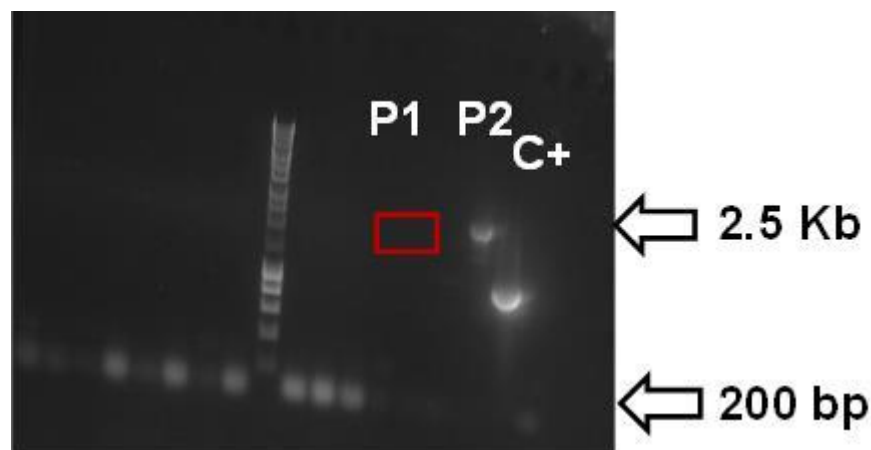
Fig 3.17 Colony PCR screen of bacteria transformed with 3:1 ligation mix (pCDFDuet + NSP450NS) showing just two positive clones (P1 and P2) with P1 as a very faint band. Notice that the entire negative clones have amplicons of 200 bp but the positive clones do not. C+ is a positive control from sequence verified pCDFDuet-NS clone

3.3.5 Analysis of DNA sequencing results

Analysis of DNA sequence was done using FinchTV version 1.4.0 and clone manager version 9 (http://www.scied.com/pr_cmbas.htm. Genes amplified with Taq DNA polymerase were seen on analysis to have synonymous and non-synonymous single nucleotide mutations as well deletions. For this reason, high fidelity DNA polymerase was used and results showed significant reduction in mutations. Further analysis using clone manager enabled identification of clones not only with the right insert but for the hexa-histidne tag to be in-frame with the protein when expressed. Figures 3.18 and 3.19 shows restriction maps for the sesquiterpene synthase and the cytochrome P450 respectively.

Figure A1 and A2 shows annotated plasmid maps with inserts in-frame with the respective N-terminal hexa-histidine sequence with the chromatogram traces seen in Figure A3 and A4 respectively.

In-silico amino acid translation of the sequences was done and further confirmed the choice of the right clones used for further work. As seen in Figures 3.20 and 3.21, the frame 1 translation of the cytochrome P450 and sesquiterpene synthaseare seen to produce the desired protein in-frame with the required histidine tag.
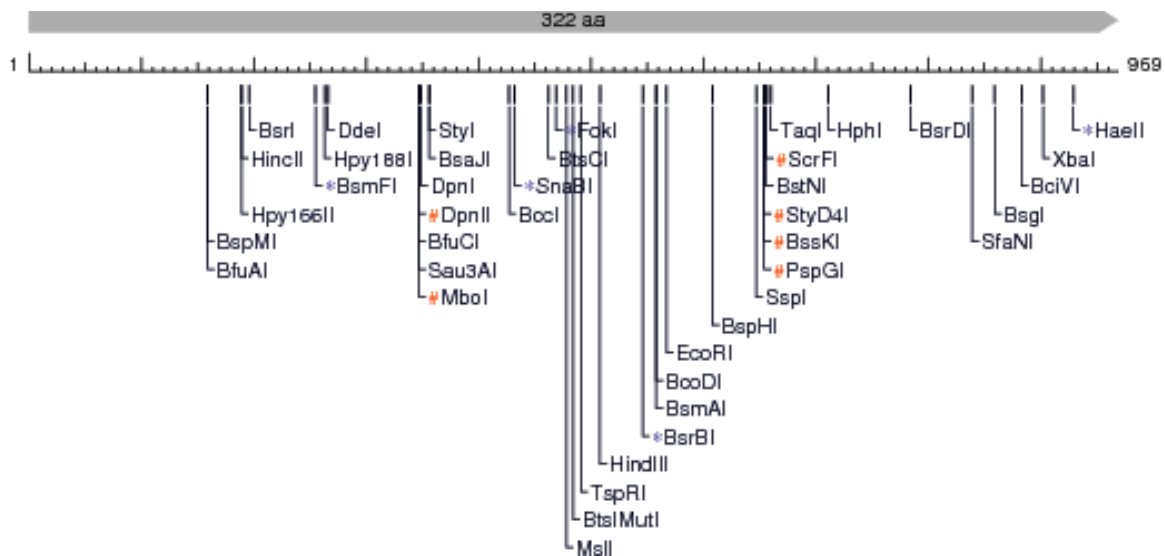
Fig 3.18 Restriction map of terpene synthase from *Nostoc* sp PCC 7120 generated using Nebcutter V2.0
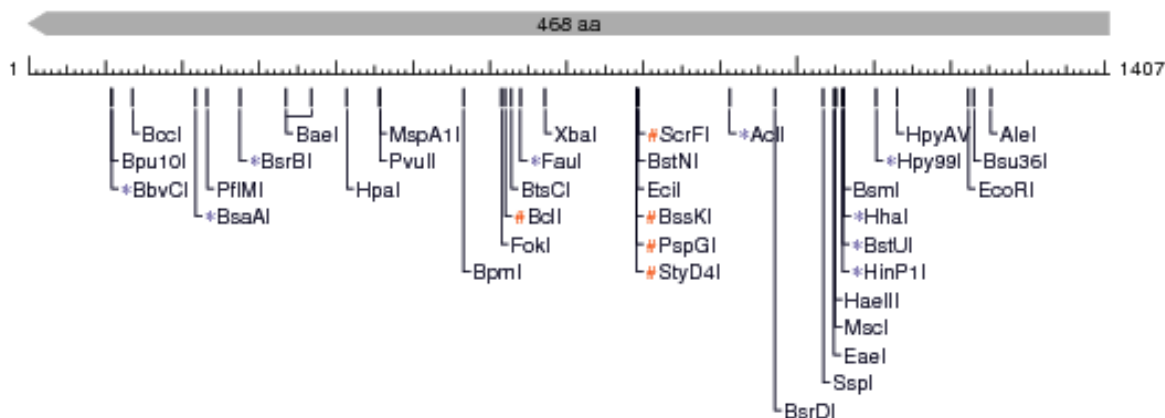
Fig 3.19 Restriction map of cytochrome P450 from *Nostoc* sp. PCC 7120 generated using Nebcutter V2.0

M   H   m

⇓   ⇓   ⇓

```
  1   agatatacca tgggcagcag ccatcaccat catcaccaca gccaggatcc tatgaaatat-
       r    y    t    m    g    s    s   h    h    h    h    h    s    q    d    p    m    k    y

 61   caaatacaga gacctaatcc actaaaaact caccctttcc tgcaaaaact acaatggata
       q    i    q    r    p    n    p    l    k    t    h    p    f    l    q    k    l    q    w    i

121   gctgaccctg tagaatacat gaaaaaagca tctctgcaac atccgatat gtttacaçca
       a    d    p    v    e    y    m    k    k    a    s    l    q    h    p    d    m    f    t    a

181   gaggtaattg gttttggtga cactgtagtg tttgttagcc accctcaggg aatccagaca
       e    v    i    g    f    g    d    t    v    v    f    v    s    h    p    q    g    i    q    t

241   cttttttgcta atgacagaaa aaagttagta gctgttgggg aagcaaacag aattttgtat
       l    f    a    n    d    r    k    k    l    v    a    v    g    e    a    n    r    i    l    y

301   cctttagtgg gtaacaattc tatgttccta ctagaaggag taaaacacaa gcagcgaaga
       p    l    v    g    n    n    s    m    f    l    l    e    g    v    k    h    k    q    r    r

361   caactcctta tgccctcctt tcatggggaa agaatgcgcg agtatggcca cttaatcaga
       q    l    l    m    p    s    f    h    g    e    r    m    r    e    y    g    h    l    i    r

421   aatattactg aaaacctttt tagtcagcta caacaggatg ttacttttttc cgctctcaca
       n    i    t    e    n    l    f    s    q    l    q    q    d    v    t    f    s    a    l    t

481   gcaatgcggg aaatctcgat gcaggttatt ttgcaagctg tctttggttt ttatgaggga
       a    m    r    e    i    s    m    q    v    i    l    ç    a    v    f    g    f    y    e    g

541   gaacgttgtc aacaattcaa acatttatta ccaattttttc tttcagaatt atttcagtcc
       e    r    c    q    q    f    k    h    l    l    p    i    f    l    s    e    l    f    q    s

601   ccattggctt ctagtatact cttttttcccct tcactccaaa aagacttagg taatttgact
       p    l    a    s    s    i    l    f    f    p    s    l    q    k    d    l    g    n    l    t

661   ccctggggaa gatttgtccg ccaaagagag aaaattgata aattgcttta tgcagaaatt
       p    w    g    r    f    v    r    q    r    e    k    i    d    k    l    l    y    a    e    i

721   gctgaacgca gacaggaaat aaattctgat cgtattgata ttctttcctt actcatatca
       a    e    r    r    q    e    i    n    s    d    r    i    d    i    l    s    l    l    i    s

781   gctagagatg agacaggtga ttccatgtca gataaagaat tacgagatga attgattacc
       a    r    d    e    t    g    d    s    m    s    d    k    e    l    r    d    e    l    i    t

841   ttaatgattt ctgçacatga aactacaggg acagctatgg catggtctct atattg
       l    m    i    s    g    h    e    t    t    g    t    a    m    a    w    s    l    y
```

Fig 3.20 Translated amino acid sequence of P450tag from pCDFDuet-P450tag showing N-terminal 6-histidine sequence (H) in-frame with the protein sequence. M shows the start of translation and m is start of P450 protein

```
                                                      rbs              M
  1   ttcccctcta gaaataattt tgtttaactt taagaaggag atataccatg ggcagcagcc
        P  L   -  K   -  F   C  L  T   L  R  R    R  Y  T  M   G  S  S
                                              m
 61   atcaccatca tcaccacagc caggatccta tggaaaaaat tactttccca aatttatatt
      ┌─────────────┐
      │ H  H  H   H  H  H │ S  Q  D  P   M  E  K   I  T  F  P   N  L  Y
      └─────────────┘
121   gcccatttcc agaaaggaaa aatcagtatt ttgaagttct acaagactat gcgcttcaat
        C  P  F   P  E  R  K   N  Q  Y   F  E  V   L  Q  D  Y   A  L  Q
181   gggtacttcg cttcaagcta attgatagtg aatcactata ccagcgtttc tcaaaagcaa
        W  V  L   R  F  K  L   I  D  S   E  S  L   Y  Q  R  F   S  K  A
241   aattttattt actcacagca ggtgcttatc ctcattgtca actacaagaa ttaaaaattg
        K  F  Y   L  L  T  A   G  A  Y   P  H  C   Q  L  E  E   L  K  I
301   ctaatgatgt aatcagctgg ttattcattt gggacgacca atgtgacatt tcagacttag
        A  N  D   V  I  S  W   L  F  I   W  D  D   Q  C  D  I   S  D  L
361   ggaaaaaacc tgaactactg aaaatctggt gtaacagatt cctagaaata ctaaatggag
        G  K  K   P  E  L  L   K  I  W   C  N  R   F  L  E  I   L  N  G
421   cagaacttac tgccgatgat ctgccccttg gatttgcatt aagagatatt agaaaccgca
        A  E  L   T  A  D  D   L  P  L   G  F  A   L  R  D  I   R  N  R
481   taattaacag aggaagcata acattcttcc atcattttgt acgtaacttt gaggattatt
        I  I  N   R  G  S  I   T  F  F   H  H  F   V  R  N  F   E  D  Y
541   tttacggatg tattgaagaa gctcataacc gtgtcactgt atcaattcct gatgttgaag
        F  Y  G   C  I  E  E   A  H  N   R  V  T   V  S  I  P   D  V  E
601   cttatatcaa aatccgtagt gcaaacgcag ctgccgctct gtgtctcaat ttaattgaat
        A  Y  I   K  I  R  S   A  N  A   A  A  A   L  C  L  N   L  I  E
661   tctgtgacag agtaatgatt ccttattctt taagaaatca tgatactctc aacaaattaa
        F  C  D   R  V  M  I   P  Y  S   L  R  N   H  D  T  L   N  K  L
721   ctcaaatgac gattaatatt cttgcctggt
        T  Q  M   T  I  N  I   L  A  W
```

Fig 3.21 Translated amino acid sequence NStag from pCDFDuet-NStag showing N-terminal 6-histidine sequence (H) in-frame with the protein sequence. M shows the start of translation and m is start of NS protein with the ribosomal binding site (rbs) nucleotide sequence.

3.4 Discussion

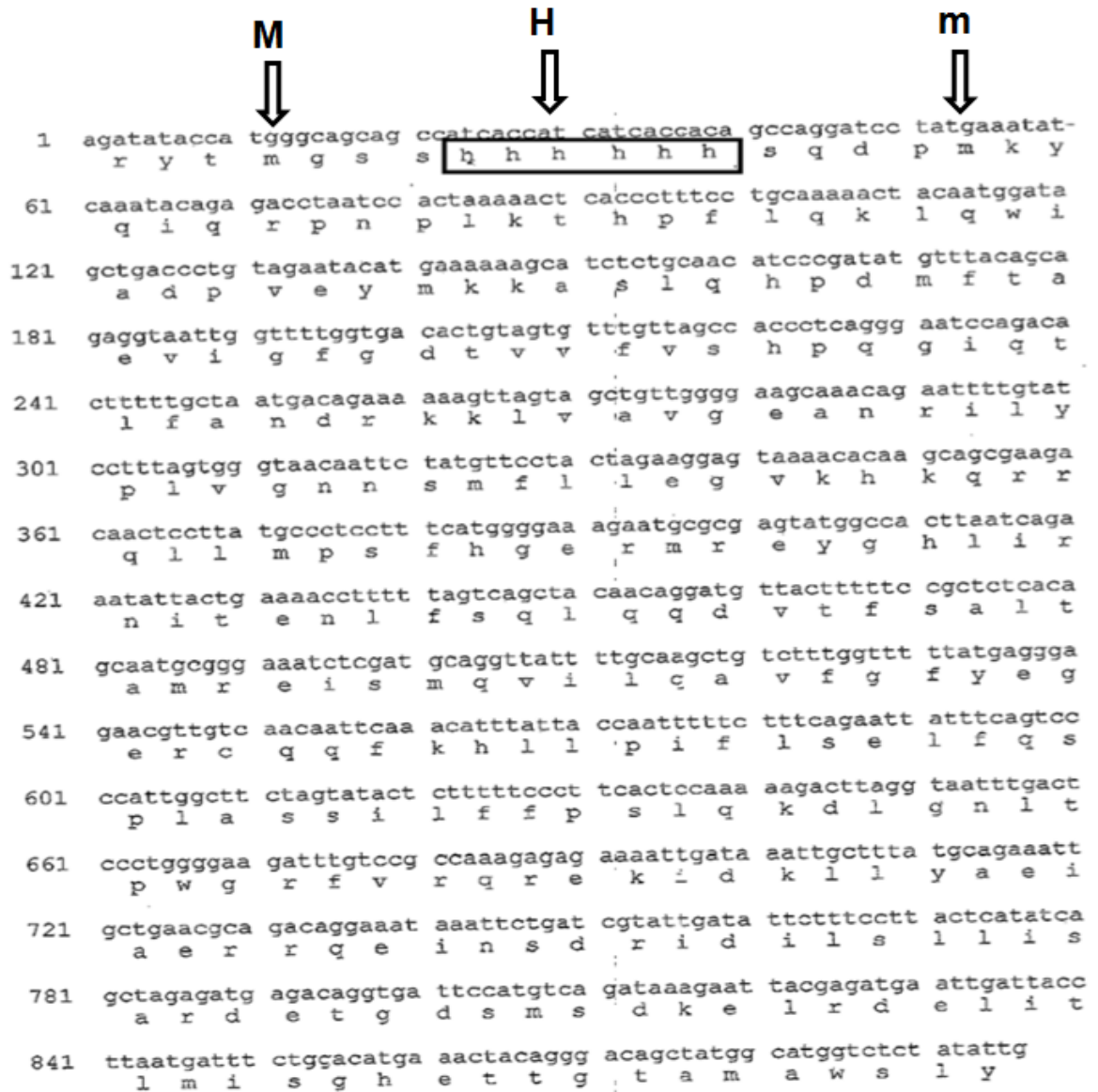PCR amplification of terpene synthase and cytochrome P450 genes from *Nostoc sp*. PCC 7120 genomic DNA was initially done without success. As a result, several optimization techniques and trouble-shooting protocols were instituted. These included graded $MgCl_2$ concentration titrations, temperature gradient PCR, variations in the primer concentration, as well as using additives such as BSA and DMSO. Other causes of PCR failure were considered including issues with primers (primer dimer formation, miss priming/mismatch), PCR program and DNA template issues. Since an in-silico template amplification of target using these primers all yielded products, it was thought that the most likely cause of failure would be issues with the template such as inhibitors. A further clean up and subsequent PCR confirmed that the amplification failure was due to inhibitors in the template. This is also seen in the significant increase of the $A_{260/A230}$ value. A further proof was demonstrated by the failure of amplification in a mix of both the initial DNA extract and the kit purified DNA as template.

For faithful amplification of target and to reduce the amount of time, energy and resources for successful cloning, a few strategies were included in the workflow. These included the substituting the Taq DNA polymerase with a high fidelity phusion polymerase, dephosphorylating the digested plasmid to reduce re-ligation and the use of ligation PCR screening. In particular, dephosphorylations of the cut plasmid greatly reduced the background of self-re-ligated plasmid and thereby increase the number of positive clones per transformation. This was further enhanced by doing a PCR screen of the ligation mix before transformation. The level of background in each case was proportional to the intensity of the 200 bp band which measured how much of the uncut plus the re-ligated plasmid was in the ligation mix. Since the ligation mix was done in ratios of 6:1, 3:1, 1:1 and 1:3, it enabled a choice of which ratio gave the least background as well as to avoid those that had multiple inserts. The 6:1 ligation ratio evidently produced less background of non-recombinants and more positive clones but also had evidence of multiple inserts. The 3:1 ligation ratio consistently produced the best results (Dardel, 1988; Topcu, 2000).

High fidelity DNA polymerase with lower error rate was substituted for Taq DNA polymerase because sequence results showed single nucleotide mutations and deletions. Of five positive clones sequenced, all had mutations or deletion. In comparison, all the positive

clones from the Phusion polymerase amplification did not have any form of mutation or deletion except for the 2.4 kb long NS-P450 amplicon. For this, of two positive clones sequenced, one had a single nucleotide synonymous substitution while the other had no form of error in the sequence result.

In conclusion, the target genes were amplified, cloned into the vectors, transformed and sequence verified. In-silico translation of the verified sequences showed that all the genes were all in-frame. These mutant plasmids would be investigated in the next chapter for the expression of the requisite terpene as well as the metabolic effect of the terpene on *E. coli*.

CHAPTER 4

Heterologous expression of sesquiterpene synthase and assay of terpene product

4.1 Introduction

The universal building blocks of terpenes, present in all organisms have been harnessed for several metabolic engineering projects using more efficient secondary hosts with varying degrees of success, as discussed in section 2.5.3 In this regard, *E. coli* and yeast have served as the main hosts for metabolic engineering of terpenes. In *E. coli,* the presence of the non-mevalonate pathway or 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate pathway (MEP/DXP) synthesizes isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) using glyceraldehyde-3-phosphate and pyruvate as source materials (Rohmer et al., 1996). These precursors for terpene biosynthesis which are inter-convertible isomers are then condensed by the farnesyl pyrophosphate synthase (*ispA*) to generate the farnesyl pyrophosphate FPP in both mevalonate and DXP pathways (Wilkin et al., 1990). This organism could therefore provide a pool of precursors which could be tapped for heterologous terpene production by inserting a down-stream terpene gene to take advantage of these precursors but yield to date has not been significant. However, the precursors are essential for the survival of the organism because they are used in the biosynthesis of several important cell wall components such as bactoprenol, glycosyl carrier lipids, dolichols and respiratory chain quinones as well as prenylation of tRNA and post translational modification (Apfel et al., 1999; Carter et al., 2003; Connolly and Winkler, 1989).

Engineering the DXP pathway to over-produce the precursors would potentially circumvent any ill effect, but as noted by Martin et al. (2003a), this approach suffers from limitations due to possible negative regulatory mechanisms in *E. coli.* Engineering the mevalonate pathway with or without the DXP pathway for the purpose of providing a flux of precursors in *E. coli* for terpene production has shown significant success and several important terpenes such as bisabolene, 900 mg/L; amorphadiene, 500 mg/L; taxadiene, 1000 mg/L; lycopene, 8 mg/g dry cell weight (Ajikumar et al., 2010; Alper et al., 2005; Peralta-Yahya et al., 2011a; Redding-Johanson et al., 2011b) have been synthesised. Some of these terpene synthetic pathways require additional down-stream processing enzymes such as cytochrome P450. Though cytochrome P450 catalyses critical steps in the pathway

of many important terpenes such as the anticancer taxol and the antimalarial artemisinin, expression of this non-native gene in *E. coli* has proved very difficult for diverse reasons such as the absence of endogenous electron tranfer protein partners, codon usage as well as stability of the expressed protein amongst others(Schuler and Werck-Reichhart, 2003). Functional assays of the sesquiterpene gene (alr4685) (NS1) and the adjacent cytochrome P450 (alr4686) (P450) products from *Nostoc* sp.7120 have been carried out by Agger et al. (2008). They functionally expressed both the terpene synthase and the adjacent cytochrome P450 with a qualitative assay of the terpene and terpenoid but could not elucidate the structure of the terpenoid because of low yield nor did they quantify the terpene. Harada et al. (2009) used a downstream mevalonate utilising plasmid to increase the titres of the target terpenoid and thus succeeded in determining this to be 1,2,3,5,6,7,8,8aoctahydro-6-isopropenyl-4,8a-dimethylnaphth-1-ol through NMR analysis, but no quantitative assay was done for either terpene or terpenoid. Peralta-Yahya et al. (2011a) developed an *E. coli* codon-optimized FPP over-producing plasmid using elements of mevalonate pathway from yeast and with the downstream *idi* (IPP isomerase)and *ispA* from *E. coli* and used this to increase the titres of the bisabolene to 900 mg/L This is a veritable tool for cross platform terpene and terpenoid production and engineering.

In moving forward, a quantitative assay method is therefore desirable which would serve as a template for further refinement. In so doing, a baseline productivity level of this system in *E. coli* is needed on which to measure subsequent perturbations. This chapter aims to express the terpene synthetic genes cloned in the previous chapter and quantitatively assay the terpene product. Subsequently, the quantitative effects of co-expression of FPP over-producing pathway and the cytochrome P450 will also be examined.

4.2 Materials and methods

4.2.1 Plasmids and strains

Three new strains were made by transforming *E. coli* BL21 (DE3) with FPP pathway, pETNS-P450, and ferrodoxin/ferrodoxin-reductase to produce the strains listed in Table 4.1. To balance the plasmid metabolic load, each strain had three plasmids with or without insert. The pBbA5c plasmid backbone was derived from pBbA5c-RFP by digesting out the

RFP gene with BamHI and BglII, purifying and re-ligating the compatible sticky ends as described in 3.2.8.

Table 4.1 Plasmids and strains

Plasmid/strain	Description	source
pET-NS1-P450NS	*NS1 and P450* inserted into *NcoI-PstI* and *NdeI-EcoRV* sites respectively of pETDuet-1	(H. Harada et al., 2011)
pETNS1	*NS1* inserted into *NotI* and *NcoI* od pETDuet-1	This study
pETNS1tag	*NS1* inserted into *NotI* and *BamHI* site of pETDuet-1	This study
pBbA5c-RFP	RFP cloned into pBb plasmid using standard BglBrick cloning strategy.	(Lee et al., 2011)
pBbA5c	RFP excised from pBbA5c-RFP and re-ligated	This study
pBbA5c-MevT(CO)-T1-MBIS(CO, ispA)	Mevalonate pathway genes from *S. cerevisiae, idi* and *ispA* from *E. coli* cloned into the pBb plasmid using standard BglBrick cloning strategy	(Peralta-Yahya et al., 2011b)
pBBR1-MCS2	Cloning vector, Kan^R	(Kovach et al., 1995)
pBBR-FER-RED	Ferredoxin I and Ferredoxin-NADP-reductase from *Nostoc* sp.7120 cloned into the HindIII and BamHI site of pBBR	(Agger et al., 2008)
N	pETNS1 + pBbA5c + pBBR1-MCS2 transformed into *E. coli* BL21 (DE3)	This study
N+	pETNS1 + pBbA5c-MevT(CO)-T1-MBIS(CO, ispA) + pBBR1-MCS2 transformed into *E. coli* BL21 (DE3)	This study
P	pETNS1-P450 + pBBR-FER-RED + pBbA5ctransformed into *E. coli* BL21 (DE3)	This study
T	pETNS1-P450 + pBbA5c-MevT(CO)-T1-MBIS(CO, ispA) + pBBR-FER-RED transformed into *E. coli* BL21 (DE3)	This study
C1	pETDuet-1 + pBBR1MCS2 + pBbA5c transformed into *E. coli* BL21 (DE3)	This study
BLK	Plasmid free *E. coli* BL21 (DE3)	Novagen

4.2.2 Expression of sesquiterpene synthase

The expression of terpene synthase gene (pETNS1tag) in *E. coli* was carried out in Terrific broth (TB) (12 g of bacto-tryptone, 24 g of yeast extract, and 4 ml of glycerol and 100 ml of phosphate buffer in 1 L of DH_2O). Overnight cultures of 5 ml of cells containing plasmid with and without insert containing ampicillin at 100 µg/ml were set up in 50 ml tubes using LB broth and grown at 37^oC. The OD_{600} was measured after overnight growth and a volume of culture enough to give a starting OD_{600} of 0.01was used to inoculate 500 ml of TB containing the same concentration of ampicillin. Culture flasks were then incubated in a rotary shaker at 30^oC and growth was monitored by OD_{600} measurements. Cultures were induced with IPTG at a final concentration of 0.4 mM at an OD_{600} of 0.6 and growth continued at 25^oC. 1 ml of culture was taken at intervals starting at induction to monitor recombinant protein accumulation with time. These were spun at 8500 x g for 10 min at 4^oC and pellet stored at -20^oC. Growth was continued for 40 hours post induction and a pellet was obtained as above. Subsequent cultures were harvested at 30 hours post induction.

4.2.3 Total protein extraction

Four different methods of protein extraction were tested in order to optimise the protein extraction process. These were bead beating, French press, nitrogen cracking and chemical lysis. Each method was done in triplicate using 500 ml cultures.

4.2.3.1 Bead beating

The crude protein lysate from the harvested cell pellets were obtained as follows: 5 ml of 0.5M TEAB and 500 mg each of glass beads 106 µm and 425-600 µm diameter (Sigma-Aldrich, UK) were added to cell pellet. This was then vortexed for 30 seconds and chilled on ice for 1 min and repeated through 5 cycles and then sonicated in a water bath containing ice for another 5 minutes. 100 µl of protease inhibitor cocktail (Sigma-Aldrich, UK) and 0.1% Benzonase (Novagen) was added and incubated on ice for 15 min. To remove cell debris, cells were centrifuged at 13,000 rpm for 30 min at 4^oC. Supernatant

containing the total protein extract (crude cell lysate) was collected into sterile microcentrifuge tubes by pipetting and stored at -20$^{\text{o}}$C until used.

4.2.3.2 French press

For French press, the cell pellet was resuspended in 5 ml of 0.5M TEAB as previously described and subjected to French pressure using standard operating procedure and collected in 50 ml tubes. 100 µl of protease inhibitor cocktail (Sigma-Aldrich, UK) and 0.1% Benzonase (Novagen) was added and incubated on ice for 15 min and clarified by centrifuging at 13,000 rpm at 4$^{\text{o}}$C for 30 min and aspirating the supernatant into sterile 1.5 ml microcentrifuge tubes.

4.2.3.3 Mechanical cracking with liquid nitrogen

For mechanical cracking with liquid nitrogen, cells were thawed on ice, protease inhibitor cocktail added and subjected to mechanical grinding and liquid nitrogen according to a standard procedure and collected in sterile 1.5 ml microcentrifuge tubes. Protease inhibitor cocktail (Sigma) and 0.1% Benzonase (Novagen) was added and incubated on ice for 15 min and clarified by centrifuging at 13, 000 rpm for 30 min at 4$^{\text{o}}$C. Supernatant containing total protein was aspirated into fresh sterile 1.5 tubes (Eppendorf) and stored at -20$^{\text{o}}$C until used.

4.2.3.4 Chemical lysis

Chemical lysis was done using BugBuster$^{\text{TM}}$ Protein extraction reagent (primary amine free) (Novagen) according to the manufacturer's protocol using Benzonase (Novagen) and protease inhibitor cocktail (Sigma) and centrifuging at 13,000 rpm at 4$^{\text{o}}$C for 30 min and supernatant was aspirated into sterile 1.5 ml microcentrifuge tubes.

Total protein quantification was done by Bradford assay kit (Sigma, UK) and absorbance at 595 nm was measured using a multi well plate reader (GENiOS-Tecan, UK).

4.2.4 SDS-PAGE of total lysate and cell pellet

A 12% bis/acrylamide gel was prepared and used to separate the proteins in both the crude protein extract and the 1 ml cell pellets from section 4.2.2.

The loading buffer (62.5 mM Tris HCl pH 6.8, 2% SDS, 25% glycerol, 0.01% bromophenol blue) was mixed with 10% β-mercaptoethanol and added into the sample in a 2: 1 ratio, heated for 5 minutes at 95°C, and loaded onto the gel. Pre-stained protein ProtoMarkers™ (Fischer Scientific) were used alongside the samples. The gel was electrophoresed at 80 volts until the dye front entered the resolving and 200 volts until the dye front ran through the resolving gel using the Protean II multicell (Bio-Rad). The gel was stained using instant blue stain solution (Expedeon, UK) in an orbital shaker for an hour. The gel was washed with de-ionized water and further immersed in de-ionized water until de-staining was achieved.

4.2.5 In-gel digestion of over-expressed band

SDS PAGE gel bands containing the over-expressed protein was cut out and subjected to in-gel protein trypsin digestion as described by Gan et al (2005). Briefly, the gel was completely de-stained using ammonium bicarbonate, reduced and alkylated using dithiothreitol and iodoacetamide respectively. Trypsin (1:40 w/w) was added and digested overnight and peptides eluted from gel using ammonium bicarbonate, formic acid and acetonitrile washing cycles. Peptide solution was vacuum dried using vacuum concentrator (Concentrator 5301, Eppendorf, UK). The dried peptides were stored at -20°C until further analysis.

4.2.6 Protein purification

Protein lysate from *E. coli* strain containing pETDuet-NS1tag from chemical lysis was purified using an affinity column.

Affinity chromatography was performedusing His.Bind purification kit (Novagen) to purify terpene synthase protein from protein lysate according to the manufacturer's instructions.

Eluted fractions from column affinity chromatography were concentrated using an Amicon Ultra-15 (Millipore) centrifugal filters according to the manufacturer's instructions. Purified protein sample were subjected to SDS-PAGE and in-gel tryptic digestion as outlined in sections 4.2.4 and 4.2.5.

4.2.7 LC-MS identification of proteins

The dried peptide fractions were resuspended in 15 µl of buffer A (3% acetonitrile and 0.1% formic acid) and analyzed on a HCT Ultra Electrospray ionization (ESI) - Ion trap (Bruker Daltonics, UK) coupled to an online nanoLC capillary liquid chromatography system (Famos, Switchos and Ultimate from Dionex/LC Packings, Amsterdam, The Netherlands). The peptides were desalted on a µ-Precolumn™cartridge of 5 µm particle size of 300 µm id x 5 mm od before separation on a PepMap C-18 RP capillary column (Dionex/LC Parkings) on a constant flow rate of 0.3 µl/min. A 45 minutes gradient starting with 3% buffer B (97% acetonitrile and 0.1% formic acid) and ramped to 35% in 45 minutes was used. Data acquisition was set in the positive ion mode with a mass range of 300 – 2000 m/z. Peptides with +2, +3 and +4 charge states were selected for further fragmentation. The protein identification search using MASCOT 2.0 (Matrix Science) was done on *E. coli* BL21 (DE3) and *Nostoc* sp. 7120 database downloaded from Uniprot (www.uniprot.org) on 14/02/2012. The search parameters were as follows: MS tolerance of 1.2 Da, MS/MS tolerance at 0.6, oxidation of methionine as fixed and cysteine modification with iodoacetamide as variable, with no missed cleavage of trypsin.

4.2.8 Growth curve measurement

Growth measurements were done to test the effect of the organic solvent, dodecane on *E. coli* pETNS1 strain and to assess growth characteristics of other mutants.

4.2.8.1 Effect of dodecane on growth

To test the effect of dodecane on the growth of *E. coli* containing pETNS1,starter culture was used to inoculate fresh TB supplemented with 100 µg/ml ampicillin at an initial concentration of 0.01 OD_{600} (optical density at 600 nm). Two hundred microliters of the

freshly inoculated media were added to each well in a 96-well plate (Fisher Scientific, UK) and half of the samples were layered with 20 µl of dodecane. Uninoculated TB broth was used as a control for contamination. The 96-well plate was incubated with aeration (orbital shaking) at 30°C in a Genios multiplate reader (Tecan, USA) and the absorbance was recorded every 30 minutes for 72 hours. Three replicates of each sample were used.

4.2.8.2 Growth measurement of mutants

Overnight cultures of each *E. coli* strain was used to inoculated fresh 5ml TB with appropriate antibiotics supplementation and grown to an OD_{600} of between 0.6 and 0.8 and induced with 0.4 mM IPTG after which 200 µl of the culture was added to 96-well plate and incubated as described above.

4.2.9 SPME-GCMS analysis of sesquiterpene

100 ml of culture of *E. coli* cells(BLK, pETNS1, N, N+, P, T and C1) were grown and induced as described in triplicates and assayed for terpene production. Cultures were tightly sealed in layers of aluminium foil and head space sampled 30 hours post induction for 30 min by solid-phase microextraction (SPME) using a 100 µM polydimethylsiloxane fiber (Supelco). After absorption, analysis was done on Finnigan Trace GC Ultra gas chromatograph coupled to a Finnigan Trace DSQ quadrupole mass spectrometer (Thermo-Fisher) operating in a split mode. Separation was done on a TR-5MS capillary column (length, 30 m; inner diameter, 0.25 mm; film thickness, 0.25 µm) (Thermo-Fisher). Polydimethylsiloxane fibre was desorbed for 5 min using an injection port of $250^{\circ}C$ and oven temperature was set at $100^{\circ}C$ for 7 minutes and ramped at $20^{\circ}C$/min to $300^{\circ}C$ and held for 2 min. Mass spectra was scanned in the range of 40 to 250 atomic mass units (amu) at 1 second interval. Spectra was searched against NIST 11 database (Scientific Instrument Services, Inc NJ, USA).

4.2.10 GCMS calibration curve for β-elemene

Calibration curve using the sesquiterpene β-elemene (International laboratory, USA) was generated for the purpose of quantifying terpene produced from biphasic cultures. β-elemene standards were made in dodecane solvent at concentrations of 5 µg/ml, 10 µg/ml, 15 µg/ml, 20 µg/ml 30 µg/ml, 40 µg/ml, 50 µg/ml and 100 µg/ml in triplicates. Samples were run on a Finnigan TraceGC Ultra gas chromatograph coupled to a Finnigan Trace DSQ quadrupole mass spectrometer (Thermo-Fisher) using a TR-5MS capillary column (length, 30 m; inner diameter, 0.25 mm; film thickness, 0.25 µm) (Thermo). The injection port was set at 250^oC with an injection volume of 0.5 µl and oven program started at 80^oC and held for min then ramped at 10^oC/min to 120^oC. It was subsequently ramped at 3^oC/min to 160^oC and then to 270^oC at 5^oC/min and held for 5 min. Ionization source was set at 230^oC with a scan rate of 5 scans per second with a full scan range of 40-250 m/z. Three injections of each sample were done. A calibration curve was generated by plotting the average peak area of each sample against its concentration.

4.2.11 Biphasic culture for sesquiterpene

Quantification of sesquiterpene produced from *E coli* mutants was done by growing the cells in a biphasic culture with an organic solvent. Three different organic solvents were tested.

4.2.11.1 Testing different organic solvents

100 ml TB were inoculated with overnight cultures of pETNS1 and grown to an OD_{600} of 0.6 at 30^oC and induced with 0.4 mM IPTG. Cultures were layered with 10% (10 ml) of octane, decane or dodecane, each in triplicate. Growth was continued at 25^oC and after 28 hours of growth, the incubation temperature was increased to 30^oC for a further 2 hours before extracting the organic solvent layer. Aqueous culture was discarded and brine was added to the organic solvent layer to further separate any residual media. Complete drying

and removal of traces of aqueous media was done by drying over anhydrous magnesium sulphate and filtered through filter paper.

4.2.11.2 Dodecane biphasic culture

For quantitative analysis of sesquiterpene production, triplicate 100 ml cultures of *E. coli* strains (BLK, N, N+, P and T) were set up as previously described. After induction with IPTG, 10 ml (10%) dodecane was layered on the culture and growth continued at 25°C and processed as described in 4.2.11.1.

4.2.11.3 GCMS quantitative analysis of sesquiterpene

Quantization of sesquiterpene from biphasic culture was done by running samples from dodecane extract on the GCMS using the same program as in section 4.2.10. Each phenotype had three biological replicates and three technical replicates from each of the biological replicates. The average peak area was used to generate the concentration using the β-elemene standard curve.

4.3 Results

4.3.1 Protein extraction and quantification

Soluble protein extract from bead beating, liquid nitrogen cracking, French pressure and chemical lysis (using BugBuster^{TM}) were performedand quantified using Bradford Ultra assay reagent. The quantification results were compared as seen in Figure 4.1. The extraction method showed that the French press gave the highest protein concentration per ml of culture and therefore the most efficient extraction method with glass bead beating giving the lowest yield. However, the French press was less reproducible compared to chemical lysis and required larger volume of cultures. Liquid nitrogen cracking was also very efficient even though it was time consuming and required some physical energy. In terms of reproducibility of results and ease of procedure, the chemical lysis method was superior as evident by the significantly smaller error bars. The lysis buffer of the BugBuster^{TM} was also compatible with the affinity chromatography kit and was therefore the method of choice for terpene synthase purification.
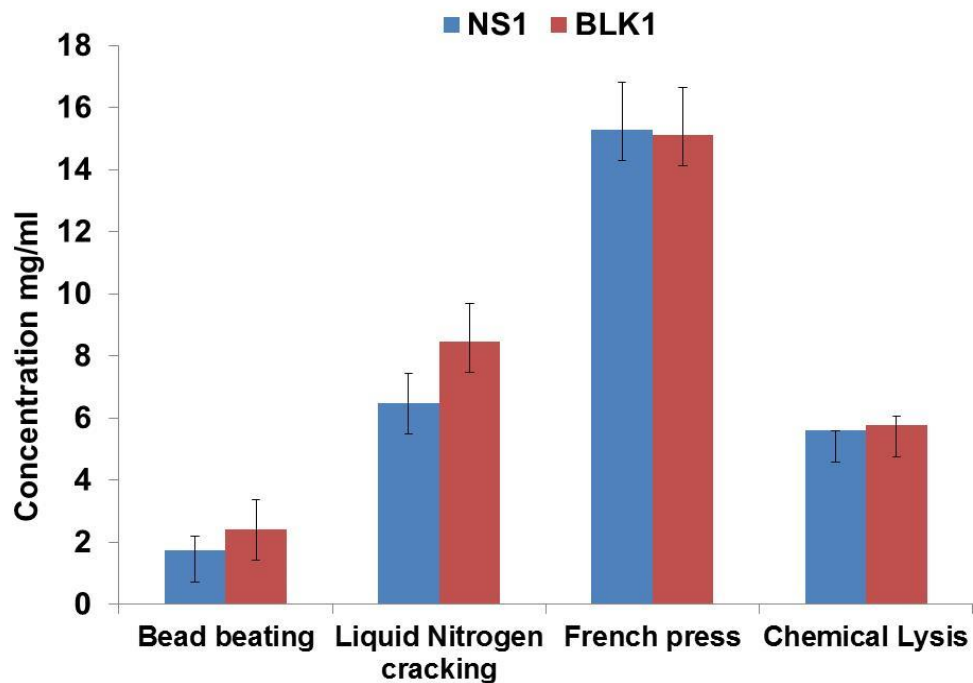
Figure 4.1 Comparison of different methods of protein extraction using plasmid free *E. coli* BL21 (DE3) (BLK1) and *E. coli* + (plasmid((NS1). Protein quantitation was done from 500 ml of culture for each method.

4.3.2 Over-expression of sesquiterpene synthase

The total protein lysate of the his-tagged terpene synthase cloned into pETDuet(pETNS1tag) and transformed into *E. coli*(NS1) and plasmid free*E. coli* cells (BLK) were run on an SDS-PAGE gel to check for over-expression of the terpene synthase protein. This was done to confirm that the protein had actually been expressed. As seen in Figure 4.2, there was an over-expressed protein band corresponding to the expected molecular mass of the protein at 37 kDa, while the wild type control did not show the same band. This showed evidence of successful protein expression.
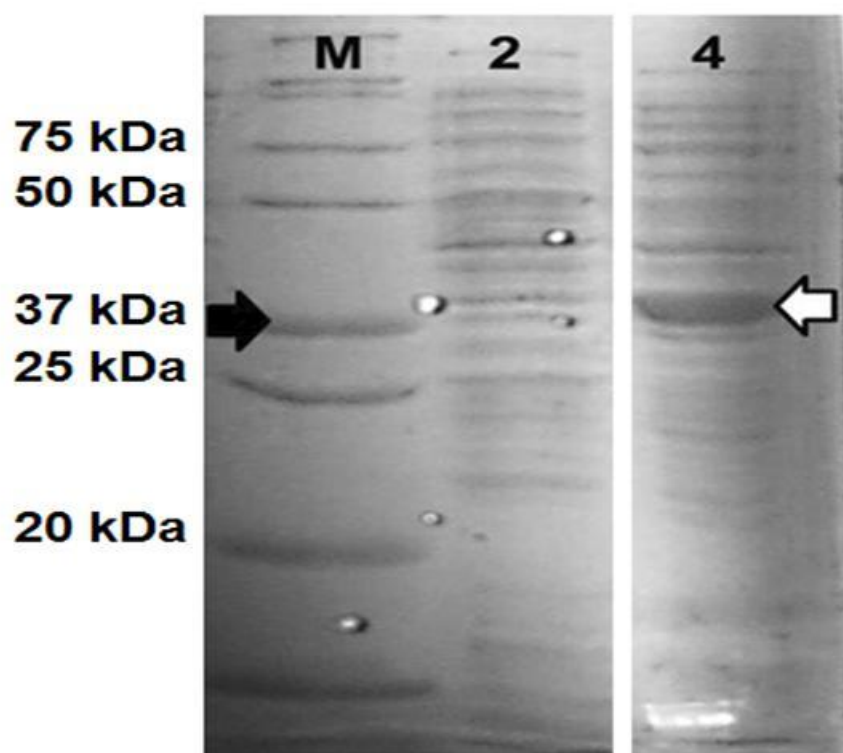
Fig 4.2 12% SDS-PAGE of 10 μg of total protein lysate plasmid free*E. coli* **(BLK1) in lane 2 and** *E. coli* **with pETNS1tag(NS1) in lane 4. The transformed** *E. coli* **lysate show over–expressed band corresponding to about 37 kDa band (hollow arrow) compared to the marker (M**)

4.3.3 Purification of expressed protein

The crude protein lysate from pETNS1tag *E. coli* was purified by nickel affinity chromatography. The terminal histidine tag on a protein can bind to the iminodiacetic acid (IDA) bound nickel and this histidine-nickel bond can be broken to elute the histidine tagged protein by increasing concentrations of imidazole. SDS-PAGE of the eluted protein fraction in Figure 4.3 shows a clear band that correspond to the theoretical size of (37kDa) of the putative protein in the mutant cells while the wild type cells did not show any band.
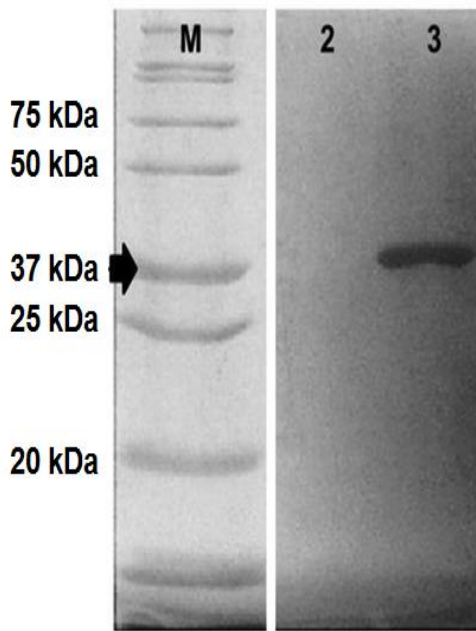
Fig 4.3 12% SDS PAGE gel of purified putative sesquiterpene synthase protein using affinity chromatography. Lane 2 is from an*E. coli* **BL21 (DE3)containing an empty pETDUET plasmid as control and lane 3 is the** *E. coli* **transformed with pETNStag showing the band of interest and M is marker.**

4.3.4 Protein expression as a function of time

Target protein (NS1) expression was surveyed over time to get the best time to harvest culture in order to achieve optimal target protein yield. A plasmid free strain without the target protein (BLK1) was used as a control. As seen in Figure 4.4, and verified by protein densitometry in Figure 4.5, target protein expression was maximal at 30 hours post induction and shows reduction from about 40 hours post induction.
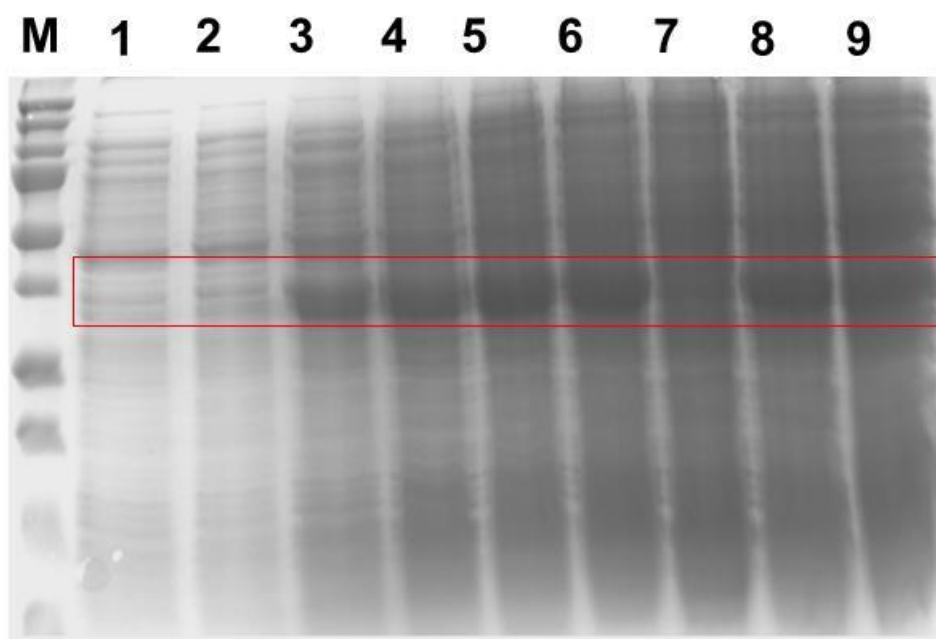
Fig 4.4 12% SDS-PAGE gel of protein expression time line using protein lysate from cells harbouring putative sesquiterpene synthase gene as well as control. Maker (M), lane 1(NS1) at induction, 2 negative control at induction, 3, 4, 5 and 6 are for expression at 10, 15, 20 and 25 hours post-induction. Lane 7 shows negative control at 25 hours post-induction and lanes 8 and 9 are for 30 and 40 hours post induction. Each lane had equal volume of culture.
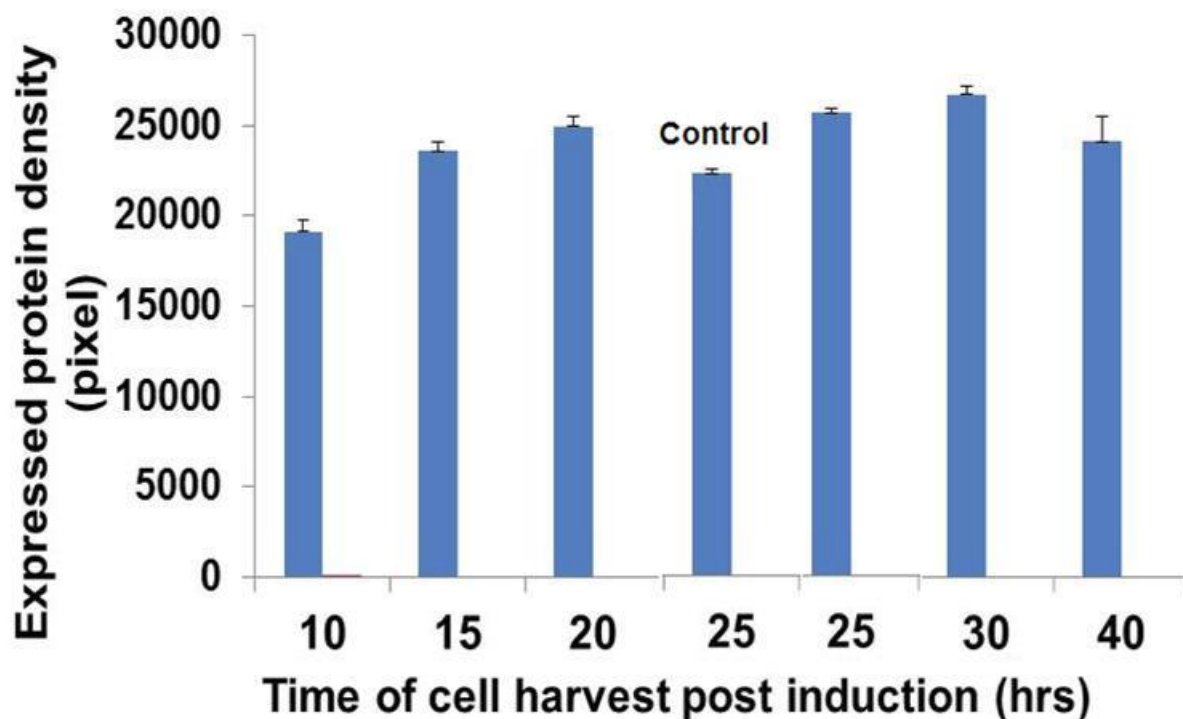
Fig 4.5 putative sesquiterpene expression from point of induction with IPTG measured by average density (pixels) of over-expressed protein band in 1 ml culture volumes during the course of time with a negative control taken at 25 hours post induction.

4.3.5 LC-MS identification of expressed protein

The purified protein from the total protein lysate from the mutant *E. coli* expressing the putative terpene synthase was identified with LC-MS after tryptic digestion. A database search on *Nostoc* sp. 7120 gave a single protein hit as shown in Table 4.2. This protein (Alr4685) is a terpene synthase.

Table 4.2 LC_MS purified protein (NS1tag) identification

Protein accession	Protein description	Protein score	Protein cover	# of unique peptides
Q8YN85_NOSS1	Alr4685 protein OS=Nostoc sp. (strain PCC 7120 / UTEX 2576) GN=alr4685	150	22%	6

4.3.6 SPME- GCMS *in vivo* test of terpene synthase activity

After the identification of the purified protein as a terpene synthase, it was necessary to test its activity *in vivo* as there is always the possibility that the protein may not be functional. There are several reasons for inactivity of expressed proteins ranging from mis-folding of protein to forming insoluble aggregates. A solid phase micro extraction (SPME) procedure combined with GC-MS was done to detect the volatile effluent from *E. coli* cultures. As seen in Figure 4.6, a single chromatographic peak was detected and the fragmentation was matched to β-elemene with a 95% confidence in the NIST database.
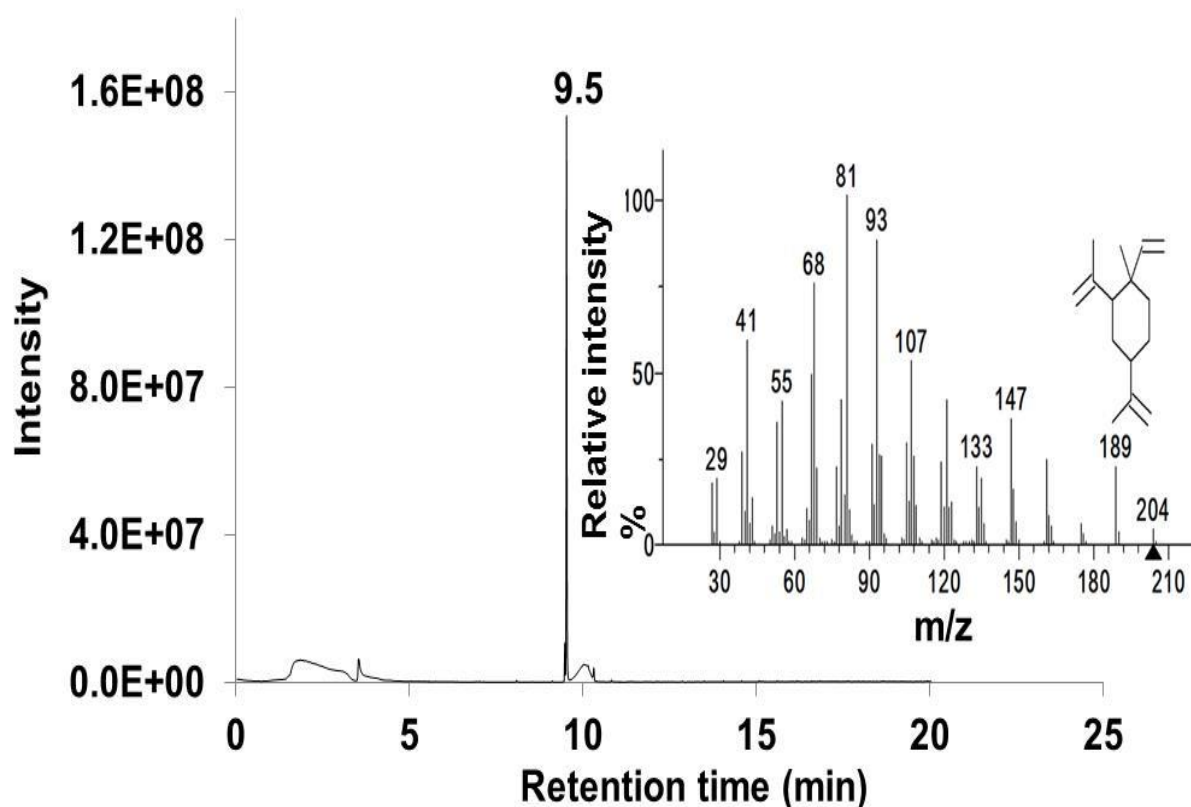
Fig 4.6GC-MS analysis of volatile organic compounds produced by mutant *E. coli* expressing sesquiterpene synthases from *Nostoc* sp. A single chromatographic peak with a retention time of 9.5 was detected. The NIST database search identified this compound as the antiproliferative β-elemene at 95% match.

4.3.7 Effect of dodecane on the growth of *E. coli*

A biphasic culture using non-miscible dodecane as a solvent to trap the volatile terpene in order to do a quantitative assay of the produced terpene.The effect of dodecane on the growth of *E. coli* cells was tested to check if dodecane had any significant impact on the growth of the cells. The productivity of the cell is linked to it growth and any factor that limits its growth is likely to affect its productivity. *E. coli* cells containing the terpene synthase were grown with and without dodecane overlay. As seen in Figure 4.7, the growth curve of both cells grown with and without dodecane did not show any significant difference over a 72 hour period. This confirms that dodecane did not have any significant effect on the growth of the cells and hence the productivity.
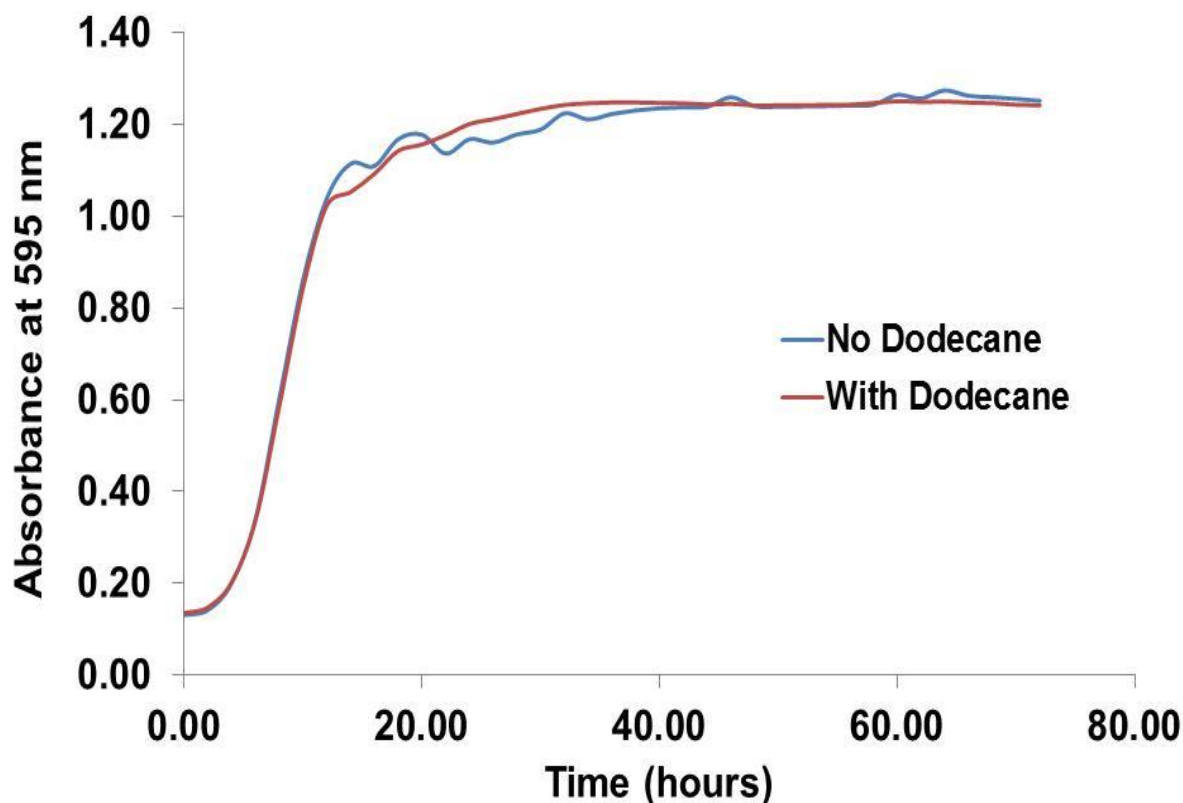
Fig 4.7 Growth assay of *E. coli* containing terpene synthase to check the effect of the organic solvent dodecane on growth over a 72 hour period at 30 ^{o}C.

4.3.8 Effect of plasmid and terpene synthase on the growth of *E. coli*

The effect of a pETDuet plasmid and the terpene synthase insert on the growth of *E. coli* cells was also tested using the the plasmid free *E. coli* BL21 cells as control. The result as seen in Figure 4.8 show a consistent growth retardation in the both the plasmid carrying and the terpene synthase carrying strains indicating that both imposed a metabolic burden on the cells.
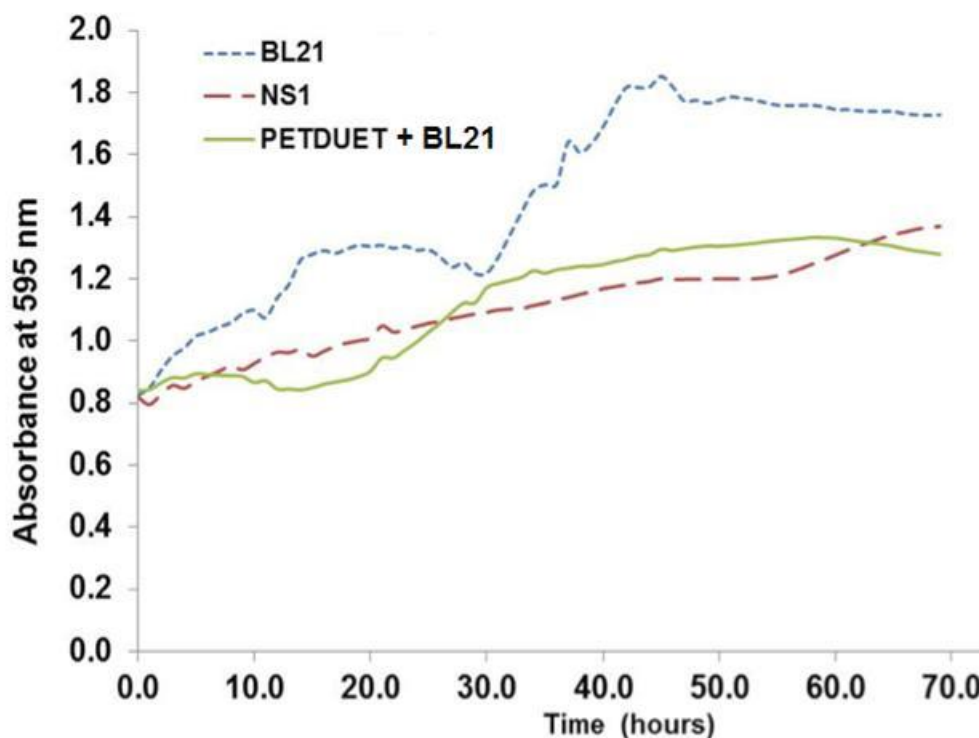
Fig 4.8 Post induction growth curve of *E. coli* cell containing an empty pETDuet plasmid, a pETDuet with a terpene synthase insert (NS1) and the plasmid free *E. coli* cells (BL21).

4.3.9 Calibration curve for terpene quantitation

To quantitatively assay terpene produced from biphasic cultures using GC-MS, an external calibration method using the sesquiterpene β-elemene was done as described in 4.2.10. The β-elemene standard was made in dodecane in increasing concentrations this was used to generate the calibration curve seen in Figure 4.9. The lower limit of detection for β-elemene in dodecane solvent was determined experimentally to be 3.5 μg/ml.
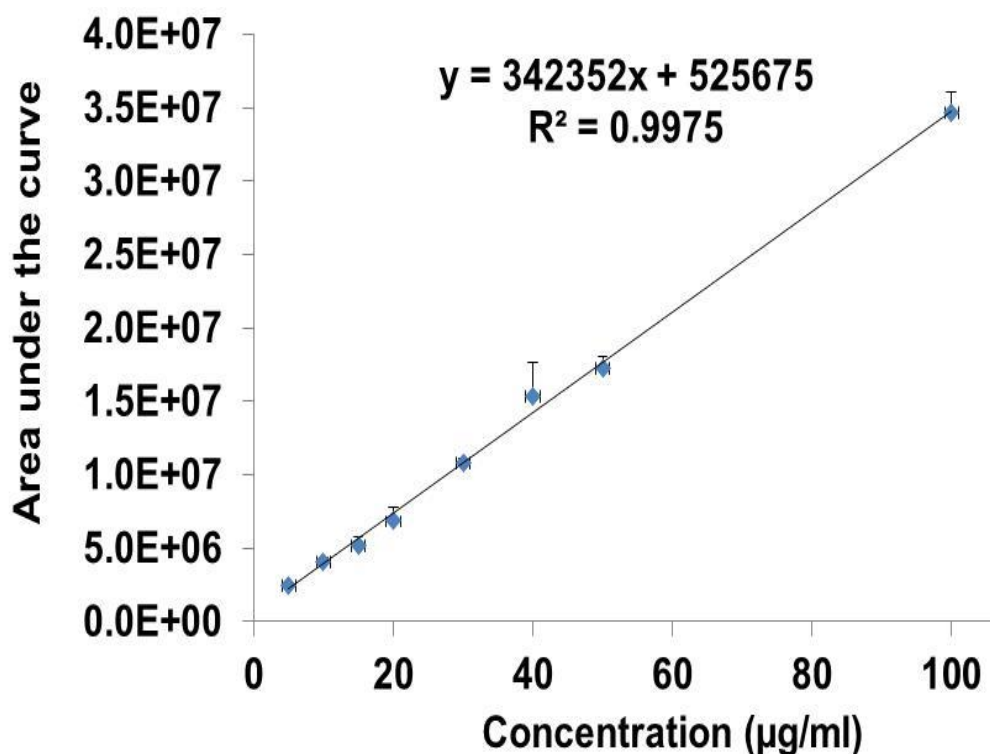
Fig 4.9 A GC-MS external calibration curve forβ-elemene standard within the dynamic range of 3.5-100 µg/ml The area under the curve for each peak intensity was plotted against the concentration of the standard producing that peak.

4.3.10 Testing different organic solvents for *in vivo* terpene production

Using SPME-GCMS analysis, the terpene synthase was shown to be active by qualitative assay. For the quantitative assay of the terpene, a biphasic culture using selected organic solvents was done to ascertain the toxicity and hence the productivity of the cells in each of these solvents. The organic layer of each culture was sampled and the GCMS analysis results for octane, decane and dodecane is shown in Figure 4.10. The results show that dodecane is the best solvent between these three for biphasic culture with a baseline terpene production at 38 µg/ml. These results show a positive correlation with the the logP values for the solvents (octane (4.9), decane (6.0) and dodecane (7.0) as previously reported (Inoue and Horikoshi, 1991; Sardessai and Bhosle, 2002; T. et al., 2002). The logP value of an organic solvent is a measure of a solvent's hydrophobicity and toxicity and the higher the logP value the less toxic it is to the cells.
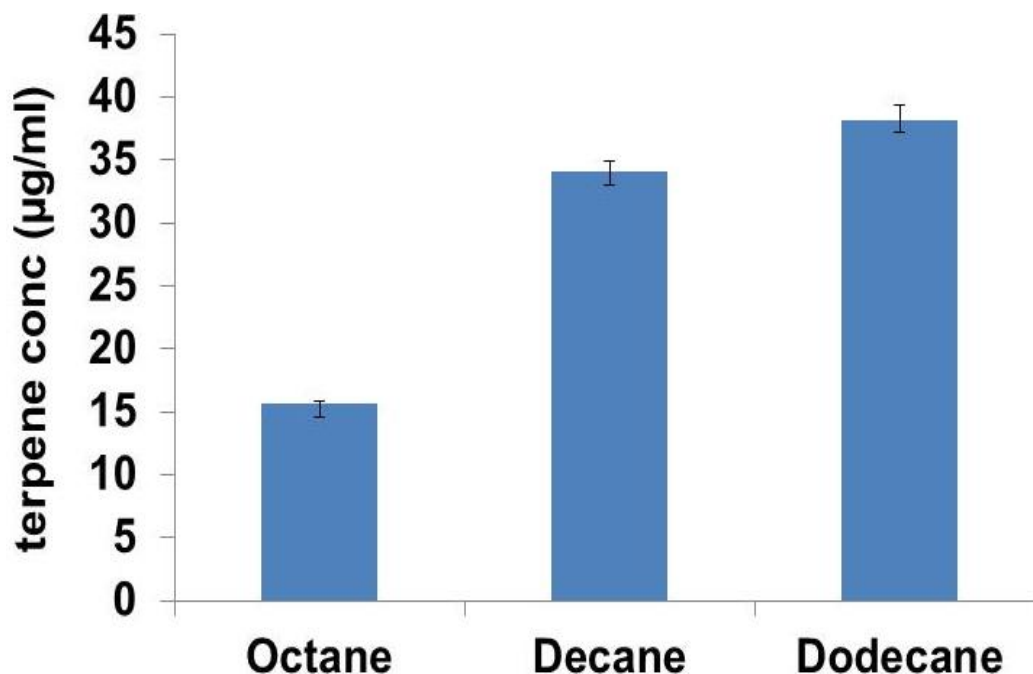
Fig 4.10 GCMS analysis of terpene produced in biphasic cultures using different organic solvents at 30 hours post induction showing the effect of the different solvents on *E. coli*terpene production.

4.3.11 Quantitative assay of terpene produced by different phenotypes

The productivity of the different strains in biphasic flask cultures with dodecane as the preferred solvent was assessed. Optimisation of the GC gradient was carried out in order to remove dodecane interference to terpene quantification (Figure 4.11). Figure 4.12 shows the terpene concentration per millilitre of dodecane solvent at the time of harvest in each strain. The N+ (terpene synthase + mevalonate pathway containing pasmid) showed a significant increase in terpene yield (more than 34 fold) compared to the mevalonate negative strain (N). The P (terpene synthase + cytochrome P450 + ferrodoxin reductase) and T (terpene synthase + cytochrome P450 +ferrodoxin reductase + mevalonate pathway containing plasmid) strains did not show any improvement over the N strain. Despite the fact that all the cultures were harvested at the same time, there were still differences in the biomass produced by each strain as a result of differences in growth and metabolism. The specific productivity of each strain with respect to terpene production was thus evaluated

and is as shown in Figure 4.13. The specific productivity of N+ was 40 fold more than that of N. There was no significant difference in productivity between N and T while P was 3 fold less productive than
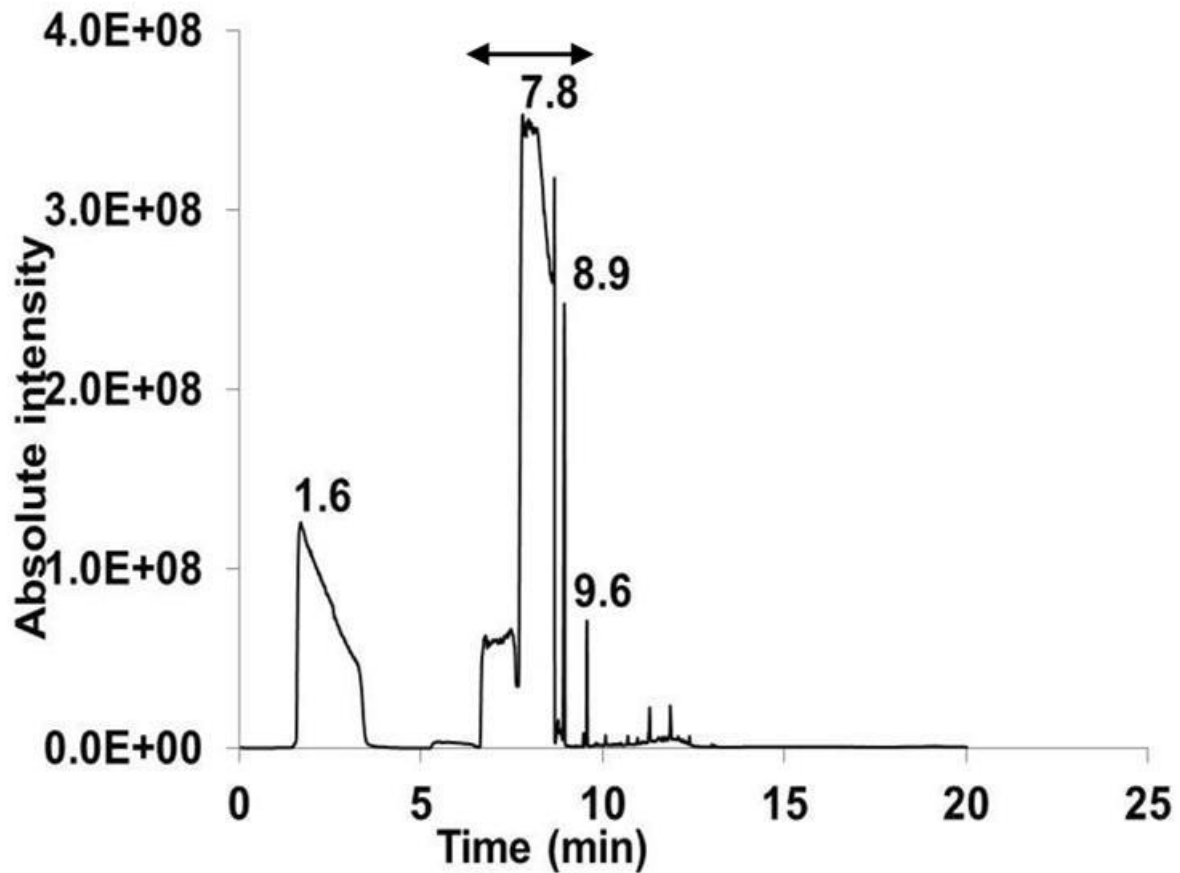
Figure 4.11 GCMS spectrum of an optimised gradient used to separate the ubiquitous dodecane solvent from the terpene. The dodecane shows a wide elution time from 6.8 to 8.9 minutes as shown by the double arrows and the terpene elution time was at 9.6 minutes.
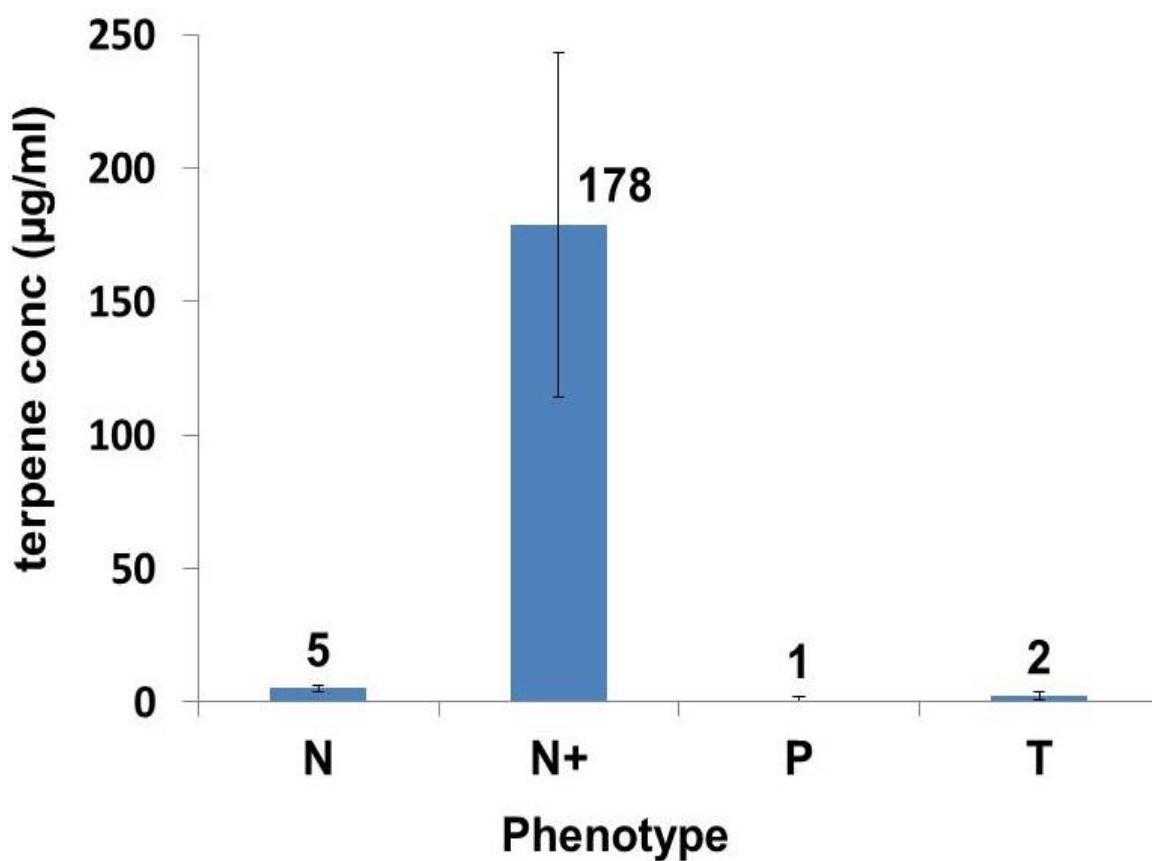
Fig 4.12 Terpene produced by the *E. coli* strains per ml of dodecane solvent. The E. coli strains all contain three plasmids each. N (pETNS1 + pBbA5c + pBBR1-MCS2),N+ (pETNS1 + pBbA5c-MevT(CO)-T1-MBIS(CO, ispA) +pBBR1-MCS-2), P (pETNS1-P450 + pBbA5c + pBBR-FER-RED), T (pETNS1-P450 + pBbA5c-MevT(CO)-T1-MBIS(CO, ispA + pBBR-FER-RED).
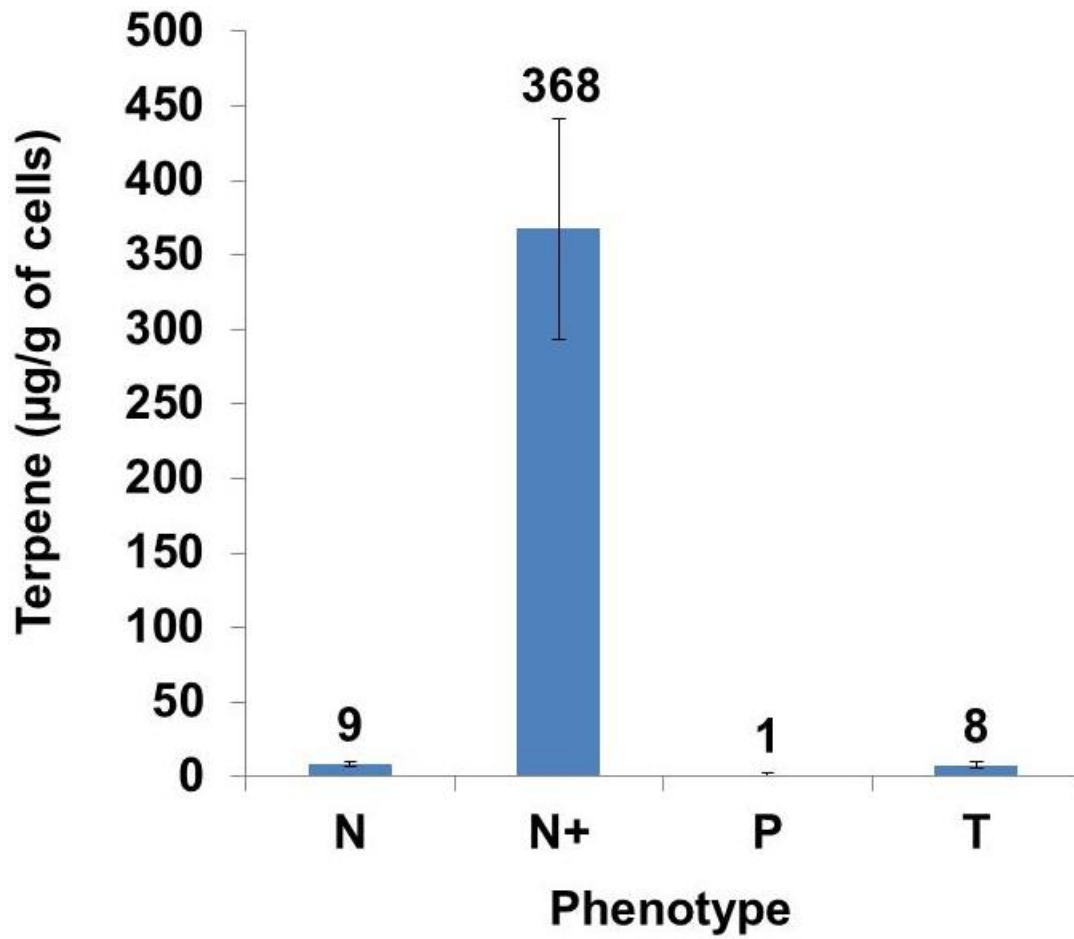
Fig 4.13 Specific productivity of *E. coli* strains per gram of cells. N (pETNS1 + pBbA5c + pBBR1- MCS2), N+ (pETNS1 + pBbA5c-MevT(CO)-T1-MBIS(CO, ispA) +pBBR1-MCS-2), P (pETNS1-P450 + pBbA5c + pBBR-FER-RED), T (pETNS1-P450 + pBbA5c-MevT(CO)-T1-MBIS(CO, ispA + pBBR-FER-RED).

4.4 Discussion

The success of any metabolic engineering project would inherently require a functional expression system upon which further improvement can be done. The functional assay of the terpene synthase and a quantitative assay of the terpene product has been carried out in this chapter. A number ofstrategies were employed in this chapter for the overall goal of achieving a quantitative assay of the terpene yeild. These include sampling time, assay method and choice of trap solvent for the terpene.

In order to quantify and thus determine the productivity of the terpene synthase, the *in vivo* turn-over of the enzyme had to be assessed through the process of expression and identification of the protein. A suitable time frame for protein harvest needed to be established by measuring the protein accumulation as a function of time because recombinant protein degradation by proteolytic enzymes is a common problem that could occur during cultivation (Hannig and Makrides, 1998; Murby et al., 1996) and this could be subject to the length of cultivation. The BL21 expression host which is known to be deficient in both ompT and lon proteases (Sørensen and Mortensen, 2005) was combined with a culture temperature set at 25°C to enhance recombinant protein yield. However, the protein time line expression measurements still showed recombinant protein reduction beyond 30 hours post-induction. Based on this, both the biomass harvest and terpene sampling was done at 30 hours after IPTG induction.

The measurable final product titre of an expression system is not only a function of pathway efficiency but also on the efficiency of the assay methods. Thus, in a system producing volatile hydrocarbons like terpenes, measurement of product titre would have to take into consideration headspace losses during cultivation. To minimize the headspace losses, biphasic culture using a non-miscible organic solvent was adopted. Of the three organic solvents tested, dodecane gave the highest yield of product and as shown by the growth curve measurement, dodecane did not have any negative effect on growth of *E. coli*. As has been shown previously, the logP value of a solvent is a measure of its hydrophobicity and the higher the logP value the more hydrophobic and hence the less toxic the solvent.

Dodecane with a logP value of 7.0 (Inoue and Horikoshi, 1991) and lower volatility showed a proportionately more yield of terpene than octane and decane. However, dodecane had a very intense solvent peak and significantly broad elution time which suppressed the detection of the terpene using GCMS. Optimisation of the GC gradient was thus carried out in order to remove dodecane interference for terpene quantification. The dodecane biphasic culture gave a yield of 38 µg/ml as quantified by GCMS compared to 16 µg/ml for octane and 34 µg/ml for decane as the basal expression level of terpene as a single plasmid expression system using the *E. coli* native precursor pool from the DXP pathway. However, using a three plasmid system, N (pETNS1 + pBbA5c + pBBR1-MCS2), the terpene yield showed more than a 7 fold decrease. This was done to set a baseline yield for which the subsequent strains with multiple plasmids would be compared. The drastic reduction in product titre is a direct consequence of metabolic burden of multiple plasmids with growth retardation as well as decrease of biomass (Diaz Ricci and Hernández, 2000; Redding-Johanson et al., 2011a; Rozkov et al., 2004). This multiple plasmid containing terpene synthase strain produced a baseline terpene titre of 5.2 µg/ml and a specific productivity of 9 µg/g of dry cell weight. The addition of FPP overproducing plasmid to the terpene synthase strain in addition to the pBBR1MCS2 plasmid, N+ (pETNS1 + pBbA5c-MevT(CO)-T1-MBIS(CO, ispA) + pBBR1-MCS2) showed a more than 34 fold increase in terpene titres and a specific productivity of more than 40 fold when compared to N. However, when the FPP over-producing plasmid was transformed into a strain harbouring the terpene synthase and cytochrome P450 on a single pETDuet plasmid with a ferrodoxin/ferrodoxin-reductase, T (pETNS1-P450 + pBbA5c-MevT(CO)-T1-MBIS(CO, ispA) + pBBR-FER-RED)the terpene product was also significantly reduced compared to N+. The titres for this system were comparable to the baseline titres for N (the three plasmid system without the FPP overproducing plasmid). This was surprising considering that the oxidation product of the terpene was almost negligible. Since cytochrome P450s are poorly expressed in *E. coli*, and was indeed observed to be true from SDS-PAGE analysis, this could not be explained by assuming that the terpene product had been transformed by the P450. This then must be the effect of expressing either the cytochrome P450 with or without the ferrodoxin/ferrodoxin-reductase. It is known that the successful expression of

multiple, sequential pathway enzymes, even with optimal level of precursor supply, restrictions may be further imposed by cofactor supply, by rapid export of intermediates, and by the limited ability of the bacterial host to traffic hydrophobic intermediates from the cytosol to membranous hydroxylation sites, and back to the cytosolic sites of subsequent redox metabolism (Carter et al., 2003). This would expected to be worse when an enzyme with a notorious poor expression in *E. coli* such cytochrome P450 is involved.

4.5 Conclusions

A robust method for the quantitation of terpenes and terpenoids was developed and optimised in this chapter. This method was successfully applied to quantify the terpenes produced by recombinant strains of *E. coli*. A baseline value of 38 mg/L of terpenes was produced using the native precursor pool in *E. coli*. This value increased to 178 mg/L of terpenes upon the introduction of the yeast-based FPP over-producing pathway in *E.coli*. This was due to the increased availability of precursors for the production of terpenes. This value translates to about 369 mg of terpenes per gram of cells. The expression of cytochrome P450 and ferrodoxin reductase required to convert the terpenes to terpenoids resulted in an expected, significant decrease in terpene production. However a concomitant increase in terpenoid production was not detected.

C̪ʜᴀᴘᴛᴇʀ5

Metabolic adaptation of *E. coli* to cyanobacterial sesquiterpene synthase

5.1 Introduction

The development of metabolic engineering has been spurred not only by the major challenges of the 21st century such as climate change, renewable energy, food security, pollution and health but also by development of new technologies(Betenbaugh and Bentley, 2008; Lindberg et al., 2010). With the advent of the post genomic era, availability of new tools and technologies and development of system biology/synthetic biology, there has been a paradigm shift from seeing the cell as a production line to that of a production factory(Cortassa, 2002; Kholodenko and Westerhoff, 2004; Smolke, 2009). The ultimate aim of using metabolic engineering from this new perspective is to achieve strain improvement and cell engineering using transcriptomics, proteomics, metabolomics, computational systems biology, protein engineering and synthetic biology.

The success of metabolic engineering has been astounding so far with many useful compounds already produced through metabolic engineering in the two decades of the discipline. In therapeutics, several drugs or drug leads such as the antimalarial compound artemesinin and the anticancerous drug taxol which hitherto were produced exclusively by plant cultivation have either been wholly heterologously synthesized in microorganisms or a biosynthesis of their direct precursor achieved(DeJong et al., 2006b; Dietrich et al., 2009b; Ro et al., 2006b). In addition several other molecules such as isoprenes, flavinoids, dihydroxyacetone and ethanol have been produced in high titres (20.8 mg/l of naringenin, 700 mg/l dihydroxyacetone, 0.46 g/g of ethanol) through metabolic engineering(Cripps et al., 2009; Nguyen and Nevoigt, 2009; Trantas et al., 2009; Watts et al., 2004). There are many variables that determine efficiency of a heterologously expressed pathway such as consumption of intermediates by native pathways, mRNA abundance and half-life, protein abundance and activity. Proteomics and or metabolomics methodologies can help to identify pathway bottlenecks which can then be metabolically engineered to increase product titres(Pandhal et al., 2013, 2011a).

E. coli has been employed as an organism of choice for production of many recombinant proteins as well as metabolites for reasons ranging from cheap and rapid growth, to abundance of genomic, transcriptomic, proteomic and metabolic data (Papagianni, 2012).

Targeted proteomics methodologies have recently been applied to metabolically engineer a more than three-fold increase of the artemisinin amorpha-4,11-diene titre to greater than 500 mg/L in *E. coli*(Redding-Johanson et al., 2011a). Several other quantitative proteomics methodologies are presently widely employed, including the popular iTRAQ(Bantscheff et al., 2007).It is the objective of this chapter to use iTRAQ based proteomics to understand the metabolic changes that accompany heterologous sesquiterpene expression in *E. coli* as a stepping stone to engineering the terpene production pathway. The expression of the sesquiterpene synthase would take advantage of the native*E. coli* DXP precursor pathway for the production of terpene as shown in Figure 5.1
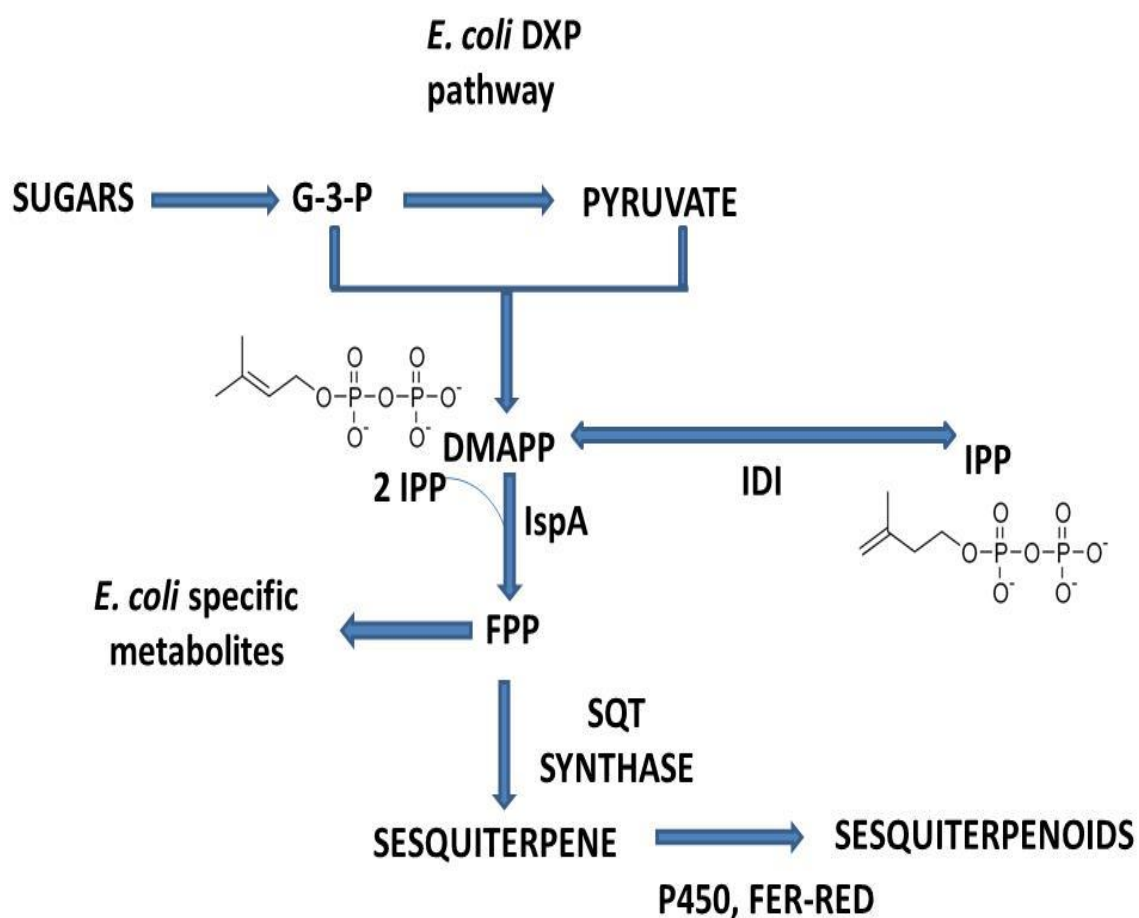
Figure 5.1 Pathway for the biosynthesis of sesquiterpenoids in *E. coli* showing the native DXP pathway for IPP/DMAPP synthesis. IspA, FPP synthase; FPP, farnesyl diphosphate; SQT synthase, sesquiterpene synthase

5.2 Materials and Methods

5.2.1 Culture conditions and cell harvest

Single colonies of mutant *E. coli* BL21 (DE3) containing pETDUET (C) and pETDUET-NS (NS) respectively was inoculated into 5 ml of LB broth containing ampicillin at 100 µg/ml concentration and grown at $37^{o}C$ in a rotary incubator for 12-16 hours. The overnight cultures were used to inoculate 100 ml of terrific broth (TB) (containing ampicillin at100 µg/ml) in 250 ml culture flasks to make a starting OD_{600} of 0.01. These were set up as biological duplicates of each phenotype. Culture flasks were then incubated in a rotary shaker at $30^{o}C$ and growth was monitored by OD_{600} measurements. Cultures were induced with IPTG at a final concentration of 0.4 mM at an OD_{600} of 0.6 and growth continued at $25^{o}C$ for further 30 hours. Cultures were harvested by centrifugation at 5400 x g for 10 minutes at $4^{o}C$. Pellet was washed twice with 500 mM triethyl ammonium bicarbonate (TEAB; pH 8.5), resuspended in 1 ml of 500 mM TEAB containing 0.1% Benzonase (Novagen) and flash frozen in liquid nitrogen and stored at $-70^{o}C$ until used.

5.2.2 Protein extraction and quantification

Thawed cells where lysed using mechanical cracking and liquid nitrogen and protein lysate was recovered by centrifugation at 21,000 x g at $4^{o}C$ for 30 minutes. The total protein lysate concentration was quantified by a Bradford assay kit (Sigma, UK) following the manufacturer's protocol.

5.2.3 SDS-PAGE

SDS-PAGE was set up as detailed in section 4.2.4 using 10 µg of each sample to check the accuracy of Bradford assay measurement and for in-gel tryptic digestion.

5.2.4 In-solution tryptic digestion

To test for the efficiency of in-solution digestion of proteins, 10 µg of each sample was digested. Prior to digestion, the protein samples were reduced with 10 mM final concentration of DTT in TEAB for 1 hour at $56^{o}C$ and alkylated with a final concentration of 55 mM iodoacetamide in TEAB for 30 minutes in the dark at room temperature. Sequencing grade trypsin (Promega, UK) was added to the protein mixture in a 1: 40 (w/w) after which acetonitrile was added to a final concentration of 5% and incubated at $37^{o}C$ for

14-16 hours. The peptide mixtures were dried in a vacuum concentrator (Concentrator 5301, Eppendorf, UK) and stored at -20°C.

5.2.5 In-gel tryptic digestion

Gel pieces from SDS-PAGE were rinsed several times with deionized water and each lane corresponding to each condition (wild type or mutant) was cut into 24 equal sized strips and processed separately. The gel strips were de-stained and trypsin digested as described in 4.2.5.

5.2.6 Stacking gel aided purification of proteins and in gel trypsin digestion

50 µg of protein from each sample was loaded onto widely separated lanes and run at 60V until the proteins were concentrated just above the resolving gel and then washed and stained with Coomassie brilliant blue. Each concentrated protein blob was then excised, diced and then digested as detailed in section 4.2.8

5.2.7 Pre-labeling SCX peptide fractionation

Dried peptides were resuspended in 70 µl of buffer A (25% acetonitrile in water, 0.1% formic acid) and separation was performed on a PolySULFOETHYL A column (PolyLC, USA) 21 cm length, 2.1 mm internal diameter and 5 µm pore size using a BioLC HPLC unit (Dionex, UK). The separation programme consisted of 100% buffer A for 5 minutes, 0-35% buffer B (25% acetonitrile in water, 0.1% formic acid, 500 mM potassium chloride) for 30 minute, 35-100% buffer B for 10 minutes, 100% buffer B for 5 minutes, and finally 100% buffer A for 10minutes. A flow rate of 0.2 ml/min was maintained with an injection volume of 70 µl. The gradient was then modified to 0-5% buffer B for 1 minute, 5-30% buffer B for 30 minutes, 30-35% buffer B in 5 minutes, 35-100% buffer B in 5 minutes, 100% buffer B for 5 minutes, and again 100% buffer A for 10minutes. A UV detector UVD170Uset at 214nm and Chromeleon software v6.50 (Dionex/LC packings, The Netherlands) was used to monitor the chromatograms and fractions were collected every minute using a Foxy Jr. fraction collector (Dionex, UK). The fractions were dried in a vacuum centrifuge and stored at -20°C prior to MS analysis.

5.2.8 LC-MS identification of proteins

The dried peptide fractions were resuspended in 15 µl of buffer A (3% acetonitrile and 0.1% formic acid) and analyzed on a HCT Ultra Electrospray ionization (ESI) - Ion trap (Bruker Daltonics, UK) coupled to an online nanoLC capillary liquid chromatography system (Famos, Switchos and Ultimate from Dionex/LC Packings, Amsterdam, The Netherlands). The peptides were desalted on a µ-Precolumn™cartridge of 5 µm particle size of 300 µm id x 5 mm od before separation on a PepMap C-18 RP capillary column (Dionex/LC Parkings) on a constant flow rate of 0.3 µl/min. A 45 minutes gradient starting with 3% buffer B (97% acetonitrile and 0.1% formic acid) and ramped to 35% was used. Data acquisition was set in the positive ion mode with a mass range of $300 - 2000$ m/z. Peptides with +2, +3 and +4 charge states were selected for further fragmentation. MS/MS data was searched on an in-house Phenyx algorithm server (Genebio, Switzerland) on and *E. coli* BL21 (DE3) and *Nostoc* sp. 7120 database downloaded from Uniprot (www.uniprot.org) on 14/02/2012.The search parameters were as follows: MS tolerance of 1.2 Da, MS/MS tolerance at 0.6, oxidation of methionine as fixed and cysteine modification with iodoacetamide as variable, with no missed cleavage of trypsin.

5.2.9 iTRAQ labeling of peptides

Digested and vacuum dried peptides were resuspended in 50 µl of 500 mM TEAB buffer and labelled according to the manufacturer's (AB-Sciex, CA, USA) instruction briefly as follows: The iTRAQ 8-plex labels (113, 114,115 and 116) were resuspended in 125 µl of 100% isopropanol and mixed with the peptides (Table 5.1). The peptides/label mix were left on the bench to incubateat room temperature for two hours, pooled and dried in a vacuum centrifuge at room temperature.

Table 5.1 Phenotypic iTRAQ labeling of samples

Phenotype	Replicate	Label	Replicate	Label
Control (C)	1	113	2	114
Mutant (NS)	1	115	2	116

5.2.9 SCX fractionation of labeled peptides

iTRAQ labeled peptides were fractionated as detailed in section 5.2.7 using a gradient of 0-5% buffer B for 1 minute, 5-30% buffer B for 30 minutes, 30-35% buffer B in 5 minutes, 35-100% buffer B in 5 minutes, 100% buffer B for 5 minutes, and again 100% buffer A for 10minutes. Fractions were then dried in a vacuum centrifuge and stored at -20°C prior to MS analysis.

5.2.10 RP-LC-MS/MS analysis

Each peptide fraction was resuspended in 20 µl of buffer A (3% acetonitrile, 0.1 % formic acid) and 8 µl was injected into a nano-LC-ESI-MS/MS system. Mass spectrometric analysis was performed using a Q-STAR XL Hybrid ESI Quadrupole Time of Flight (ESI-qQ-TOF) mass spectrometer (Applied Biosystems, MDS-Sciex) coupled to an online capillary nanoLC system (Ultimate 3000, Dionex/LC packings, The Netherlands). Separation was done on a PepMap C-18 RP column (LC packings) at a constant flow rate of 0.3µl/min using a gradient of 97% buffer A and 3% buffer B (97% acetonitrile in water, 0.1% FA) for 30 minutes, followed by 3-25% buffer B for 120 minutes, 90% buffer B for 5 minutes and 100% buffer A for 5 minutes. Data acquisition was set in the positive ion mode with a selected mass range of 300 – 2000 m/z. MS/MS was performed on peptides with 2+, 3+ and 4+ charge states. A total of two injections were performed.

5.2.11 Data analysis

Protein identification and quantification was done on a Phenyx algorithm server (binary version 2.6;Genebio, Switzerland). Search was performed on an *E. coli* BL21 (DE3) protein database downloaded from Uniprot (www.uniprot.org) on 14/02/2012 using MS tolerance of 0.6 Da, MS/MS tolerance of 0.3 Da, and one missed cleavage of trypsin. Peptide modification was set as oxidation on methionine as variable, cysteine modification as fixed and 8-plex iTRAQ mass shifts (+304 Da, K and N-term).

5.2.12 Data analysis

A careful analysis of the relative abundance ratio between phenotypes was carried out using the significance testing algorithm, Signifiquant(S. Y. Ow et al., 2009). Quantification was

obtained at a 95% confidence level from proteins identified by two or more peptides. An increased abundance of protein is indicated by fold changes above a factor of 1 and a decrease in the abundance of proteins is indicated by fold changes below a factor of 1. Multiple test correction was employed on the first search to rule out the possibility of false positive results (Pham et al., 2010). The first search was used, therefore, to identify the "hotspots" in the bacterial metabolism that undergoes highly differential regulation. In the subsequent comparison, the multiple test correction was relaxed to include false negative results omitted in the first search. A further reduction of to 90 and 70% confidence level was done for pathway enrichment in specific pathways and peptides manually curated. The metabolic proteins identified at 95% confidence were mapped onto KEGG Mapper metabolic pathways curated by the Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/tool/map_pathway2.html; organism code – ebe)

5.3 Results

5.3.1 Overview of the proteomic analysis workflow and optimisations

The general proteomic workflow employed in this chapter is presented in Fig 5.2. Optimisations were carried out for protein extraction methods (detailed in 4.2.2), trypsin digestion as well as SCX-HPLC fractionation of peptides.
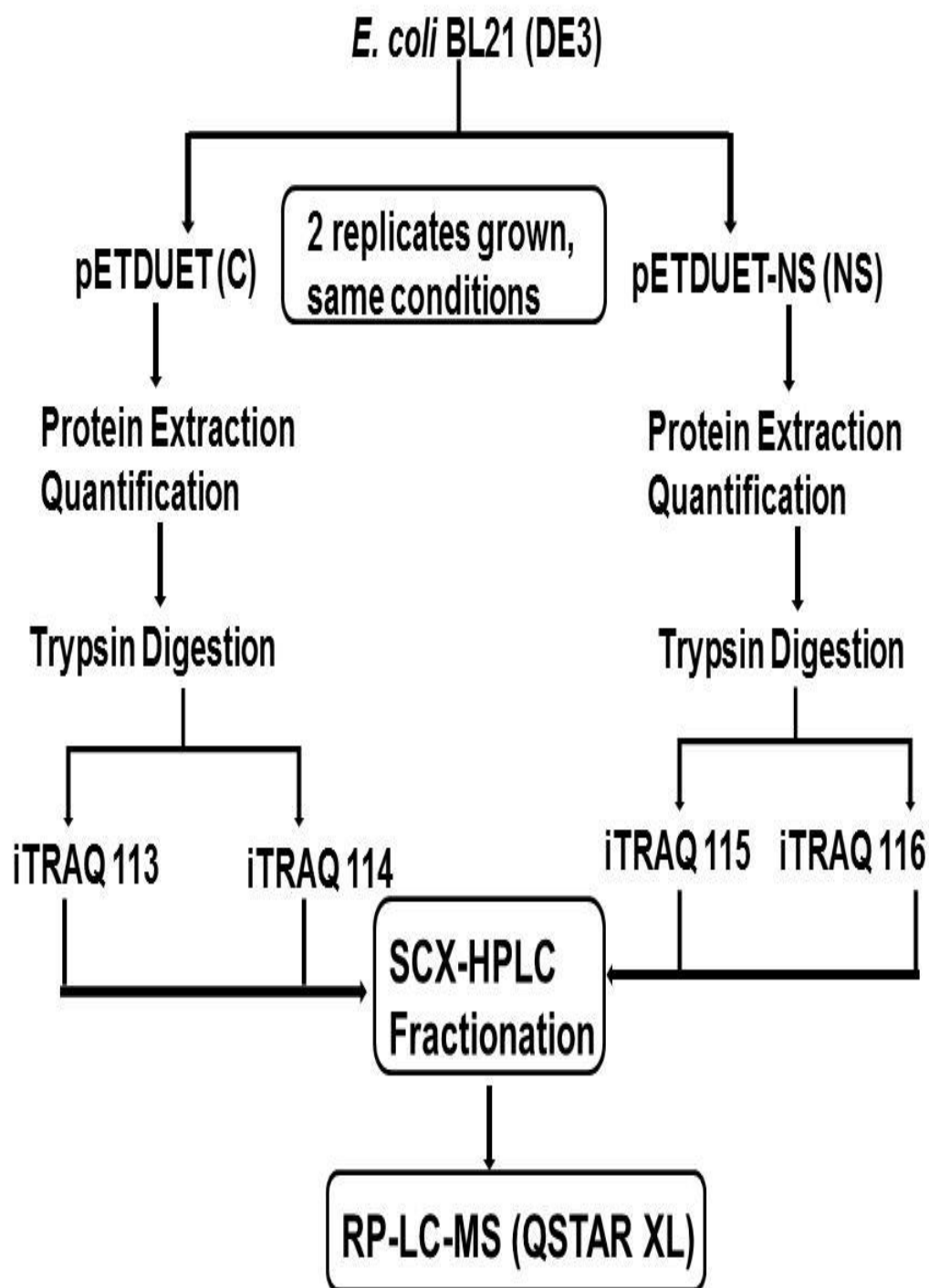
Fig 5.2 An overview of the iTRAQ proteomics workflow for the iTRAQ experiment carried out for this chapter

5.3.1.1 Optimisation of protein extraction method

Different protein extraction methods were tried as discussed in chapter 4 and mechanical cracking with liquid nitrogen was seen to be the optimal method for iTRAQ experiments, even though the French press did prove a more efficient cell lysis method, but required larger volume (> 5 ml) of cell suspension which could be achieved either by growing larger cell culture volumes (500 ml and above) or diluting the cell suspension significantly. Chemical lysis was also very efficient and much more reproducible compared to liquid nitrogen cracking, but the detergents in this buffer are not compatible with the iTRAQ workflow. For these reasons, mechanical cracking with liquid nitrogen was used as the preferred protein extraction method.

5.3.1.2 SDS PAGE analysis of extracted proteins

SDS-PAGE was used to analyze extracted and quantified proteins to further validate the quantitative methods. 10 µg of protein of sample was used for this purpose and visually inspected for equal loading across replicates as seen Fig 5.3. This also enabled direct visual comparison of phenotypes to see any differences in the protein expression pattern. As a further preliminary check on phenotypic differences in the expressed proteins, one lane of representative phenotype was digested in 24 equal sized strips. The results, as seen in Table 5.2 serves as a non-quantitaive crude screening test to show that some proteins were relatively over-expressed in between the phenotypes.

Table 5.2 Abundant proteins identified by the mass spectrometerin (NS) but not detected in control (with empty vector) *E. coli* cells

AC	Score	#pept	Protein ID
A1AAG4	114.2	21	CoA-linked acetaldehyde dehydrogenase.
A1AIF3	92.1	19	Translation elongation factor EF-Tu.
A2UCU8	73.4	16	Phosphoglycerate kinase
A7ZMP9	58.6	16	GIyceraldehyde-3-phosphate dehydrogenase (Phosphorylating)
A2UKH3	40.6	7	Tryptophanase
A2UGU7	39.7	5	Ribosomal protein L9.
A2UPI8	35.7	6	Translation elongation factor G.
A2UKF2	35.4	5	Phosphate ABC transporter, periplasmic phosphate-binding protein precursor
A7ZYQ9	30.9	5	Outer membrane protein A.
Q6J0R8	30.4	7	Chloramphenicol acetyltransferase
A2UCU9	29.1	6	Fructose-bisphosphate aldolase, class II
A1A8N6	28.1	6	Alkyl hydroperoxide reductase, C22 subunit.
A2UMT9	28.1	5	Ribosomal protein L1.
A2UKQ4	28.0	4	FAD-dependent pyridine nucleotide-disulphide oxidoreductase.
A2UL29	20.4	2	Chaperone protein DnaK.
A2UD96	19.6	4	Porin, Gram-negative type precursor
A2UKZ7	17.4	4	Purine nucleoside phosphorylase.
A2UMT8	17.3	5	Ribosomal protein L11.
A2UNW1	16.6	2	Ribosomal protein L2.
A2UPE0	16.2	3	50S ribosomal protein L5 (Fragment).
A1A9I0	15.7	2	PflB.
A1AIC1	15.5	2	Peroxidase/catalase
Q1R486	13.3	2	Uridine phosphorylase
Q68QU6	11.4	3	Malate dehydrogenase (Fragment).
Q53X36	10.9	2	Elongation factor G (Fragment).
A2ULA0	9.3	2	Phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent
A1AAT6	9.0	2	Thiol peroxidase.
A2UPE3	8.3	2	50S ribosomal protein L6.
Q52JH4	34.0	6	PstS
Q9ETZ1	27.0	4	Malate dehydrogenase
A2UL48	23.0	5	Uridine phosphorylase.
A2UKE5	16.6	2	ATP synthase F1, alpha subunit
A2UN91	15.5	2	Trigger factor.
A2UP41	49.3	12	Glyceraldehyde-3-phosphate dehydrogenase (Phosphorylating)
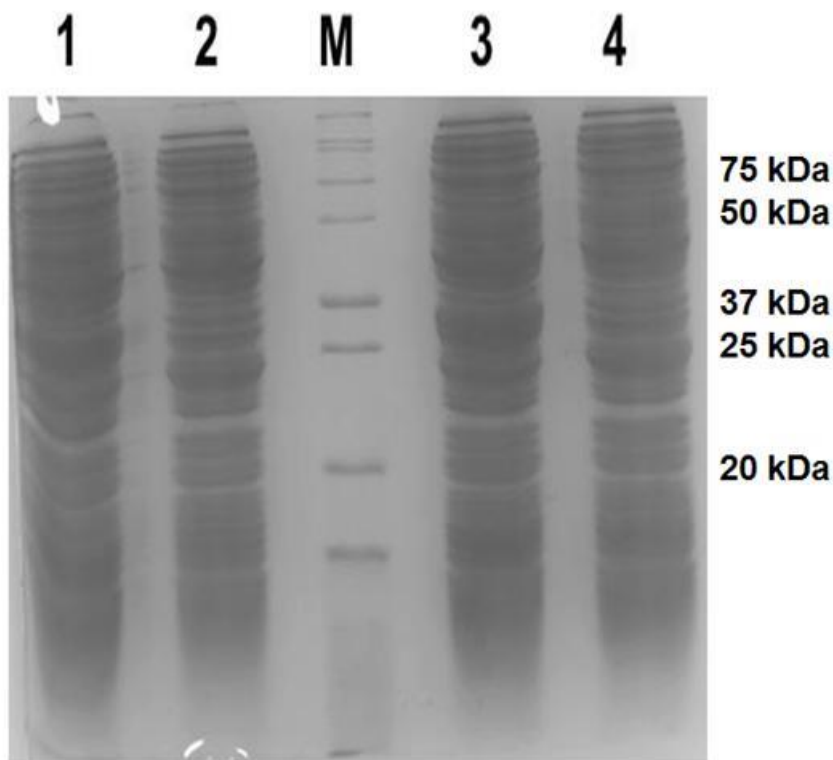
Figure 5.3 SDS-PAGE analysis of 10 µg of each protein sample used in iTRAQ experiment, C (*E. coli* transformed with empty pETDuet plasmid) in lanes 2 and 4 and NS1 (*E. coli* transformed with pETNS1) lanes 1 and 3. Marker in lane M

5.3.1.3 Testing tryptic digestion methods

Trypsin digestion of protein sample was tested before proceeding to fractionation and iTRAQ labeling in the workflow. Initial in-solution digestion of protein samples was accompanied by significant protein precipitation on reduction with DTT at 56°C. Subsequent digestion of these samples and RP-LC-MS analysis yielded few protein hits. In order to increase the number of identified proteins hits on LC-MS, the proteins were concentrated in an SDS-PAGE stacking gel and subsequently digested using an in-gel digestion method. To check for the efficiency of digestion, a volume of the digest equivalent to 5 µg of each protein sample was separated on an SDS-PAGE gel. As seen in Fig 5.4, the SDS-PAGE gel of a representative

stacking gel and post digestion gel confirm protein concentration and purification as well as completion of digestion.
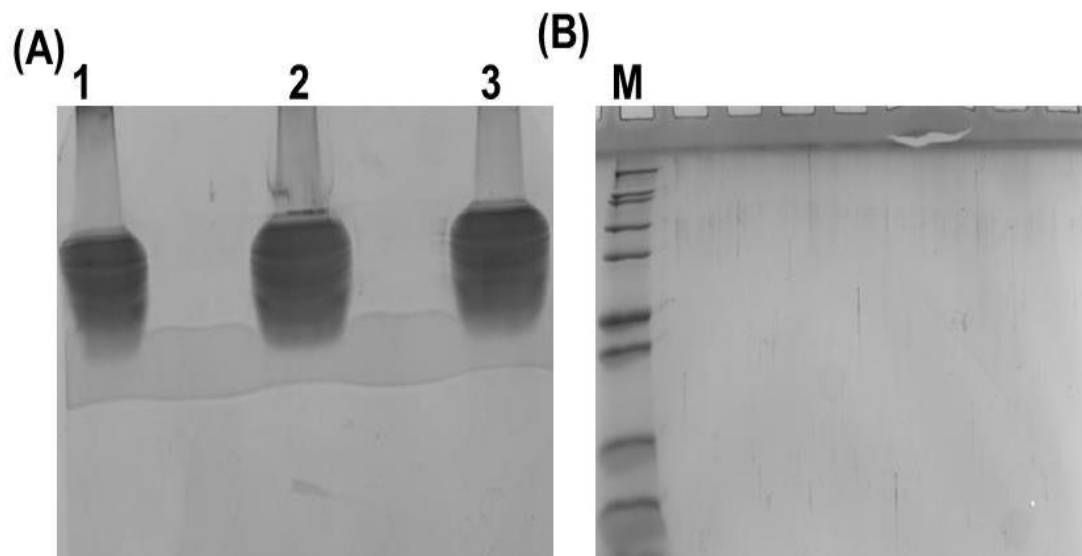
Fig 5.4 SDS-PAGE of A) stacking gel showing50 ug of concentrated proteins just above the resolving gel in each lane and B) post digestion of protein. Notice that in B, there is no visible protein band in all lanes.

5.3.1.4 Optimisation of fractionation methods

To achieve optimal identification and quantification of peptides in RP-LC-MS analysis, a two-dimensional chromatographic separation was applied. As a result, the number and complexity of peptides presented to the mass spectrometer for fragmentation and identification at any given time was reduced with significant improvements in protein numbers and especially for low abundance proteins. An efficient offline first dimensional HPLC separation of peptides significantly enhances the number of peptides identified. The efficiency of HPLC separation not only depends on the method employed, but significantly on the choice of mobile and stationary phases but also on the gradient employed. For any given system with fixed mobile and stationary phases, the choice of gradient could make a significant difference to

the outcome of the iTRAQ experiment. In Fig 5.5, the initial SCX chromatogram is seen alongside the fractional distribution of peptides. A significant elution of peptides is seen between 22 min and 26 min on the gradient. This observation served as a template for subsequent optimisation. The resolution of the SCX separation, defined by the frequency of the unique occurrence of peptides in a SCX fraction, was calculated by plotting the number of peptides seen against the frequency of peptide occurrence. As seen in Fig 5.6, the change in peptide resolution efficiency was significantly increased from 60% to 79% as a result of the gradient optimisation.
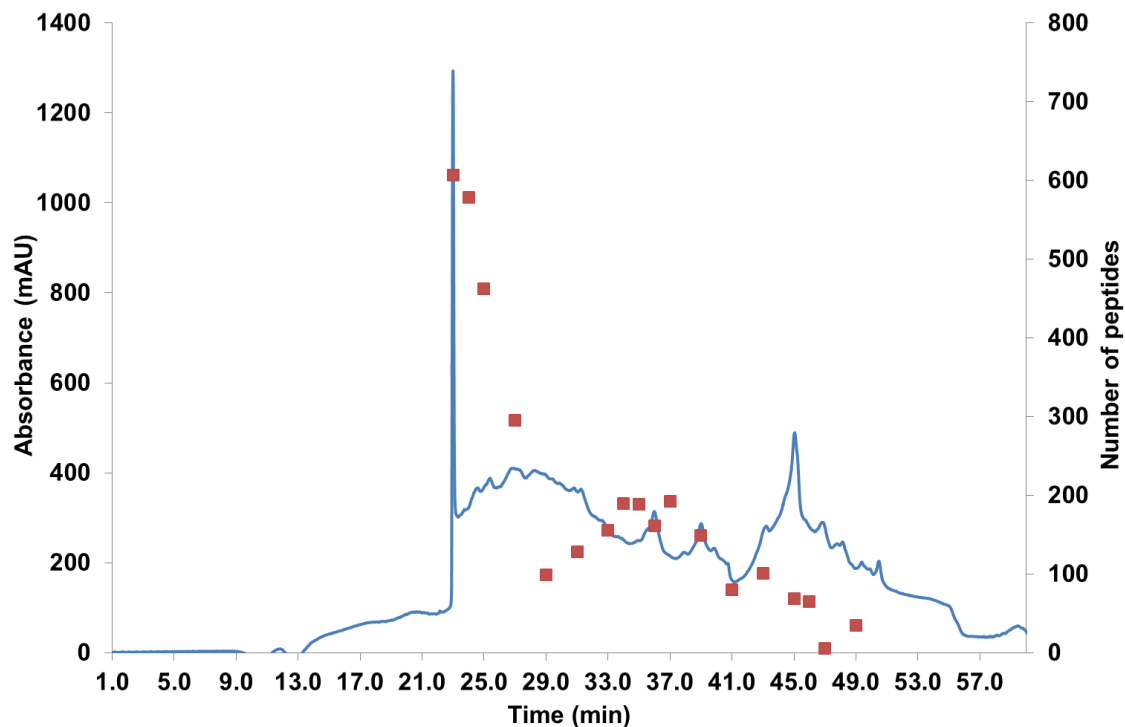
Fig 5.5 Distribution of peptides across fractions collected during SCX fractionation. The number of peptides per fraction is shown in red squares and the chromatographic trace in absorbance (mAU) units is in blue. A significant number of the total peptides were seen to be eluted between the short window of 22 to 26 min which indicates a poor peptide resolution.
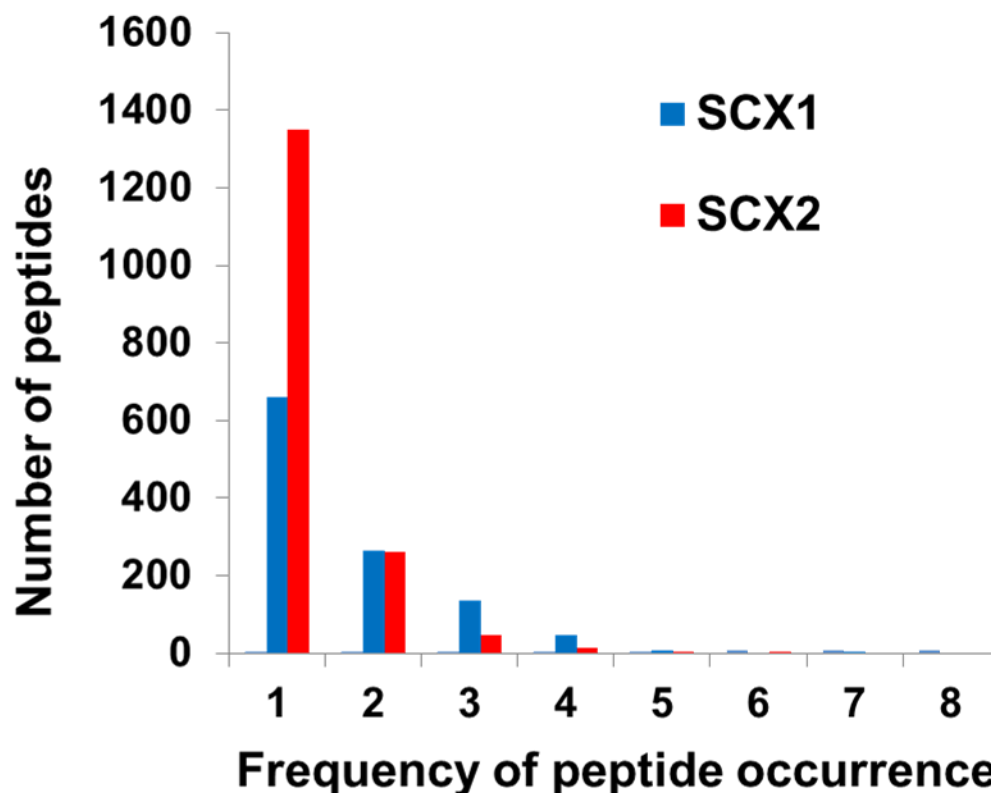
Fig 5.6Comparison SCX resolution efficiency of initial run (SCX1) and optimized gradient run (SCX2) showing better peptide resolution as a result of gradient optimisation.

The optimized SCX gradient parameters were subsequent applied for the fractionation of the iTRAQ experiment discussed in the following section.

5.3.2 Investigation of*E. coli* cellular adaptation strategies

The abundance of proteins present in control cells relative to that of sesquiterpene expressing cells was compared. A total of 5467 peptides and 410 unique *E. coli* proteins with iTRAQ labels were found at 1 % false discovery rate (FDR) and 218 of these had regulatory information. The list of proteins with differential intracellular abundance as identified from Signifiquant is presented in Tables 5.3

and 5.4. Figure 5.7 shows the distribution of the regulated proteins into the different functional groups with significant up regulation of the protein synthetic machinery and metabolic stress response.

Table 5.3 List of proteins differentially regulated between pETDuet (empty) carrying and pETNS1 (terpene synthase)*E. coli* at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change	Function
C6EB40	Chaperone protein DnaK (HSP70)	3.50	Stress protein
C6EGF2	30S ribosomal protein S3	2.65	Translation
C6EJY6	50S ribosomal protein L19	2.37	Translation
C6EGF8	50S ribosomal protein L5	2.27	Translation
C6EE36	50S ribosomal protein L11	2.24	Translation
C6EGG8	30S ribosomal protein S13	2.22	Translation
C6EGE5	30S ribosomal protein S10	2.17	Translation
C6EGG1	50S ribosomal protein L6	2.06	Translation
C6EGH0	30S ribosomal protein S4	2.05	Translation
C6EK89	AhpC component, subunit of alkylhydroperoxide reductase	1.81	Oxidative stress protein
C6EGG4	50S ribosomal protein L30	1.79	Translation
C6EG95	Tryptophanase	1.38	Amino acid metabolism
C6EG74	Glutamine--fructose-6-phosphate aminotransferase	-1.56	Carbohydrate metabolism
C6EEY2	6-phosphofructokinase	-1.67	Energy metabolism
C6EBG1	Transaldolase	-1.72	Pentose phosphate pathway
C6EK07	Autonomous glycyl radical cofactor	-2.04	Energy metabolism

Table 5.4 List of proteins differentially regulated between wild and mutant *E. coli* at 95% confidence level without multiple test correction

Uniprot accession Number	Protein ID	Fold change	Function
C6EI53	OmpF, subunit of outer membrane porin F	4.24	Ion transport
C6EB40	Chaperone protein DnaK (HSP70)	3.50	Stress protein
C6EGF0	30S ribosomal protein S19	2.94	Translation
C6EGF2	30S ribosomal protein S3	2.65	Translation
C6EGE6	50S ribosomal protein L3	2.50	Translation
C6EGG3	30S ribosomal protein S5	2.39	Translation
C6EJY6	50S ribosomal protein L19	2.37	Translation
C6EGF8	50S ribosomal protein L5	2.27	Translation
C6EE36	50S ribosomal protein L11	2.24	Translation
C6EGG8	30S ribosomal protein S13	2.22	Translation
C6EGE5	30S ribosomal protein S10	2.17	Translation
C6EGE7	50S ribosomal protein L4	2.15	Translation
C6EGG1	50S ribosomal protein L6	2.06	Translation
C6EGH0	30S ribosomal protein S4	2.05	Translation
C6EGE9	50S ribosomal protein L2	1.84	Translation
C6EK89	AhpC, subunit of alkylhydroperoxide reductase	1.81	stress protein
C6EGG4	50S ribosomal protein L30	1.79	Translation
C6EEX9	Triosephosphate isomerase	1.72	Glycolysis/Gluconeogenesis
C6EGG5	50S ribosomal protein L15	1.68	Translation
C6EGF5	30S ribosomal protein S17	1.54	Translation
C6EII4	Fructose bisphosphate aldolase monomer	1.53	Glycolysis/Gluconeogenesis
C6EG97	tRNA modification GTPase MnmE	1.38	Translation
C6EG95	Tryptophanase	1.38	Amino acid metabolism
C6EAU1	Dihydrolipoyl dehydrogenase	1.36	Pyruvate metabolism
C6EAB5	Elongation factor Ts	1.34	Translation
C6EE66	Catalase-peroxidase	1.27	Amino acid metabolism
C6EEX2	Glycerol kinase	-1.27	Glycerol metabolism
C5W3D5	Oligopeptide transporter subunit	-1.33	Transporter
C6EKF6	Serine hydroxymethyltransferase	-1.35	Amino acid metabolism

Uniprot accession Number	Protein ID	Fold change	Function
C6ECY6	Aspartate ammonia-lyase	-1.39	Amino acid metabolism
C6EKI4	Peptidase B	-1.43	Proteolysis
C6ED05	Alpha-galactosidase monomer, subunit of alpha-galactosidase	-1.49	Carbohydrate metabolism
C6EI52	Asparagine--tRNA ligase	-1.54	Translation
C6EG74	Glutamine--fructose-6-phosphate aminotransferase	-1.56	Carbohydrate metabolism
C6EEY2	6-phosphofructokinase	-1.67	Glycolysis/Gluconeogenesis
C6EBG1	Transaldolase	-1.72	Pentose phosphate pathway
C6EIG9	S-adenosylmethionine synthase	-1.79	Amino acid metabolism
C6EBX3	Pyruvate kinase	-1.82	Glycolysis/Gluconeogenesis
C6EK07	Autonomous glycyl radical cofactor	-2.04	Energy metabolism
C6EJ92	CTP synthase	-2.78	Nucleotide metabolism
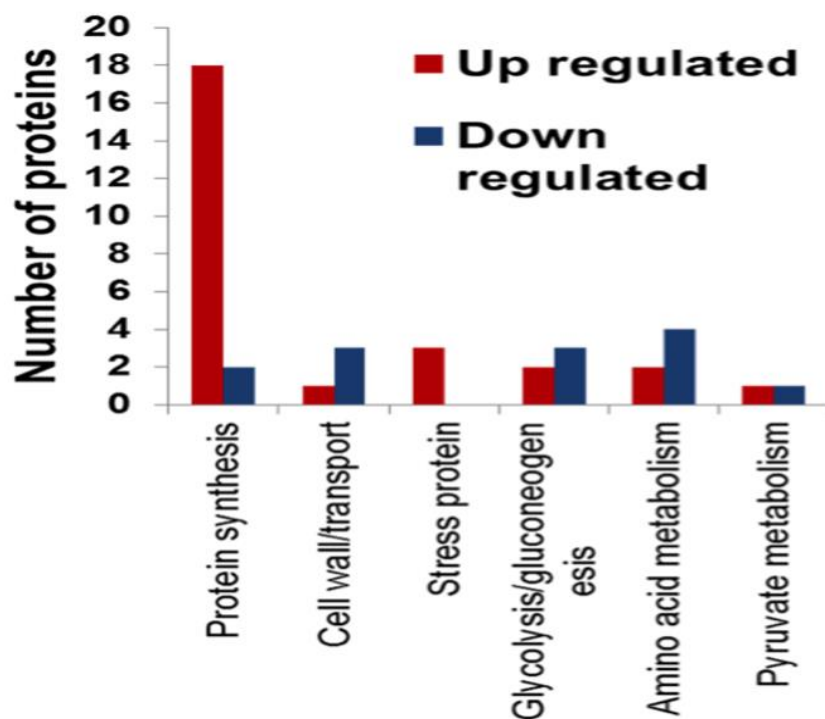
Figure 5.7 Functional annotation of differentially regulated proteins expressed between wild type and mutant *E. coli* expressing sesquiterpene synthase using iTRAQ. Protein annotationwas performed using KAAS (Kegg Automatic Annotation Server).

5.3.2.1 Stress response

Cells are exposed to a variety of environmental stress factors especially during growth. These stress factors are usually caused by changes in pH, temperature, nutrient level, dissolved oxygen, osmolality, as well as chemicals and production of heterologous proteins. Additionally, the metabolic stress level in a recombinant expression system depends also on how much recombinant protein is produced, the size and copy number of the plasmid (Glick, 1995). As seen in Table 5.5, there was significant elevation (> 3 fold) of the heat shock protein DnaK (HSP70). This protein with other heat shock proteins, are involved in proper protein folding especially during heterologous protein over-expression. Additionally, there was an increase in alkylhydroperoxide reductase (1.81) and catalase-peroxidase (1.27). These proteins are responsible for mopping up oxygen free radicals produced during increased respiratory stress, which usually accompanies an increased metabolic load (Christman et al., 1985; Hassan and Fridovich, 1977; Yamamoto et al.,

139

1999). The relative metabolic stress here is due mainly to production of the heterologous sesquiterpene synthase as other factors were constant in both phenotypes.

Table 5.5 List of differentially regulatedproteins characterised as stress proteins

Uniprot Accession Number	Protein ID	Fold change
C6EB40 C6EK89 C6EE66	Chaperone protein DnaK (HSP70) AhpC component, subunit of alkylhydroperoxide reductase Catalase-peroxidase	3.50 1.81 1.27

5.3.2.2 Central carbon metabolism

The central carbon metabolism (CCM) of *E. coli* plays a dominant role in almost all metabolic processes of the organism. Growth and survival are key themes that serve as modulators of metabolism. The CCM is a combination of pathways that are responsible for utilization of carbon sources and energy generation/conservation and these include glycolysis/gluconeogenesis, pentose phosphate pathway and the tricarboxylic acid cycle (TCA) with the glyoxylate shunt. The major carbon source was primarily supplied by the glycerol in the TB growth media. Figure 5.6 shows the regulated enzymes that are involved in the CCM. There is a significant up regulation of triosephosphate isomerase (1.72) which is a bidirectional enzyme that interconverts dihydroxyacetone phosphate (DHAP) to glyceraldehyde 3-phosphate (G3P). Additionally, fructose bisphosphate aldolase, another bidirectional enzyme that converts fructose bisphosphate to DHAP and G3P is also up-regulated (1.53). This enzyme can increase the flux of DHAP and G3P in glycolysis by hydrolysis of fructose bisphosphate or reduce flux of these molecules by forming fructose bisphosphate in the gluconeogenetic pathway. Evidence of repression of glycolysis is seen in the down regulation of the regulatory points (irreversible) at 6-phosphofructokinase (-1.67) and pyruvate kinase (-1.82) and up regulation of the aldolase and triosephosphate isomerase. The up-regulation of triosephosphate

isomerase and aldolase enzymes, point to increase utilization of glycerol (glycerol to DHAP) which is subsequently converted to G3P.

Table 5.6 List of differentially regulatedproteins involved in central carbon metabolism

Uniprot accession Number	Protein ID	Fold change
C6EEX9	Triosephosphate isomerase	1.72
C6EII4	Fructose bisphosphate aldolase monomer	1.53
C6EAU1	Dihydrolipoyl dehydrogenase	1.36
C6EEX2	Glycerol kinase	-1.27
C6ED05	Alpha-galactosidase monomer, subunit of alpha-galactosidase	-1.49
C6EG74	Glutamine--fructose-6-phosphate aminotransferase	-1.56
C6EEY2	6-phosphofructokinase	-1.67
C6EBG1	Transaldolase	-1.72
C6EBX3	Pyruvate kinase	-1.82
C6EK07	Autonomous glycyl radical cofactor (PFL)	-2.04

Dihydrolipoyl dehydrogenase up regulation (1.36) translates to increase generation of acetyl CoA from pyruvate as well as increase in acetoacetate/acetyl CoA from catabolism of amino acids-leucine, valine, isoleucine as well as glycine, serine and threonine. These feed into the TCA cycle which generatesfurther acetoacetate or directly from acetoacetate to gluconeogenesis pathway.

There is also a significant down regulation of the pentose phosphate pathway enzyme transaldolase (-1.72) which is responsible for the reversible conversion of G3P and sedoheptulose 7-phosphate (Sh7P) to fructose 6-phosphate and erythrose 4-phosphate (E4P) which would also serve to increase flux towards G3P and reduce formation of the E4-P needed for downstream synthesis of DNA. There is also a significant down regulation of autonomous glycyl cofactor of the pyruvate formate lyase (-2.04). This enzyme catalyzes the reversible conversion of pyruvate and coenzyme-A to formate and acetyl -CoA

5.3.3.3 Amino acid metabolism

Protein biosynthesis is a complex process that involves many steps from initiation, elongation and termination. Each of these steps requires many factors such as mRNAs, tRNAs, ribosomes, elongation factors and numerous enzymes. As seen in Table 5.7 significant up-regulation of the translation machinery is seen in the subunits of both the 30S and 50S. Increased expression of the protein synthetic machinery is not only needed for making of recombinant protein but also for synthesis of stress response proteins.

There is an increase in tryptophan catabolism as seen in the up regulation of both tryptophanase and catalase-peroxidase. These generate pyruvate, indole and ammonia.
There is a down regulation of Serine hydroxymethyltransferase, Aspartate ammonia-lyase and S-adenosylmethionine synthase which metabolizes serine, aspartate and methionine respectively. Serine hydroxymethyltransferase is also involved in serine and glycine metabolism Aspartate ammonia lyase converts aspartate to fumarate so a down regulation will probably favour the alternate route to oxaloacetate.

Table 5.7 List of differentially regulated proteins involved in protein biosynthesis and amino acid metabolism

Uniprot accession Number	Protein ID	Fold change
C6EGF0	30S ribosomal protein S19	2.94
C6EGF2	30S ribosomal protein S3	2.65
C6EGE6	50S ribosomal protein L3	2.50
C6EGG3	30S ribosomal protein S5	2.39
C6EJY6	50S ribosomal protein L19	2.37
C6EGF8	50S ribosomal protein L5	2.27
C6EE36	50S ribosomal protein L11	2.24
C6EGG8	30S ribosomal protein S13	2.22
C6EGE5	30S ribosomal protein S10	2.17
C6EGE7	50S ribosomal protein L4	2.15
C6EGG1	50S ribosomal protein L6	2.06
C6EGH0	30S ribosomal protein S4	2.05
C6EGE9	50S ribosomal protein L2	1.84
C6EGG4	50S ribosomal protein L30	1.79
C6EGG5	50S ribosomal protein L15	1.68
C6EGF5	30S ribosomal protein S17	1.54
C6EG97	tRNA modification GTPase MnmE	1.38
C6EG95	Tryptophanase	1.38
C6EAB5	Elongation factor Ts	1.34
C6EE66	Catalase-peroxidase	1.27
C6EKF6	Serine hydroxymethyltransferase	-1.35
C6EI89	Serine--tRNA ligase	-1.37
C6ECY6	Aspartate ammonia-lyase	-1.39
C6EKI4	Peptidase B	-1.43
C6EI52	Asparagine--tRNA ligase	-1.54
C6EIG9	S-adenosylmethionine synthase	-1.79

5.3.3.4 Cell envelope

The cell envelope, which comprises of the cell wall and membrane play significant roles in the metabolism of all organisms. The cell envelope is responsible for influx and efflux of substances (nutrients, ions, and molecular species such as oxygen), structural integrity as well as signaling. There is a significant up-regulation of the outer membrane porin OmpF (4.24) which responds to reduction in nutrients and growth retardation as well as acting as an osmoregulator (Buckler et al., 2000; Kenney, 1997; Liu and Ferenci, 2001). This protein is a two component sensor system that helps to maintain intracellular osmolality. However, the oligopeptide ABC transporter that transports oligopeptides into the cell was down-regulated. This protein does not however transport exogenous peptides or protein sourced peptides but, is involved in cycling of cell wall peptidoglycan proteins (Goodell and Higgins, 1987).

The down regulation of glutamine metabolism through Glutamine-fructose-6-phosphate aminotransferase decreases synthesis of amino sugars and peptidoglycan of cell wall backbone. Peptidoglycan synthesis is a complex process that is catalysed by several enzymes. The first committed step involves the conversion of fructose-6 phosphate into glucosamine-6 phosphate by the dimeric enzyme glutamine-fructose-6-phosphate aminotransferase using glutamine as a nitrogen source (Badet-Denisot et al., 1995; Barreteau et al., 2008). The down- regulation of this enzyme (-1.56) indicates decreased synthesis of peptidoglycan. Down regulation of Glutamine-fructose-6-phosphate aminotransferasewould indicate reversal of the reaction and towards glycolysis and to G3-P and pyruvate required by *E. coli* for synthesis of IPP/DMAP through the non-mevalonate pathway.

Table 5.8 List of differentially regulatedproteins involved in cell wall metabolism

Uniprot accession Number	Protein ID	Fold change
C6EI53	OmpF, subunit of outer membrane porin F and The Colicin A Import System	4.24
C5W3D5	Oligopeptide transporter subunit	-1.33
C6EG74	Glutamine-fructose-6-phosphate aminotransferase	-1.56

5.4 Discussion

The aim of this chapter was to examine the metabolic changes that accompany the expression of the sesquiterpene from *Nostoc* sp. PCC 7120 in *E. coli.*

Expression of a heterologous proteins such as sesquiterpene synthase, induces a metabolic burden on the host cell and causes significant perturbation of the host metabolism (Glick, 1995; Hoffmann and Rinas, 2004; Sano, 1990). An iTRAQ-based proteomic assessment of a recombinant sesquiterpene synthase gene containing *E. coli* BL21 (DE3) has been carried out. The results show that metabolic changes in the cell envelope construction, central carbon metabolism as well as amino acid and protein metabolism occur as a result of the expression of sesquiterpene synthase in *E. coli* as depicted in Figure 5.8.

The cell wall of both Gram positive and negative bacteria is made up significantly of peptidoglycan which is heteropolymer complex of N-acetylmuramoyl-peptides and N-acetylglucosamine (GlcNAc) monomers whose synthesis is complex and involves many cytoplasmic and membrane enzymes (Vollmer et al., 2008). The cytoplasmic aspect of the synthetic process involves the synthesis of the monomer component which is subsequently linked to C_{55} undecaprenyl pyrophosphate by the synthase to form lipid II. The C_{55}undecaprenyl pyrophosphate chain is made up of isopentenyl pyrophosphate and dimethylallyl pyrophosphate monomers whose synthetic precursors are generated by the DXP pathway (Bouhss et al., 2008; Pandhal et al., 2011b). These monomers are apparently being hijacked from cell wall synthesis to make the sesquiterpene. As a result, there was a significant up regulation of OmpF signifying nutrient limitation, growth impairment and reduced intracellular osmolality (Kenney, 1997). This might also serve as a route of exporting the potentially toxic sesquiterpene from the cell. In an attempt to meet the increased energy requirements of the cell, other available energy sources are mobilized. This is seen in the reduction of cell wallsynthesis as shown by the down regulation of Glutamine-fructose-6-phosphate aminotransferase (*glmS*). This assertion is further made evident by the significant down-regulation of tetrahydrodipicolinate N-succinyltransferase seen at the 90% confidence level. Tetrahydrodipicolinate N-succinyltransferaseis required in the

lysine/diaminopimelic acid (DAPA) biosynthesis, an important component of the peptidoglycan layer of *E. coli* (Barreteau et al., 2008; Born and Blanchard, 1999). This indicates a redirection of synthetic flux from cell wall synthesis to terpene synthesis.
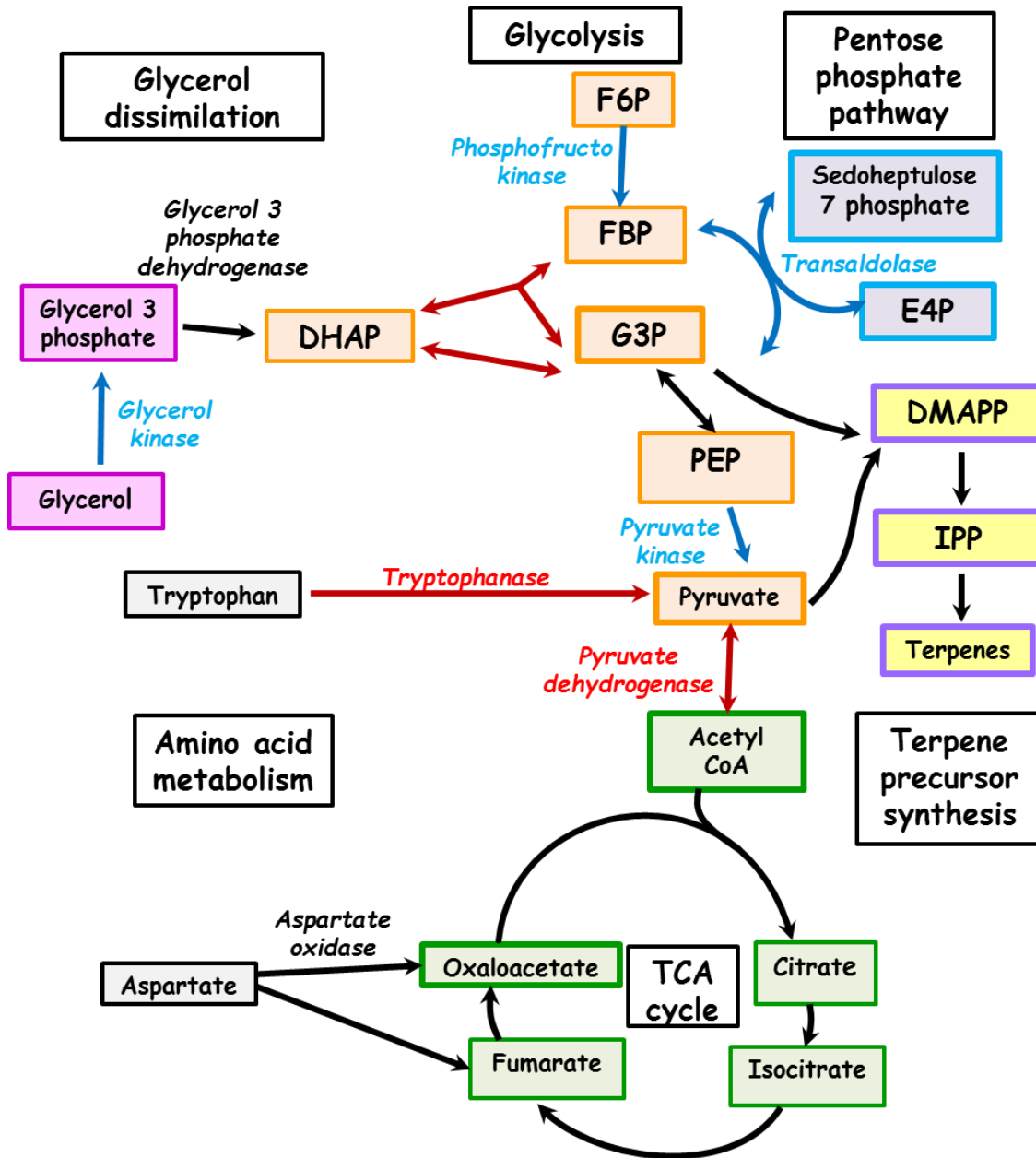
Figure 5.8 Metabolic changes in *E. coli* pETNS relative to pETDUET using iTRAQ based quantitation at 30 h post induction. Red and blue arrows show increased and decreased protein abundance respectively. Red and blue outlined boxes show increased and reduce flux of metabolites and purple boxes show precursors for terpene production. F6P, fructose-6-phosphate; FBP, fructose1,6 bisphosphate; G3P, glyceraldehyde-3-phosphate; DHAP, dihydroxy acetone phosphate; PEP, phosphoenol pyruvate; E4P, Erythrose-4-phosphate; DMAPP, dimethylallyl pyrophosphate; IPP, isopentenyl pyrophosphate.

At the 90% confidence level, Glycerol-3-phosphate transporter is also seen to be significantly up regulated. This protein is responsible for the transport of glycerol-3 phosphate but not glycerol. The glycerol-3 phosphate in this case is sourced from the cell wall and a high concentration of glycerol-3 phosphate reduces the need for the glycerol kinase enzyme (Lemieux et al., 2004; Maurel et al., 1994; Voegele et al., 1993). This is in agreement with the down regulation of glycerol kinase which is responsible for converting intracellular glycerol to glycerol-3 phosphate for subsequent conversion to DHAP. This is in accordance with the postulation of Noor et al (2010) that if a given precursor becomes redundant (with alternative environmental source(s)), a shortcut would evolve to bypasses that precursor and hence its enzyme.This route is likely due to exhaustion of glycerol in the media at 30 hours post induction cultivation time and also leading to the significant up-regulation of OmpF protein.. There is therefore not only a reduction in cell wall synthesis but an active degradation of cell components and mobilization of metabolites from the cell wall.UDP-N-acetylglucosamine pyrophosphorylase (*glmU*) which catalyzes the last two steps in the synthesis of UDP-N-acetyl-alpha-D-glucosamine is in the same operon as *glmS*.

Tryptophanase up regulation leads to increase in pyruvate as well as indole which has an inhibitory role not only in cell growth but in terpenoid biosynthesis as well. Indole has been implicated as an inhibitor of both the isoprenoid pathway as well as growth by Ajikumar et al (2010). The up regulation of tryptophanase which converts tryptophan to pyruvate and indole is seen. The pyruvate from the activity of tryptophanase contributes to the flux of pyruvate.

There is a likely increase flow of metabolites through the CCM towards synthesis of the foreign protein and in so doing; there is consequential reduction in normal metabolic processes of the cell including cell wall synthesis. This affects the growth and viability of the cell. The biosynthesis of the foreign protein is not only energy consuming (in form of ATP and GTP), but also uses up pools of synthetic and metabolic precursors such as amino acids and aminoacyl-tRNAs
There was a relative depression of biosynthetic activity at the time of sampling (30 hours post induction) as shown by the growth curve data. This may be as a result of many factors

including the decrease in the synthesis of aromatic amino acid through the shikimate pathway. The down regulation of the oxidative phase of the pentose phosphate pathway which supplies biosynthetic precursors for nucleotide synthesis will affect replication and growth.

The increase abundase of the unidirectional aspartate oxidase which converts asparttate to oxaloacetate, down regulation of aspartate ammonia-lyase which converts aspartate to fumarate would favour a flux towards oxaloacetate. An increase in the generation of oxaloacetate should favour the gluconeogenetic pathway. If this were true, then, there should be an up regulation of phosphoenolpyruvate carboxykinase enzyme which converts oxaloacetate to PEP. This enzyme was not among regulated proteins at the 95% confidence level. However, on further reduction of the confidence level to 70%, it was seen to be up-regulated (1.34). Since there is no evidence of up-regulation of gluconeogenesis enzymes beyond fructose bisphosphate aldolase, the significance of this would be to increase the flux of G3P for making terpene precursors needed in the DXP pathway for ultimate synthesis of the recombinant sesquiterpene. The down regulation of the both 6-phosphofructokinase and pyruvate kinase which are the major controlling enzymes in the glycolytic pathway indicates a repression of glycolysis.

5.5 Conclusions

In this chapter, iTRAQ-based proteomic method was used to study the effect of a heterologous expression of sesquiterpene synthase form *Nostoc* sp. PCC 7120in *E. coli* compared to pETDuet plasmid control. The findings and conclusions are limited by the poor iTRAQ data and also the lack of metabolomics data.

In order to increase proteome coverage and reduce protein under-estimation, the first dimentional offline SCX-HPLC fractionation was first optimised.

The proteomic data showed evidence of possible growth retardation in the mutant which is in aggrement with the growth curve data in Figure 4.8. This is likely the result of reduced availability of precursors (IPP and DMAPP) which are essential for the biosynthesis of cell wall components. These precursors which are synthesized from pyruvate and G-3-P through the DXP pathway are channelled to

the synthsis of terpene.The utilisation of the energy and metabolic resources of the cell for the synthesis of a foreing protein and metabolite and the subsequent metabolic adaption is responsible for growth retardation that was observed compared to the control.This growth retardation and mobilization of cell resources towards recombinant protein synthesis at the expense of the cell is likely to be ameliorated if an external source of precursors (IPP/DMAPP) could be supplied to the cell. Alternatively, the synthesis of these precursors could be done by supplying the alternate mevalonate pathway to the cells which should relief the burden on the native *E. coli* DXP pathway.In the next chapter, this hypothesis would be tested using an adaptation of the iTRAQ-based proteomics methods used in this chapter.

Systemic profile of a terpenoid biosynthetic pathway in *E. coli*

6.1 Introduction

The heterologous expression of several classes of terpenes has been achieved in *E. coli* and *S. cerevisiae* with subsequent optimisation or metabolic engineering for improved yield. As previously noted, in chapter 5, the expression of terpene synthase in *E. coli* induces a significant burden on the cell with deleterious consequences (Barkovich and Liao, 2001b). In order to circumvent the excess burden and drain that the terpene production imposes on the host cell DXP pathway, several groups have engineered exogenous mevalonate pathway genes from *S. cerevisiae* into hostswith or withoutadditional DXP pathway genes(Atsumi and Liao, 2008; Enfissi et al., 2005; Martin et al., 2003b; Peralta-Yahya et al., 2011a) and with modifying cytochrome P450 (Hisashi Harada et al., 2011). These efforts have been rewarded with improved yield of products, but further refinement and metabolic engineering is still needed in order to achieve a truly commercially viable strain for many of these products. This would indeed require a better understanding of the metabolic perturbations of the exogenous mevalonate pathway genes and indeed other modifying enzymes such as cytochrome P450 and its ferrodoxin-reductase partner. Indeed, pathway performance is a complex interplay of transcription regulation, mRNA abundance, protein abundance, and native pathway consumptions of intermediates/precursors to which final product titres may not be truly representative of (Redding-Johanson et al., 2011a). To this end, a new paradigm of a systemic approach to rational metabolic engineering of terpene/terpenoid is required. This is even more imperative considering the known difficulties of heterologous expression of the downstream processing cytochrome P450 enzymes (Agger et al., 2008). A SRM based proteomic approach to metabolic engineering of a mevalonate overproducing terpene synthetic system was reported by Redding-Johanson et al (2011a) with a more than three-fold increase inamorpha-4,11-dienetiters. Even though a significant improvement was recorded, the objective was centered on optimizing the expression of both mevalonatekinaseandphosphomevalonatekinase which were found to be potential bottlenecks in the expression of the mevalonate pathway. This approach therefore ignored the potentially more important central carbon metabolism, the regulatory mechanisms of *E. coli* and other important global metabolic changes that may affect the final product. It is therefore the objective of this chapter to examine the systemic

changes that are activated in this system which can then serve as tools for further rational engineering. Using an 8-plex iTRAQ, the proteome of four phenotypes are compared using the mutant (N) from chapter 5 as control to check for the effect of the exogenous mevalonate pathway (N+), cytochrome P450 and the ferrodoxin-reductase (P) and also the combined effect of expressing the mevalonate pathway and the P450-ferrodoxin-reductase (T). The pathway for the biosynthesis of sesquiterpenoids with an exogenous mevalonate from yeast is shown in Figure 6.1.
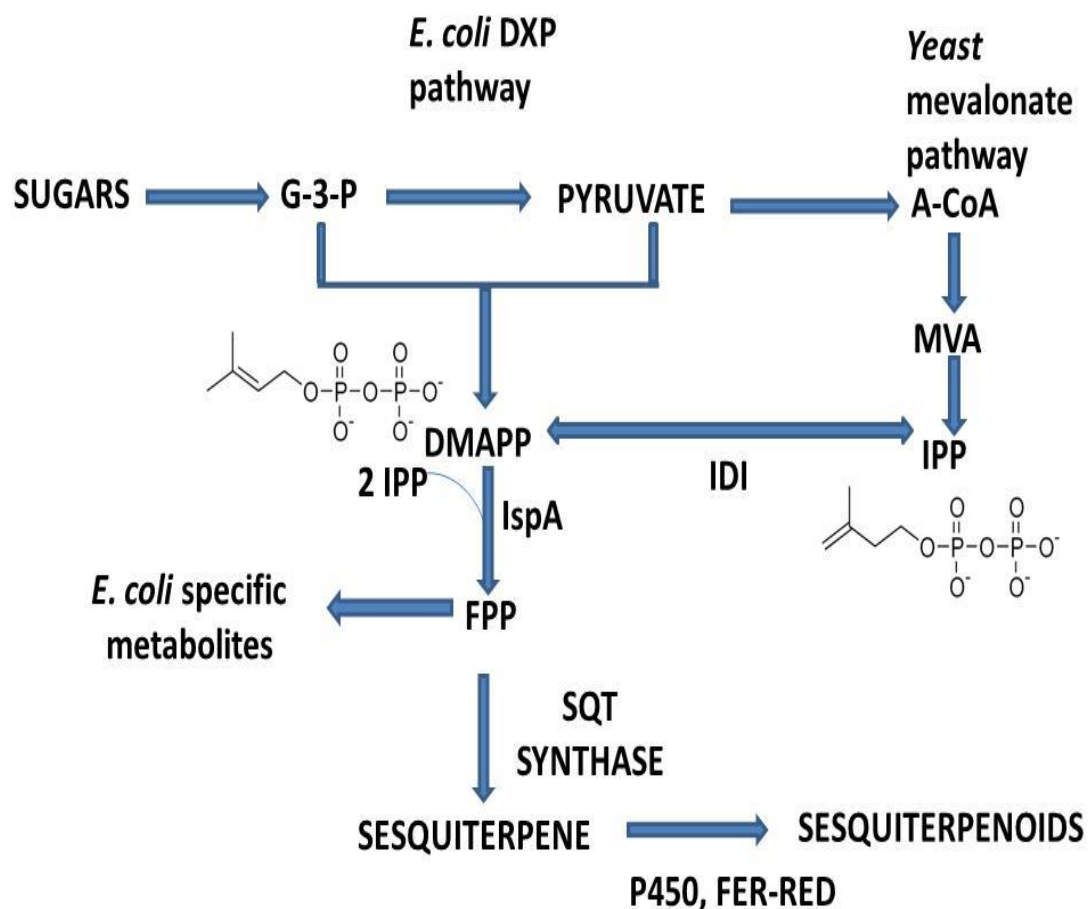
Figure 6.1 Pathway for the biosynthesis of sesquiterpenoids in *E. coli* showing the exogenous mevalonate pathway from yeast and the native DXP pathway for IPP/DMAPP synthesis. IspA, FPP synthase; FPP, farnesyl diphosphate; SQT synthase, sesquiterpene synthase

6.2 Materials and Methods

Several of the methods used in this chapter are identical to those used in chapter 5. However, a few modifications were made for the purpose of further optimization of the entire system.

6.2.1 Culture conditions and cell harvest

Culture was set up and grown as described in section 5.2.1 using phenotypes N (pETDUETNS1 + pBbA5c + pBBRIMCS-2), N$^+$(pETDUETNS1 + pBbA5c-MevT-MBIS + pBBRIMCS-2), P (pETDUETNS-P450NS + pBbA5c + pBBR-FER-RED) and T (pETDUETNS-P450NS + pBbA5c-MevT-MBIS + pBBR-FER-RED) as described in section 4.2.1. Two biological replicates of each phenotype were set up.

6.2.2 Protein extraction and quantification

Proteins from the cultures were extracted and quantified as previously described in section 5.2.2.

6.2.3 SDS-PAGE

SDS-PAGE gel electrophoresis for the purpose of checking for accuracy of protein quantification and post-trypsin digestion assessment was performed as described in section 5.2.3.

6.2.4 In-gel tryptic digestion

In-gel tryptic digestion was performed as described in section 5.2.5.

6.2.5 Solution-phase tryptic digestion

Solution phase tryptic digestion was performed as described in section 5.2.4.

6.2.5.1 RapiGest aided solution-phase tryptic digestion

Biological replicates of each phenotype were digested using 100 µg of each protein sample as detailed in 5.2.4. The digestion was carried out using a final concentration of 0.1% RapiGest surfactant solution (Waters, UK) in 500 mM TEAB according to the manufacturer's instructions. Reduction and alkylation was performed using a final concentration of 10 Mm methyl methanethiolsulfonate (MMTS) and 2.5 mMTris (2-carboxyethyl)phosphine (TCEP) respectively. A protein to trypsin ratio of 1:40 (w/w) was used and incubated at 37°C for 16 hrs after adding a 5% (v/v) of acetonitrile. Each one of the 8 peptide digest solutionswere centrifuged and then divided into two equal parts; one for fractionation optimization and the other was used for iTRAQ labeling.

6.2.6 SCX peptide fractionation

The pH of each peptide solution was reduced to <2 to precipitate the surfactant by adding an appropriate volume of formic acid and centrifuged at 13, 000 rpm for 10 minutes and the supernatant further divided into two equal parts of 25 µg each. These were separately pooled to make two lots of 200µg and vacuum centrifuged. One set of the vacuum dried lot was fractionated using the optimized SCX gradient as detailed in section 5.2.7 while the other was reserved for HILIC fractionation.

6.2.7 Pre-labeling HILIC peptide fractionation

The second set of 200µg of dried peptides were dissolved in buffer A (80% acetonitrile, 10 mM ammonium formate, pH 3) and separated on an Agilent 1100-series HPLC (Agilent, Berkshire UK), coupled to a 200mm PolyHYDROXYETHYL-A (5 µm, 4.6mm ID, 200Å, PolyLC) analytical HILIC column. Using a constant flow rate of 0.5 ml/min, peptides were isocratically eluted on buffer A for 45 min and gradient was started with 0-20% buffer B (5% acetonitrile, 10 mM ammonium formate, pH 5) for 5 minutes, 20-60% B for 50 minutes and 60-100% for 10 minutes and maintained at 100% B for another 10 minutes. The chromatography was monitored at 280 nm and fractions were collected at 1 minute

interval using a Foxy Jr. fraction collector. The fractions were dried in a vacuum concentrator and subjected to LC-MS analysis as detailed in 5.2.8

6.2.8 iTRAQ labeling of peptides

Peptide samples (50 µg each) from section 6.2.5.1 were each labeled with iTRAQ 8-plex labels as described in section 5.2.8. After labeling, the surfactant was removed as described in section 6.2.6. The label for each of the 8 samples is shown in Table 6.1.

Table 6.1 Phenotypic iTRAQ labeling of samples

Phenotype	Replicate	Label	Replicate	Label
N	1	113	2	117
N^+	1	114	2	118
P	1	115	2	119
T	1	116	2	121

6.2.9 HILIC fractionation of labeled peptides

Peptides were then pooled and vacuum dried and subjected to HILIC separation as described in section 6.2.7 and the fractions vacuum dried. Each vacuum dried fraction was dissolved in 25 µl of buffer A (3% (v/v) acetonitrile, 0.1% (v/v) formic acid) centrifuged at 13,000 rpm for 5 minutes and supernatant aspirated for RP-LC-MS/MS analysis.

6.2.10 RP-LC-MS/MS and protein identification

Tandem mass spectrometry was performed on a Quadrupole time-of-flight (Q-ToF) QSTAR XL (Applied Biosystems, MDS-Sciex) equipped with and electrospray ion source fitted with a 15 µm fused-silica emitter tip, PicoTip FS360-75-15-N-5-C15 (New Objective, Woburn, MA). This was coupled to an Ultimate 3000 HPLC system (Dionex, Surrey, UK). Fractions were desalted online using a300 µm i.d x 5 cm trap column packed with C18, 5 µm, 100 Å particles (LC Packings, CA, USA) under 0.1% (v/v) TFA and 3%

(v/v) acetonitrile for 15 minutes, and eluted to a75 µmx15 cm analytical column packed with C18, 5 µm, 100 Å particles (LC Packings, CA, USA) in 0.1%(v/v) formic acid with an acetonitrile gradient extending from 3% to 90%. A 150 min LC program was run at a constant flow rate of 0.3 µl/min starting with 3% buffer B (97% (v/v) acetonitrile in water, 0.1% (v/v) FA) for 20minutes followed by 3- 35% B in 95 minutes then ramped to 90% B in 0.5 minute and held at 90% B for 7 minutes and 3% B for another 28 minutes. A total of two injections of 8 µl each were done to increase coverage and confidence of results. Data acquisition was set in the positive ion mode with a selected mass range of 300 – 2000 m/z. MS/MS was performed on peptides with 2+, 3+ and 4+ charge states and all other parameters as described by Ow et al. (2008b).

6.2.11 Data and quantitative analysis

Tandem MS data generated was converted to MGF using Analyst QS version 1.1 (ABSciex, Foster City, CA) and protein identification and quantification was performed on a Phenyx algorithm server (binary version 2.6;Genebio, Switzerland). Search was performed on an *E. coli* BL21 (DE3) protein database downloaded from Uniprot (www.uniprot.org) on 14/02/2012 using MS tolerance of 0.6 Da, MS/MS tolerance of 0.3 Da, and one missed cleavage of trypsin. Peptide modification was set to include 8-plex iTRAQ mass shifts (+304 Da, K and N-term), methylthiol (+146 Da, C) as fixed modifications, and oxidation of methionine (+116 Da, M) as a variable modification.Peptide level filters were set to a *z*-score of 5.0 and a *p*-value significance of 1×10^{-5}. Analysis of the relative abundance ratio between phenotypes was carried out using the significance testing algorithm, Signifiquant (S. Y. Ow et al., 2009) and only proteins identified by two or more peptides at 95% confidence were considered with or without Bonferonni correction. Relative increase or decrease in abundance of protein is measured by fold changes above and below a factor of 1. The identified proteins were mapped onto colour coded pathways using the search and colour pathway of KEGG mapper (http://www.genome.jp/kegg/tool/map_pathway2.htmlorganism code – ebe). All identified proteins were organised into functional groups using annotation based on cluster of orthologous groups (COG).

6.3 Results

6.3.1 Overview of the proteomic analysis workflow and optimisations

The general proteomic workflow for the 8-plex iTRAQ is presented in Fig 6.2. An optimized protein extraction method as described in 4.2.2 was used and further optimisations were carried out for trypsin digestion and peptide fractionation.
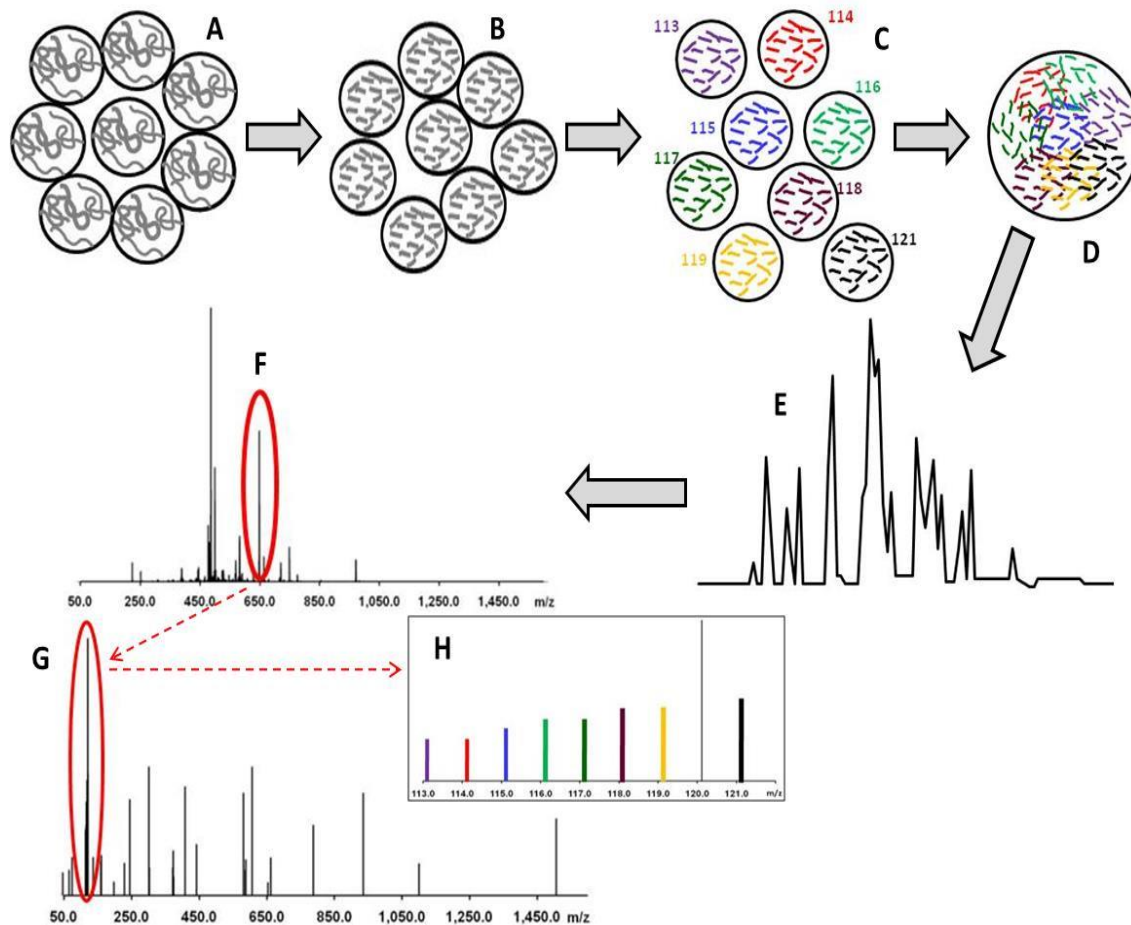
Fig 6.2 An overview of 8-plex iTRAQ proteomics workflow for the with points of optimization at B and E. A, protein extraction and quantitation; B, trypsin digestion; C, peptide labelling of individual samples; D, pooling of samples and vacuum drying; E, offline HILIC fractionation; F, Q-ToF MS scan; G and H, low mass zone showing 8-plex reporter ions

6.3.1.1 SDS PAGE analysis of extracted proteins

Extracted protein samples for 8-plex iTRAQ were subjected to SDS-PAGE analysis as described in 5.3.1.2 using 10 µg aliquots. As seen in Fig 6.3, there is uniformity in the intensity of the lanes showing the accuracy of the concentration measurements. There is also evidence of phenotypic differences in the banding pattern of the samples.
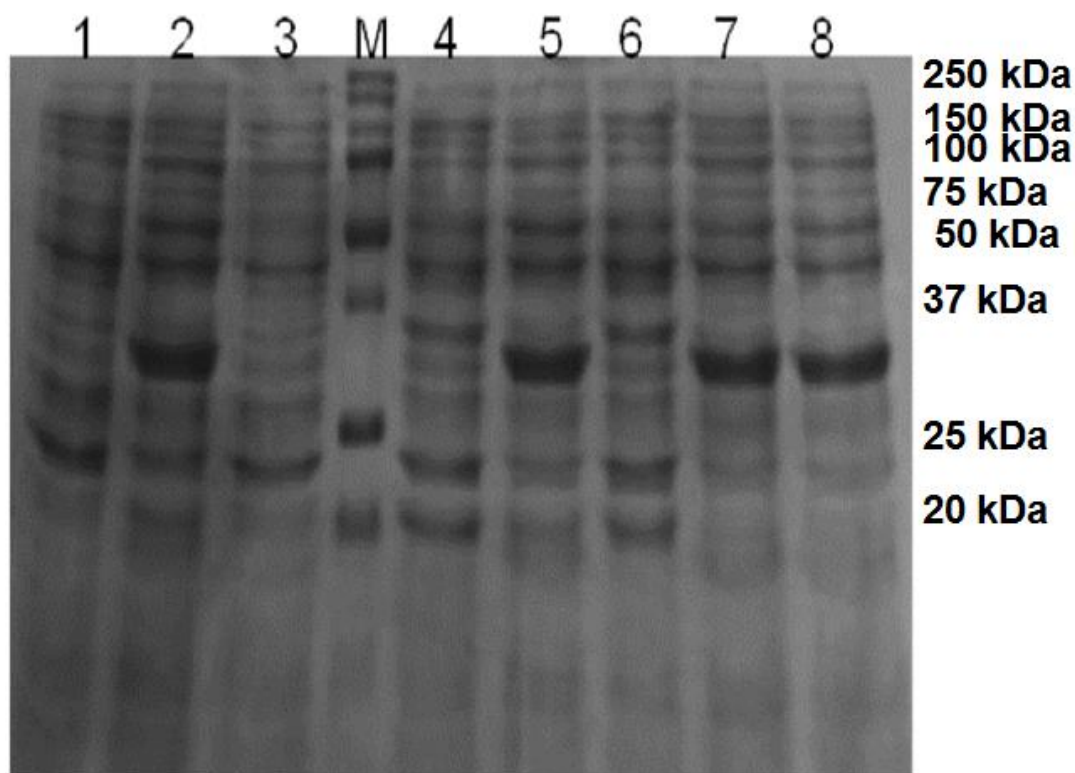
Fig 6.3 SDS-PAGE Analysis of 10 µg of each protein sample from transformed *E. coli* cells showing equal loading. N (1 & 3), N+ (2 & 5), P (4 & 6), T (7 & 8)

6.3.1.2 Testing tryptic digestion methods

Further optimization of tryptic digestion was carried out by comparing in-gel, solution-phase and surfactant aided in-solution digestion protocols. As seen in Table 6.2, the surfactant aided in-solution digestion method yielded significantly more protein hits of the representative phenotype as compared both to in-gel digestion and solution-phase tryptic digestion without surfactant. To ensure that the surfactant

aided solution-phase tryptic digestion was complete, aliquots equivalent to 5 µg of each representative phenotype was run on an SDS-PAGEwith evidence of complete digestion of each sampleas seen in Fig 6.4. The stacking gel aided tryptic digestion showed significantly good protein hits on LC-MS analysis. In the subsequent experiment, the in-gel digestion was again compared to in-solution digestion with the aid of surfactant. The use of surfactant in in-solution digestion gave a remarkable improvement over the stacking gel aided digestion. This method was used subsequently.
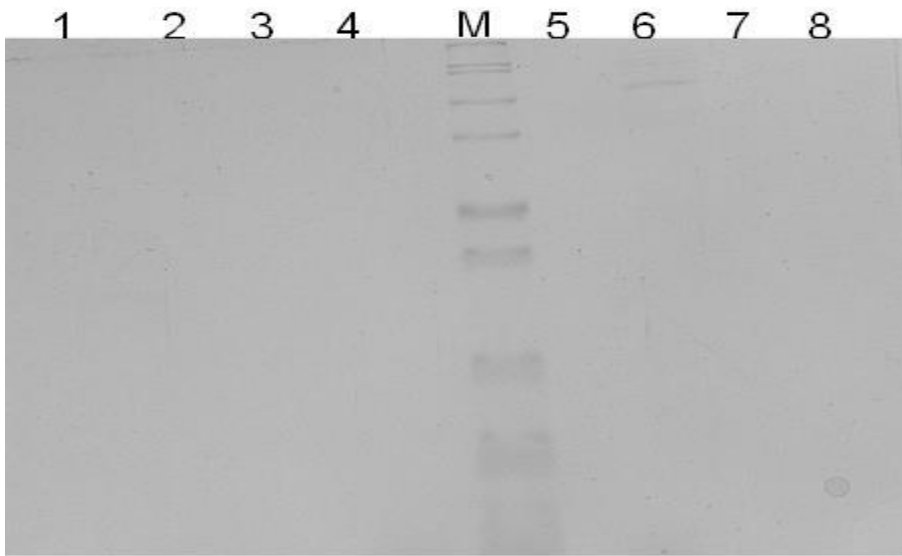
Fig 6.4 SDS-PAGE analysis of a volume of digestion solution equivalent to 5 µg. lanes 1-8 (protein digest), M (marker).

6.3.1.3 Optimization of fractionation methods

SCX was used in the first experiment as detailed in 5.3.1.4 and subsequently compared with HILIC in this experiment. A significant improvement was observed as measured by the separation and resolution efficiency, peptide distribution profile as well as the proteome coverage. Fig 6.5 shows the peptide distribution of HILIC fractionation and Fig 6.6 shows a direct comparison of peptide distribution profile from the same sample between HILIC and SCX. The HILIC fractionation shows significantly better distribution of unique peptides across the spectrum as well as

greater number of unique peptides in each fraction. In terms of resolution efficiency, HILIC was seen to have a significantly better resolution compared to SCX. As seen in Fig 6.7, more peptides where resolved into a single fraction (81%) using HILIC fractionation as opposed to SCX (71%). This is significant enough to affect under-estimation as a result of compression ratios (S. Y. Ow et al., 2009). Even though the SCX gradient has been so optimized to increase the resolution efficiency significantly as seen in Fig 5.6, the difference in resolution efficiencies between HILIC and SCX (10%) in this experiment gave a significant difference in the total peptide and protein numbers as seen in Table 6.2
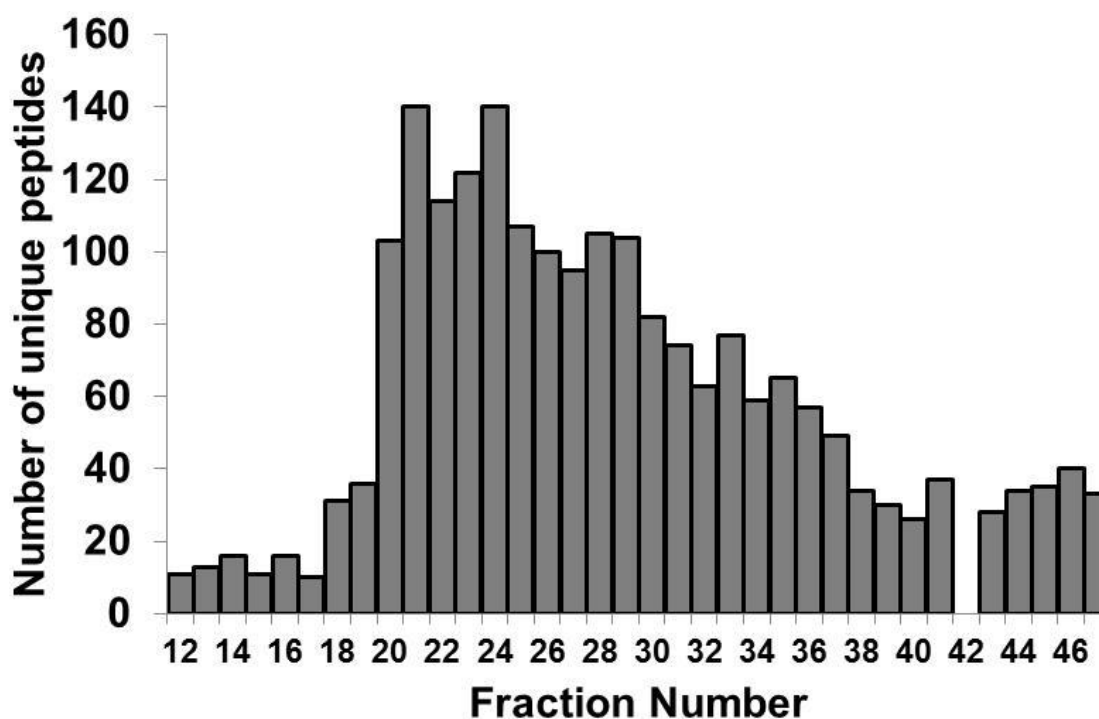
Fig 6.5 Distribution of unique peptides from HILIC separation showing number of unique peptides per fraction collected. Fractions were collected every 1 minute interval. Fraction 42 was lost during processing.
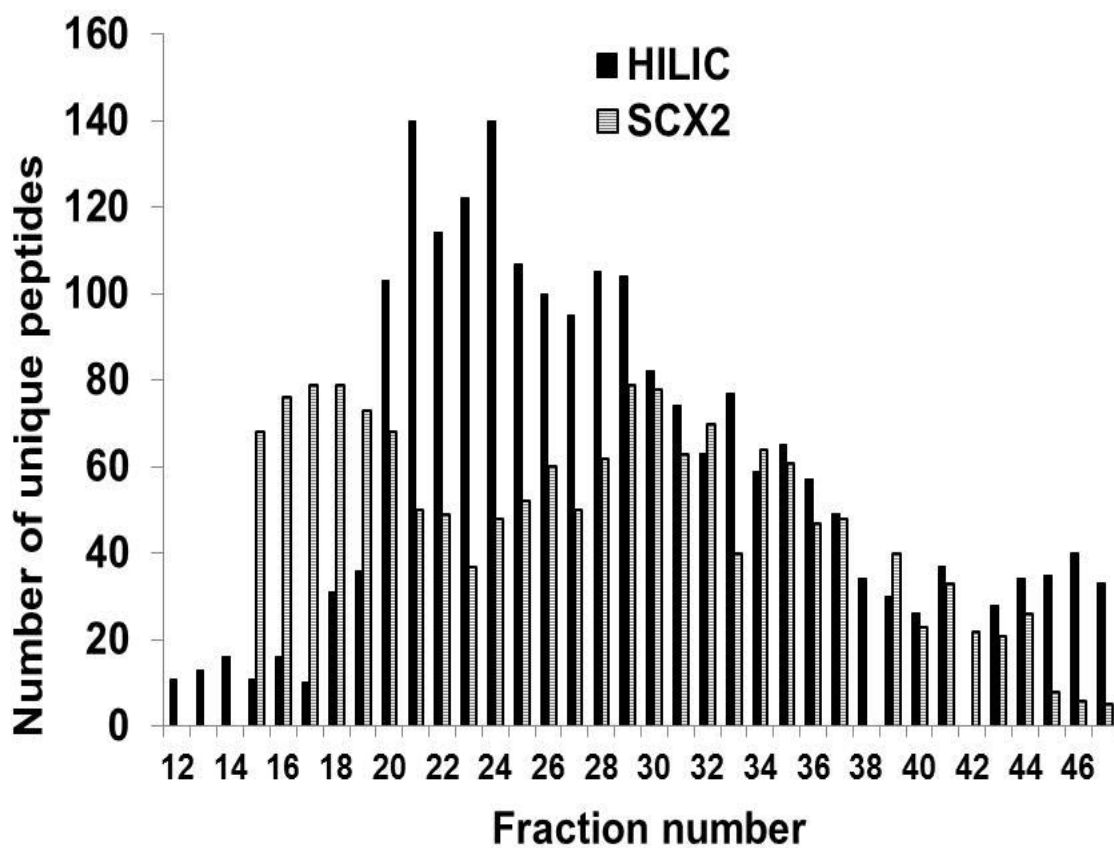
Fig 6.6 comparison of SCX and HILIC peptide distribution showing better peptide distribution profile and increased number of peptides per fraction in HILIC.
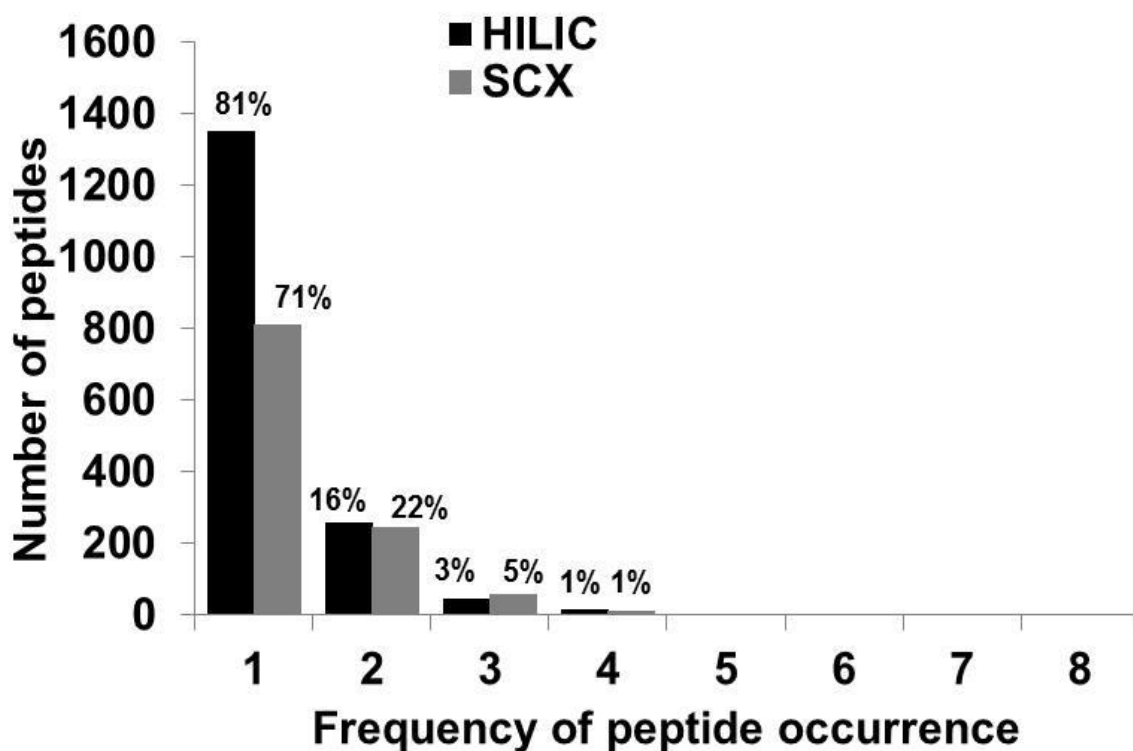
Fig 6.7 comparison between HILIC and SCX resolution efficiencies showing that most peptides were resolved into one fraction in HILIC (81%) compared to SCX fractionation (71%)

Table 6.2 comparison of performance parameters between HILIC and SCX fractionation methods

Parameter	HILIC	SCX	Difference in performance
Resolution efficiency	81%	71%	10%
Number of peptides	2097	575	3.6 fold
Total proteins	646	244	2.6 fold

6.3.2 Overview of iTRAQ results and presentation outline

Four phenotypes in duplicates were examined and compared in order to understand the adaptation strategies and metabolic changes induced by the heterologous protein expression. A total of 16613 peptides and 714 unique *E. coli* proteins with iTRAQ labels were found at 1% false discovery rate (FDR) with FDR Slice (version 4.0). From this, 582 proteins were found to have regulatory information as seen in the global metabolic map in Figure 6.8, using KEGG mapper to show pathways that are covered by these proteins.

In order to understand the metabolic changes and the dynamics of these changes between phenotypes, comparisons were done between the phenotypes to answer specific questions. To understand the metabolic effects of the mevalonate pathway and the P450/ferrodoxin-reductase on the terpene producing mutant (N), each one was independently compared to N. To further our understanding of the metabolic spectrum, N+ was compared to T and finally a comparison was done between P and T. Since the objective is to get a better understanding of the metabolic and regulatory mechanisms responsible for the expression or otherwise of the desired products as seen in chapter 3; the comparison between N and N+ was used as a reference point for reporting the other comparisons and as template for optimization.

The cluster analysis dendogram seen in Figure 6.9 shows a clear aggregation of all biological replicates as well as wide phenotypic differences between the control (N) and the test samples. The phenotypic divergence is greatest between N and N+ followed by N and T which further validates the rationale for using N vs N+ as a reference point.
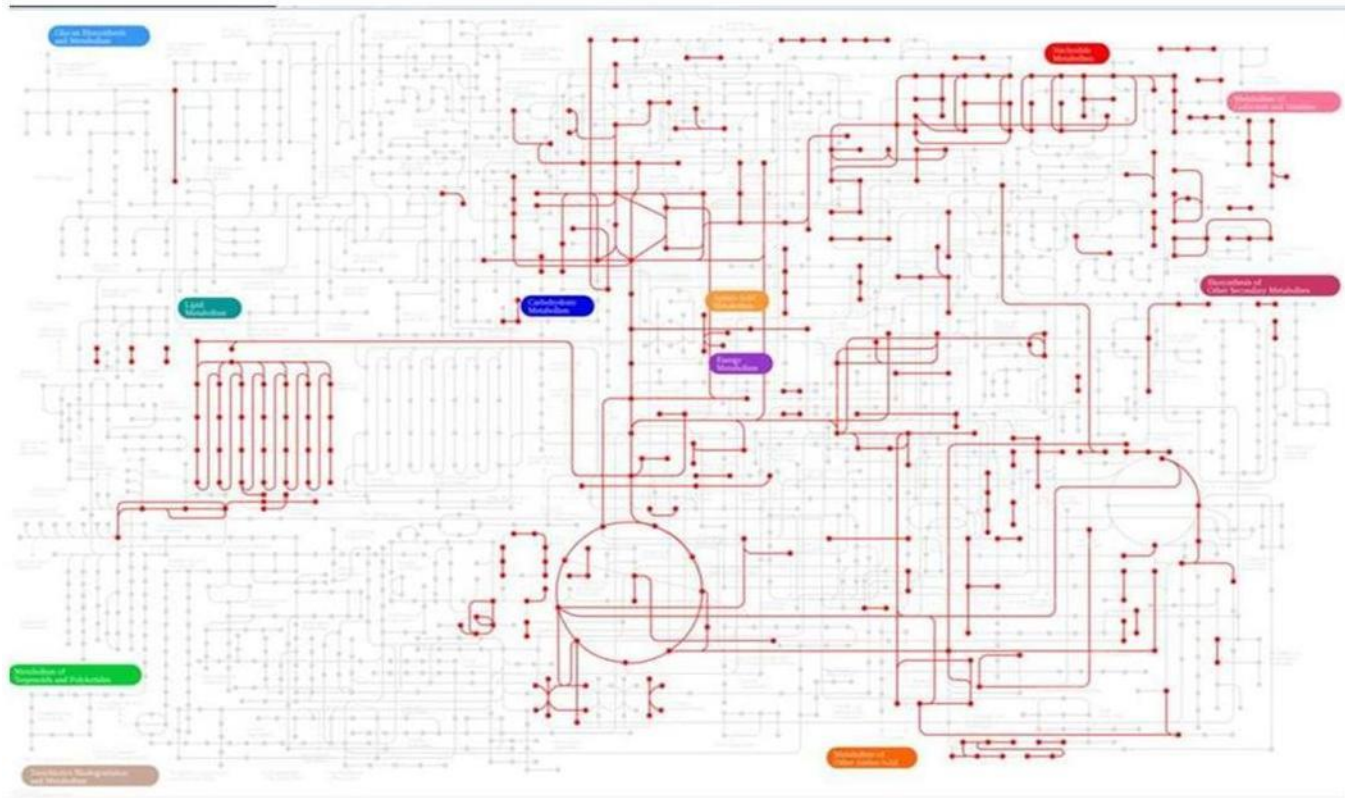
Fig 6.8 Global map of pathwaysenzymes in _E. coli_ with all differentially regulated proteins identified in the study shown in red. Mapping was done using KEGG colour mapper.
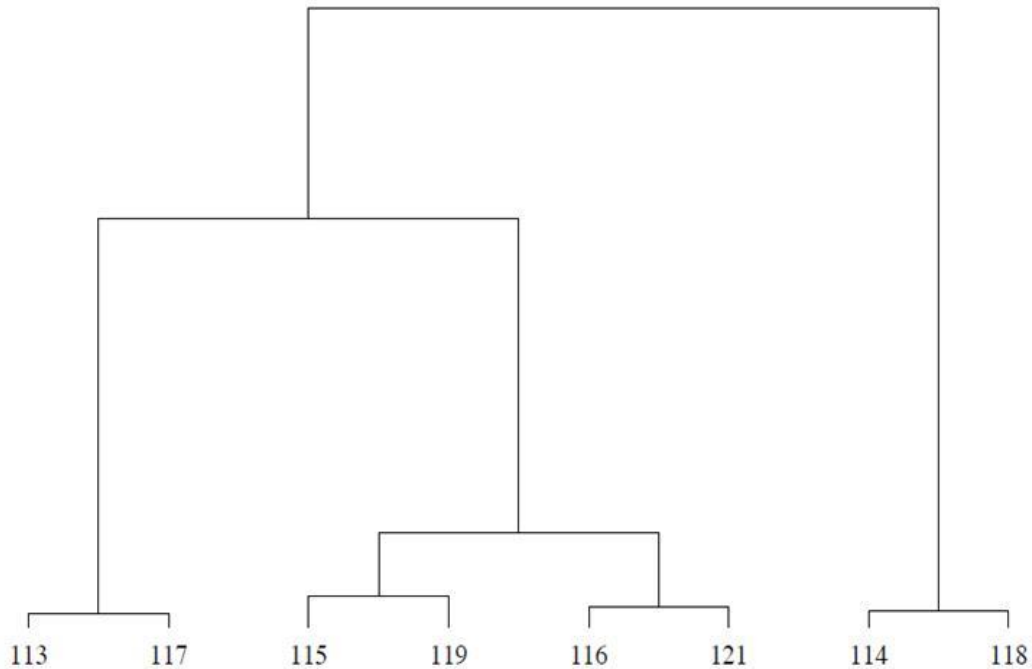
Fig 6.9 cluster analysis data of 8-plex iTRAQ labeled biological phenotypes showing biological replicates (113/117, 115/119, 116/121 and 114/118) clustered together as well as phenotypic segregation.

6.3.3 Investigating *E. coli* adaptation strategies to mevalonate expression

In order to investigate the adaptation strategies of a terpene expressing *E. coli* mutant to the addition of a heterologous eukaryotic mevalonate pathway, both phenotypes were compared. The comparison between N and N+ would show the metabolic and regulatory changes needed for increased production of the sesquiterpene as well as possible optimization steps or points to explore for further strain improvement. A list of differentially regulated proteins between N and N+ is seen in Table 6.3. A total of 104 proteins were differentially regulated between the control terpene mutant (N) and the mevalonate pathway mutant.

6.3.3 Investigation of *E. coli* cellular adaptation strategies to mevalonate expression

Cellular adaptation strategies of *E coli* mutant (N+) expressing terpene synthase and the eukaryotic mevalonate pathway from *S. cerevisiae* against a mevalonate negative strain (N) was investigated using an 8-plex iTRAQ. The comparison between N and N+ would show the metabolic and regulatory changes for increased production of the sesquiterpene as well possible optimization steps or points to explore for further strain improvement. A total of 104 proteins as seen in Table 6.4 either showed increased or reduced abundance at the 95% confidence level with multiple test correction applied. These proteins are further classified into major functional groups and metabolic pathways as seen in Figure 6.10. A further 147 unique differentially regulated proteins were discovered at the 95% confidence when the Bonferroni correction was turned off, as seen in Table 6.5.

Table 6.3 List of proteins differentially regulated between N and N+ at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change	Function
C6EIL9	Isopentenyl-diphosphate Delta-isomerase (IPP isomerase)	29.01	Terpenoid biosynthesis
C6EL48	Farnesyl diphosphate synthase/geranyl diphosphate synthase	21.24	Terpenoid biosynthesis
C6EL58	Predicted protein (Putative uncharacterized protein yfeK)	10.35	Uncharacterised
C6EHL8	Cold shock protein CspG	5.29	Stress
C6E9W2	Glycerophosphoryl diester phosphodiesterase	5.23	Glycerol metabolism
C6ECC2	Inorganic pyrophosphatase	4.98	Oxidative phosphorylation
C6EGB7	Peptidyl-prolyl cis-trans isomerase	4.94	Protein folding
C6EK14	RNA polymerase sigma factor	4.65	Transcription regulation
C6EDY9	Extracellular solute-binding protein family 1 (MalE, subunit of maltose ABC transporter)	4.46	Transporter
C6ED05	Alpha-galactosidase monomer, subunit of alpha-galactosidase	4.36	Galactose metabolism
C6EEN8	NAD-dependent malic enzyme (NAD-ME)	4.12	TCA
C5W9Z4	Phosphorylase	4.11	Starch and sucrose metabolism

Uniprot Accession Number	Protein ID	Fold change	Function
C6ECU7	Superoxide dismutase	3.08	Stress
C6EHL0	Molybdopterin guanine dinucleotide-containing S/N-oxide reductase	3.05	Two-component/membrane
C6EIK5	Glycine dehydrogenase [decarboxylating]	2.86	Amino acid metabolism
C6EAU3	Pyruvate dehydrogenase E1 component	2.67	Pyruvate metabolism
C6EKP9	Malate dehydrogenase (NADP(+))	2.51	Pyruvate metabolism
C6EJL8	Citrate synthase	2.47	TCA
C6EJJ0	Phospho-2-dehydro-3-deoxyheptonate aldolase (3-deoxy-D-arabino-heptulosonate 7-phosphate synthase)	2.47	Amino acid biosynthesis/shikimate pathway
C6EGM8	Acetaldehyde dehydrogenase/PFL-deactivase	2.35	Pyruvate metabolism
C6EGZ6	Malate dehydrogenase	2.30	TCA
C5W1L3	Lac repressor (LacI transcriptional repressor)	2.28	Lactose metabolism
C6ECZ5	Lysine--tRNA ligase	2.23	Translation
C6EHG4	Glucans biosynthesis protein G	2.12	stress
C6EKC3	6,7-dihydropteridine reductase	2.12	Nitrotoluene degradation
C6ELD0	3-oxoacyl-(Acyl carrier protein) synthase	2.11	Fatty acid biosynthesis
C6ECP3	Phenylalanine--tRNA ligase alpha subunit	2.07	Translation
C6EAU2	Dihydrolipoamide acetyltransferase	2.04	Pyruvate metabolism
C6ELF7	Acetate kinase (Acetokinase)	1.98	Pyruvate/acetate metabolism
C6EAL5	D-tagatose-1,6-bisphosphate aldolase subunit GatZ	1.97	Galactose metabolism
C6EGF8	50S ribosomal protein L5	1.97	Translation
C6EIH4	Transketolase (Transketolase 1, thiamin-binding)	1.96	Pentose phosphate pathway
C6EB00	Chaperone SurA (Peptidyl-prolyl cis-trans isomerase	1.93	stress
C6EEC4	Cold-shock DNA-binding domain protein (CspA transcriptional activator)	1.91	Stress
C6EAH8	D-lactate dehydrogenase NADH independent)	1.84	Pyruvate metabolism
C6ECA9	Alpha,alpha-phosphotrehalase (Trehalose-6-P hydrolase)	1.82	Starch metabolism
C6ED16	Alkylphosphonate utilization operon protein PhnA	1.81	Phosphonate/phosphinate metabolism
C6EAL9	Alcohol dehydrogenase GroES domain protein (Galactitol-1-phosphate dehydrogenase)	1.81	Galactose metabolism
C6EAU1	Dihydrolipoyl dehydrogenase	1.81	Pyruvate metabolism
C6ECY6	Aspartate ammonia-lyase	1.79	Amino acid metabolism

Uniprot Accession Number	Protein ID	Fold change	Function
C6EGE6	50S ribosomal protein L3	1.77	Translation
C6EBG1	Transaldolase	1.76	Pentose phosphate pathway
C5W9B3	Polyribonucleotide nucleotidyltransferase	1.73	RNA degradation/DNA metabolism
C6ELF6	Phosphate acetyltransferase	1.70	Pyruvate metabolism
C6EDN5	Mannose-6-phosphate isomerase	1.69	Fructose metabolism
C6EI89	Serine--tRNA ligase	1.67	Translation
C6ECS9	Pyruvate kinase	1.65	Glycolysis/gluconeogenesis
C5W0U2	AhpF component, subunit of alkylhydroperoxide reductase	1.59	Stress
C6EJT4	Alanine--tRNA ligase	1.55	Translation
C6EGV5	Isocitrate dehydrogenase [NADP]	1.55	TCA
C6EII3	Phosphoglycerate kinase	1.54	Glycolysis/gluconeogenesis
C6ECP1	50S ribosomal protein L20	1.53	Translation
C6EKJ5	Histidine--tRNA ligase	1.52	Translation
C6EGG2	50S ribosomal protein L18	1.45	Translation
C6EKI4	Peptidase B	1.44	Proteolysis
C6EIK4	Glycine cleavage system H protein	1.41	Amino acid metabolism
C6EAB9	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase	1.40	Amino acid/peptidoglycan biosynthesis
C6EAA0	Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha	1.37	Pyruvate metabolism
C6EK89	AhpC component, subunit of alkylhydroperoxide reductase	1.36	Stress
C6EBI7	Purine nucleoside phosphorylase DeoD-type (PNP)	1.31	Nucleotide metabolism
C6EIL8	Lysine--tRNA ligase	1.29	Translation
C6EE39	Elongation factor Tu (EF-Tu)	1.14	Translation
C6EE33	50S ribosomal protein L7/L12	-1.18	Translation
C6EH56	NusA antitermination factor	-1.23	Transcription regulation
C6EH57	Translation initiation factor IF-2	-1.27	Translation
C6EGF5	30S ribosomal protein S17	-1.28	Translation
C6EE32	DNA-directed RNA polymerase subunit beta	-1.32	Transcription regulation
C6EG71	ATP synthase subunit beta (ATP synthase F1 sector subunit beta) (F-ATPase subunit beta)	-1.33	Oxidative phosphorylation
C6EGF3	50S ribosomal protein L16	-1.37	Translation
C6EB56	6-phosphogluconate dehydrogenase, decarboxylating	-1.39	Pentose phosphate pathway
C6EG69	ATP synthase subunit alpha (ATP synthase F1 sector subunit alpha) (F-ATPase subunit alpha)	-1.41	Oxidative phosphorylation
C6EL60	Crr, subunit of enzyme II [glc], trehalose PTS permease,	-1.41	Glycolysis/gluco

Uniprot Accession Number	EIIBCMalX and N-acetylmuramic acid PTS permease		neogenesis
	Protein ID	**Fold change**	**Function**
C6EGG4	50S ribosomal protein L30	-1.43	Translation
C6EGE8	50S ribosomal protein L23	-1.45	Translation
C6EGG5	50S ribosomal protein L15	-1.45	Translation
C6EGE7	50S ribosomal protein L4	-1.47	Translation
C6EGC3	30S ribosomal protein S7	-1.47	Translation
C6EB28	Bifunctional riboflavin kinase / FMN adenylyltransferase	-1.50	Riboflavin metabolism
C6EG32	Thioredoxin	-1.50	Stress
C6EDF4	50S ribosomal protein L28	-1.52	Translation
C6EGG3	30S ribosomal protein S5	-1.52	Translation
C6EGH2	50S ribosomal protein L17	-1.54	Translation
C6EJY3	30S ribosomal protein S16	-1.56	Translation
C6EI70	Integration host factor subunit beta (IHF-beta)	-1.59	Transcription regulation
C6EI25	OmpA domain protein transmembrane region-containing protein	-1.61	Transporter
C5WBA2	Thiol:disulfide interchange protein DsbA	-1.67	Stress/protein folding
C6EJT9	S-ribosylhomocysteine lyase (AI-2 synthesis protein) (Autoinducer-2 production protein LuxS)	-1.67	Amino acid metabolism/quorum sensing
C6EBI8	Phosphopentomutase	-1.67	Pentose phosphate pathway
C6EHT0	30S ribosomal protein S21	-1.75	Translation
C6ED65	Single-stranded DNA-binding protein	-1.75	DNA replication
C6EC07	Cold-shock DNA-binding domain protein (Stress protein, member of the CspA family	-1.75	Stress
C6ECE9	30S ribosomal protein S6	-1.79	Translation
C6EE18	HU, DNA-binding transcriptional regulator, alpha subunit	-1.89	Transcription regulation
C6EJ93	Enolase (2-phospho-D-glycerate hydro-lyase) (2-phosphoglycerate dehydratase)	-2.00	Glycolysis/gluconeogenesis
C6EGH0	30S ribosomal protein S4	-2.00	Translation
C6EFW4	Predicted protein (Putative uncharacterized protein yciN)	-2.17	Uncharacterised
C6ECP5	Integration host factor subunit alpha (IHF-alpha)	-2.33	Transcription regulation
C6EKZ0	Adenylate kinase (ATP-AMP transphosphorylase)	-2.38	Energy metabolism/Nucleotide synthesis
C6EJ54	CSD sulfur transfer protein	-2.5	Amino acid metabolism
C6EC51	Peptide methionine sulfoxide reductase MsrB	-2.5	Protein repair
C6EIX0	DNA protection during starvation protein	-3.03	DNA damage
C6EG51	D-ribose transporter subunit	-5.00	Transporter

Table 6.4 List of additional proteins that are differentially regulated between N and N+ at 95% confidence level without multiple test correction

Uniprot Accession Number	Protein ID	Fold change
C6EAC3	Protease Do (Serine protease Do)	5.29
C6EJZ3	Chorismate mutase	4.35
C6EAL7	Galactitol-specific enzyme IIB component of PTS (GatB, subunit of galactitol PTS permease)	4.33
C6EBZ9	Carboxy-terminal protease for penicillin-binding protein 3	2.96
C6EGF9	30S ribosomal protein S14	2.88
C6EAL6	Galactitol-specific enzyme IIA component of PTS (GatA, subunit of galactitol PTS permease)	2.68
C6EJ77	Conserved protein (Putative uncharacterized protein ygdH)	2.57
C6EAJ8	Antiporter inner membrane protein (Mrp protein)	2.24
C6EK45	Pyrimidine-specific ribonucleoside hydrolase RihA	2.18
C6EJV6	DNA binding protein, nucleoid-associated (H-NS-like DNA-binding protein	2.12
C6EEX0	Cell division protein ZapB	2.11
C6EKJ1	Nucleoside diphosphate kinase (NDK) (NDP kinase)	2.09
C6EDI0	Protein-export protein SecB	2.02
C6EGY5	Beta-hexosaminidase (N-acetyl-beta-glucosaminidase)	2.02
C6EHB9	3-oxoacyl-[acyl-carrier-protein] synthase 2	2.00
C6EB72	Exodeoxyribonuclease I	2.00
C6EHZ8	Cytochrome-c3 hydrogenase	1.93
C6EIJ6	D-3-phosphoglycerate dehydrogenase	1.91
C6EH47	Predicted RNA-binding protein	1.88
C6EAS3	3-methyl-2-oxobutanoate hydroxymethyltransferase	1.86
C6EHW4	Antibiotic biosynthesis monooxygenase (Quinol monooxygenase))	1.84
C6EKK2	GMP synthase [glutamine-hydrolyzing] (Glutamine amidotransferase)	1.80
C6EGG7	50S ribosomal protein L36	1.80
C6EGR9	Dihydroxyacetone kinase subunit M, subunit of dihydroxyacetone kinase	1.79
C6EHC1	3-oxo-acyl-[acyl-carrier-protein] reductase	1.77
C6EGA1	50S ribosomal protein L34	1.76
C6EAB7	Methionine aminopeptidase (EC 3.4.11.18)	1.75
C6EK18	Elongation factor 4 (EF-4) (Ribosomal back-translocase LepA)	1.74
C6ELH6	Protein translocase subunit SecD	1.64
C6ECP4	Phenylalanine--tRNA ligase beta subunit	1.62
C6EIL5	DsbC[reduced], subunit of protein disulfide oxidoreductase	1.61
C6EKN1	Peroxiredoxin	1.59
C6EL61	Phosphoenolpyruvate-protein phosphotransferase (Phosphotransferase system, enzyme I)	1.54
C6EK23	Pyridoxine 5'-phosphate synthase (PNP synthase)	1.53
C6EFU4	Enoyl-[acyl-carrier-protein] reductase [NADH]	1.52
C6ELL5	Delta-aminolevulinic acid dehydratase	1.49
C6EEI7	Oligopeptidase A	1.47
C6EJL1	2-oxoglutarate dehydrogenase, E2 subunit, dihydrolipoamide succinyltransferase	1.47
C6EBZ8	ProP effector	1.45
C6EIK3	Aminomethyltransferase (Glycine cleavage system T protein)	1.45
C6EH42	50S ribosomal protein L27	1.44

Uniprot Accession number	Protein ID	Fold change
C6VCK1	ATP-dependent helicase	1.42
C6EE53	Phosphoenolpyruvate carboxylase (PEPC)	1.40
C6EGL0	Cell shape determining protein, MreB/Mrl family	1.38
C6EI53	OmpF, subunit of outer membrane porin F and The Colicin A Import System	1.36
C6EIG9	S-adenosylmethionine synthase (AdoMet synthase)	1.36
C6EBF4	Ferritin iron storage protein (Cytoplasmic)	1.35
C6EKJ4	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	1.34
C6EKZ1	Chaperone protein HtpG (Heat shock protein HtpG)	1.33
C6EDZ9	Glucose-6-phosphate isomerase (GPI)	1.32
C6ECJ2	Nitroreductase (Predicted oxidoreductase)	1.30
C6ED67	Conserved protein (Putative uncharacterized protein yjbR)	1.30
C6EI75	Phosphoserine aminotransferase	1.28
C6EG30	Transcription termination factor Rho (ATP-dependent helicase Rho)	1.27
C6EII4	Fructose bisphosphate aldolase monome	1.27
C6EBI0	ABC transporter related (Fused predicted transporter subunits of ABC superfamily: ATP-binding components) (YjjK)	1.24
C6EJT2	Protein RecA (Recombinase A)	1.24
C6EGA6	CRP transcriptional dual regulator (DNA-binding transcriptional dual	1.23
C6EJL0	Succinyl-CoA ligase [ADP-forming] subunit beta	1.22
C6EGF7	50S ribosomal protein L24	1.21
C6ELG9	6,7-dimethyl-8-ribityllumazine synthase	1.20
C6EAA7	Chaperone protein skp	1.15
C6EEX9	Triosephosphate isomerase	1.14
C6EJP6	Phosphoglucomutase	1.14
C6EAB5	Elongation factor Ts (EF-Ts)	1.11
C6EL63	Cysteine synthase	-1.09
C6EA12	50S ribosomal protein L25	-1.14
C6EGH1	DNA-directed RNA polymerase subunit alpha	-1.15
C6E9U5	Bifunctional polymyxin resistance protein ArnA	-1.16
C6EDH1	ADP-L-glycero-D-manno-heptose-6-epimerase	-1.18
C6EFL3	Peptidyl-prolyl cis-trans isomerase	-1.18
C6EF45	Conserved protein (Putative uncharacterized protein yihD)	-1.18
C6EAJ7	Methionine--tRNA ligase (Methionyl-tRNA synthetase)	-1.19
C6EJL4	Succinate dehydrogenase flavoprotein subunit	-1.19
C6EJY6	50S ribosomal protein L19	-1.19
C6EE31	DNA-directed RNA polymerase	-1.20
C6ED11	BasR transcriptional regulator	-1.22
C6EC17	L-serine deaminase I	-1.23
C6EI80	Pyruvate formate lyase activating enzyme 1	-1.23
C6EGE5	30S ribosomal protein S10	-1.23
C6EB40	Chaperone protein DnaK (HSP70)	-1.23
C6EGK4	AccC, subunit of biotin carboxylase and acetyl-CoA carboxylase	-1.25
C6EC50	Glyceraldehyde 3-phosphate dehydrogenase-A monomer	-1.25
C6EDU7	3-hydroxy acid dehydrogenase monomer	-1.25
C6ECH2	Adenylosuccinate synthetase	-1.25
C6EEY2	6-phosphofructokinase	-1.27
C6EAB3	Ribosome-recycling factor (RRF) (Ribosome-releasing factor)	-1.27
C6EH41	50S ribosomal protein L21	-1.28

Uniprot Accession Number	Protein ID	Fold change
C5W6A7	UPF0265 protein YeeX (UPF0265 protein B21_01896/ECBD_1650)	-1.28
C6EGF2	30S ribosomal protein S3	-1.30
C6EGT6	5-carboxymethyl-2-hydroxymuconate Delta-isomerase	-1.30
C5W723	NADH-quinone oxidoreductase subunit C/D	-1.30
C6EBW1	Aspartate--tRNA ligase	-1.30
C6ECX8	Elongation factor P (EF-P)	-1.32
C6EL62	Phosphohistidinoprotein-hexose phosphotransferase component of PTS system (Hpr)	-1.32
C6ECH0	Ribonuclease R (RNase R)	-1.35
C6EHF4	N-methyl-L-tryptophan oxidase	-1.35
C6EGG8	30S ribosomal protein S13	-1.35
C6EDH9	Glutaredoxin 3 (Reduced glutaredoxin 3)	-1.35
C6EH58	Ribosome-binding factor A	-1.35
C6EH03	Glutathione S-transferase domain protein (Stringent starvation protein A)	-1.37
C6EI52	Asparagine--tRNA ligase	-1.37
C6EF71	Uridine phosphorylase	-1.37
C6ELD4	Aspartate-semialdehyde dehydrogenase	-1.39
C6EF35	Coproporphyrinogen III dehydrogenase	-1.39
C6EIH1	Biosynthetic arginine decarboxylase	-1.39
C6EIX1	Cationic amino acid ABC transporter, periplasmic binding protein (GlnH, subunit of glutamine ABC transporter)	-1.41
C6EAC1	UPF0325 protein YaeH	-1.41
C6EH02	30S ribosomal protein S9	-1.41
C6ELG8	N utilization substance protein B homolog (Protein NusB)	-1.43
C6EAI8	Conserved protein (Putative uncharacterized protein yehS)	-1.43
C6EGF4	50S ribosomal protein L29	-1.43
C6EAT1	Carbonic anhydrase	-1.43
C6EH35	MlaC, subunit of phospholipid ABC transporter (Toluene tolerance family protein)	-1.45
C6EGC2	30S ribosomal protein S12	-1.45
C6EHM5	Phosphotyrosine-protein phosphatase	-1.47
C6EDU2	Uncharacterized protein YdfK	-1.52
C6EDR4	Diamine N-acetyltransferase (Spermidine N1-acetyltransferase)	-1.56
C6EAR5	Predicted subunit with GalU (Predicted subunit with GalU, subunit of glucose-1-phosphate uridylyltransferase)	-1.59
C6EI79	Formate acetyltransferase (Pyruvate formate lyase I)	-1.61
C6EK50	Conserved protein (Putative uncharacterized protein ybeL)	-1.64
C6EAV7	Protein translocase subunit SecA	-1.67
C6EG68	ATP synthase subunit delta	-1.67
C6EI98	Translation initiation factor IF-1	-1.67
C6EGK5	Acetyl-CoA carboxylase (Biotin carboxyl carrier protein)	-1.69
C6EFG0	Fe/S biogenesis protein NfuA	-1.69
C6EH24	Predicted ribosome-associated, sigma 54 modulation protein	-1.75
C6EAD0	Glutamate-1-semialdehyde 2,1-aminomutase	-1.82
C6EK62	UPF0250 protein YbeD	-1.82
C5WA02	Glycerol 3-phosphate dehydrogenase, aerobic	-1.82
C6ECW3	Oligoribonuclease	-1.89
C6EIE8	Probable Fe(2+)-trafficking protein	-1.92

Uniprot Accession Number	Protein ID	Fold change
C6EEY9	Superoxide dismutase	-1.92
C6EA47	Sigma factor-binding protein Crl	-1.96
C6ECD1	2',3'-cyclic-nucleotide 2'-phosphodiesterase	-1.96
C6EGD0	Bacterioferritin	-2.08
C6EG70	ATP synthase gamma chain	-2.08
C6EEW7	Pyruvate-flavodoxin oxidoreductase	-2.17
C6EBP2	Conserved inner membrane protein (Putative uncharacterized protein yjiN)	-2.17
C6EJL3	Succinate dehydrogenase and fumarate reductase iron-sulfur protein	-2.27
C6EG59	Aspartate--ammonia ligase	-2.44
C6EIU5	Glutathione S-transferase domain protein	-2.44
C6EHD7	Flagellar basal body FlaE domain protein	-3.13
C6EG50	Ribokinase	-3.57
C6EF32	NtrB (Sensory histidine kinase in two-component regulatory system with GlnG)	-3.70
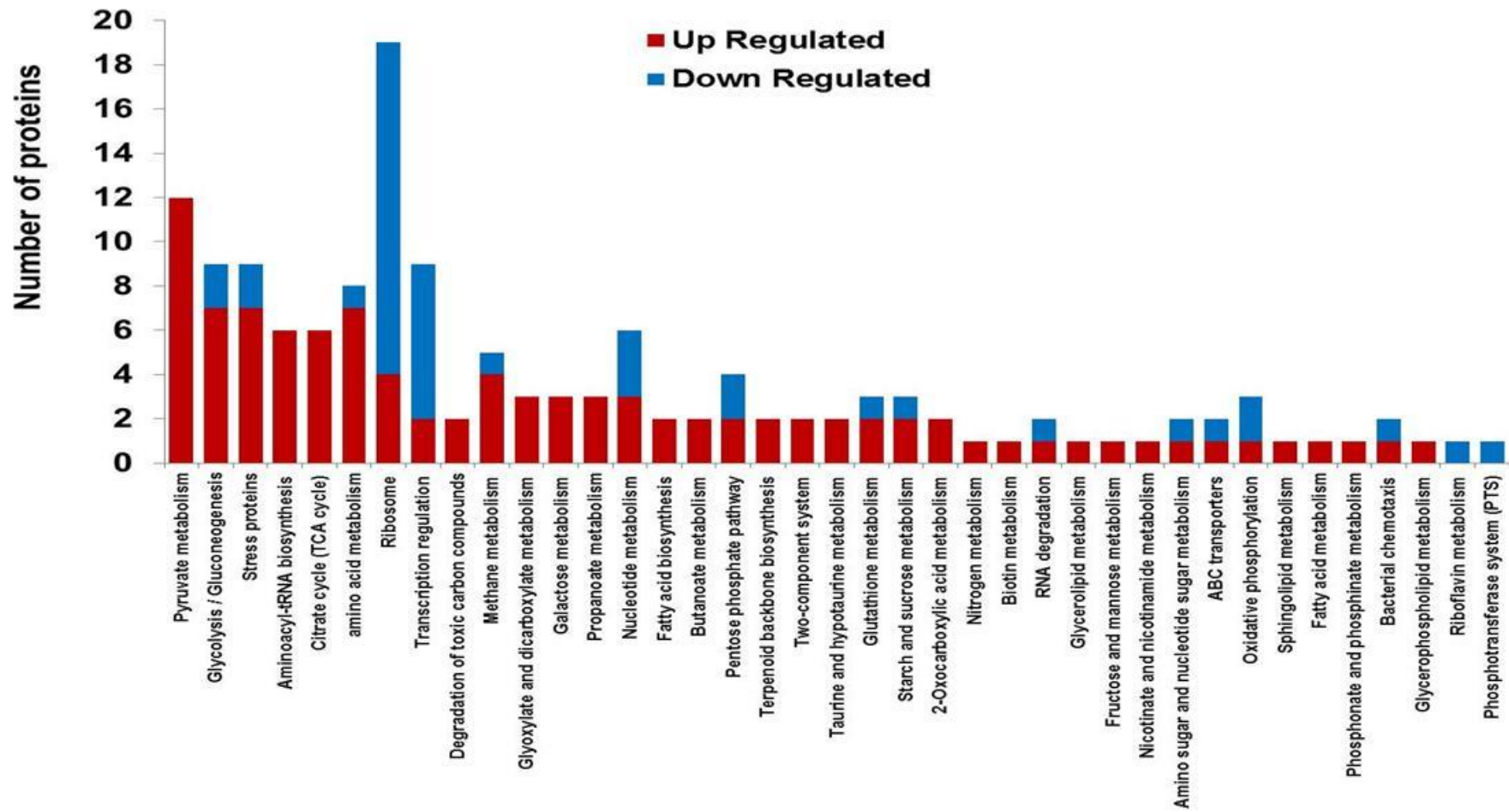
Figure 6.10 Functional annotation of regulated proteins between N and N+ was performed using KAAS. For annotation, the bi-directional best hit (BBH) method was used

6.3.3.1General stress response

The expression of the heterologous mevalonate pathway genes in *E. coli* induces and exacerbates different stressful conditions on the cell as seen in Table 6.5. There is an increase in the cold shock response as a result of growing cells at a temperature of 30°C even though the cold shock protein response is not expected at this temperature (Goldstein et al., 1990). The major cold shock proteins, cold shock protein CspG (5.29), cold shock protein CspA (1.91) are up regulated though cold shock protein CspC (-1.75) is down regulated. These proteins especially CspA, CspB and CspG are responsible for acting as chaperones to RNA polymerase and ensure efficient translation of proteins (Jiang et al., 1997; Jones et al., 1996; Nakashima et al., 1996). And in order to ensure proper folding of the proteins, there is an up-regulation of the peptidyl-prolyl cis-trans isomerase (4.94) and Chaperone SurA (Peptidyl-prolyl cis-trans isomerase SurA) (1.93)

Oxidative stress is induced by the normal metabolic processes of the cell that produce reactive oxygen species (ROS). The metabolic burden imposed by the heterologous proteins production further increases the oxidative stress as evidenced by the increase in abundance of superoxide dismutase (3.08),AhpF component, subunit of alkylhydroperoxide reductase (1.59) and AhpC component, subunit of alkylhydroperoxide reductase (1.36). These proteins which serve to protect DNA, lipids and other proteins from damage are under the regulation of the OxyR regulon (Bauer et al., 1999; Kim et al., 2002; Lynch, 1996). There is a down regulation of thioredoxin (-1.50),Integration host factor subunit alpha (IHF-alpha) (-2.33), Integration host factor subunit beta (IHF-beta) (-1.59), Thiol:disulfide interchange protein (DsbA) (-1.67) as well as the DNA protection during starvation protein (Dps) (-3.03). Dps is regulated by rpoS and IHF in the stationary phase and the OxyR and IHF in the exponential phase and serve to protect DNA from ROS (Bauer et al., 1999). The up-regulation of both RNA polymerase sigma factor (rpoE) (4.65) and Alpha,alpha-phosphotrehalase (1.82) would suggest that the is significant periplasmic stress as well as osmotic stress. Accumulation of trehalose which is an

osmo-protectant is an indication of high osmotic stress (Boos et al., 1987; Strom and Kaasen, 1993a). It has been noted that the osmo-protectiveness of trehalose is not due to externally supplied trehalose (Strom and Kaasen, 1993b). There is also an increase in abundance of proteins involved in xenobiotic metabolism as seen by the up regulation of 6,7-dihydropteridine reductase subunit of dihydropteridine reductase (2.12) and acetaldehyde dehydrogenase/alcohol dehydrogenase/PFL-deactivase (2.35). These proteins are responsible for degrading nitrotoluene, chloroalkanes/alkenes, xylene and benzoates which are toxic to the cell (Choi et al., 2008).

There is a down regulation of both NusA antitermination factor (Transcription elongation factor NusA) (-1.23) and DNA-directed RNA polymerase subunit beta (RNAP subunit beta) (-1.32) which are responsible for transcription of DNA and an up-regulation of the RNA degrading Polyribonucleotide nucleotidyltransferase (1.73), indicating a repression of protein synthesis. However, Purine nucleoside phosphorylase DeoD-type (PNP) (1.31) is up-regulated, showing likely active synthesis of nucleotide. HU, DNA-binding transcriptional regulator (-1.89) binds to curved DNA and represses transcription by reducing RNA polymerase access to the condensed DNA(Atlung and Ingmer, 2003; Dame, 2005a; Dame et al., 2006; Luijsterburg et al., 2006).

 There is up-regulation of the Lac repressor (LacI transcriptional repressor) (2.28) which represses the lactose operon in the phase of glucose abundance appears anomalous. There is concurrent up regulation the cAMP-receptor protein (CRP) (1.23), which would indicate catabolite activation and activation of the cAMP-CRP regulon.

Table 6.5 List of differentially regulated proteins involved in transcriptional control and stress response at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change
C6EHL8	Cold shock protein CspG	5.29
C6EGB7	Peptidyl-prolyl cis-trans isomerase	4.94
C6EK14	RNA polymerase sigma factor	4.65
C6EDY9	Extracellular solute-binding protein family 1 (MalE, subunit of maltose ABC transporter)	4.46
C6ECU7	Superoxide dismutase	3.08
C5W1L3	Lac repressor (LacI transcriptional repressor) (Lactose operon repressor)	2.28
C6EKC3	6,7-dihydropteridine reductase (Dihydropteridine reductase monomer, subunit of dihydropteridine reductase)	2.12
C6EB00	Chaperone SurA (Peptidyl-prolyl cis-trans isomerase SurA)	1.93
C6EEC4	Cold-shock DNA-binding domain protein (CspA transcriptional activator)	1.91
	Alpha,alpha-phosphotrehalase (Trehalose-6-P hydrolase)	1.82
C5W9B3	Polyribonucleotide nucleotidyltransferase	1.73
C5W0U2	AhpF component, subunit of alkylhydroperoxide reductase	1.59
C6EKI4	Peptidase B	1.44
C6EK89	AhpC component, subunit of alkylhydroperoxide reductase	1.36
C6EBI7	Purine nucleoside phosphorylase DeoD-type (PNP)	1.31
C6EH56	NusA antitermination factor (Transcription elongation factor NusA)	-1.23
C6EE32	DNA-directed RNA polymerase subunit beta (RNAP subunit beta)	-1.32
C6EL60	Crr, subunit of enzyme II (Glucose-specific PTS system enzyme IIA component)	-1.41
C6EG32	Thioredoxin	-1.50
C6EI70	Integration host factor subunit beta (IHF-beta)	-1.59
C5WBA2	Thiol:disulfide interchange protein DsbA	-1.67
C6EJT9	S-ribosylhomocysteine lyase (AI-2 synthesis protein)	-1.67
C6ED65	Single-stranded DNA-binding protein	-1.75
C6EC07	Cold-shock DNA-binding domain protein (Stress protein, member of the CspA family; predicted DNA-binding transcriptional regulator)	-1.75
C6EE18	HU, DNA-binding transcriptional regulator, alpha subunit	-1.89
C6ECP5	Integration host factor subunit alpha (IHF-alpha)	-2.33
C6EKZ0	Adenylate kinase (ATP-AMP transphosphorylase)	-2.38
C6EC51	Peptide methionine sulfoxide reductase MsrB	-2.5
C6EIX0	DNA protection during starvation protein	-3.03

6.3.2.2 Central carbon metabolism (CCM)

There is evidence of an increase in glycolytic activity, as seen in Table 6.6and 6.7 as a result of the mevalonate pathway. There is up regulation of pyruvate kinase (1.65), phosphoglycerate kinase (1.54) and phosphoglycerate mutase (1.92). Pyruvate kinase is responsible for the final irreversible step in the glycolytic pathway by converting PEP to pyruvate. Enolase (-2.00) is, however, down-regulated. The significance of this may be related to its participation in RNA degradosome complex. Also up regulated are fructose bisphosphate aldolase monomer (1.27) and triosephosphate isomerase (1.14). The down regulation of 6-posphofructokinase (1.27) is in keeping with repression of gluconeogenesis and up regulation of glycolysis.

Other carbon sources are seen to have been mobilized in the mevalonate mutant. Both MalE (subunit of maltose ABC transporter), which actively transports maltose and maltodextrin into the cell and malP (the maltodextrin phosphorylase) which releases glucose-1-phosphate from maltodextrin are up-regulated by more than 4 fold. Glucose-1-phosphate is converted to glucose-6-phosphate by phosphoglucomutase (1.14) before entering the glycolytic pathway. Other sources of hexose sugars for the glycolytic pathway that are up-regulated include gat A (2.68), gatB (4.33), gatD (1.81) and gatZ (1.97) the gat operon for galactitol utilization which generates G-3-P and DHAP from galactitol. manA (mannose -6-phosphate isomerase) (1.69) which converts mannose-6-phosphate to fructose-6-phosphate as well as treC (trehalose-6-phosphate) (1.82) which hydrolyses trehalose-6-phosphate to glucose-6-phosphate. Crr, (Glucose-specific PTS system enzyme IIA component) (-1.41) is, however, seen to be down regulated. This enzyme functions in the PEP: PTS system for import and phosphorylation of extracellular glucose and is known to negatively regulate rpoS a down regulation of this enzyme is therefore in keeping with the up regulation of the maltose transporter, as an up regulation will directly inhibit the maltose transporter. This would indicate exhaustion of available extracellular glucose and therefor activation of alternative carbon source utilization (Higgins, 1992).

There is also a general up-regulation of pyruvate metabolic enzymes as well as the TCA cycle enzymes.There is up-regulation of the three components of the pyruvate dehydrogenase enzyme complex (E1p component) (2.67), dihydrolipoamide acetyltransferase (2.04) and dihydrolipoyl dehydrogenase (1.81)] which in co-operation

179

converts pyruvate to acetyl CoA. A further source of pyruvate is provided by the up regulation of the malic enzymes [NAD-dependent malic enzyme (4.12) and malate dehydrogenase (Oxaloacetate-decarboxylating) (NADP(+)) (2.51)] which decarboxylates malate and oxaloacetate to pyruvate. Both these enzymes are positively regulated by acetyl-CoA and acetyl-P respectively(Bologna et al., 2007). The up-regulation of both acetate kinase (1.98) and phosphate acetyl transferase (1.70), as well as acetaldehyde dehydrogenase (2.35) suggests conversion of acetyl-CoA to acetate through acetyl-P and acetaldehyde respectively, which would indicate a flux of pyruvate and acetyl CoA (El-Mansi and Holms, 1989). Also, there is up-regulation of lactate dehydrogenase (1.84) forming lactate. A further evidence of increase in acetyl-CoA is seen in the up regulation of the irreversible acetyl-CoA carboxytransferase (1.37) which converts acetyl-CoA to malonyl-CoA for synthesis of fatty acids. In the TCA cycle, four enzymes are up-regulated including citrate synthase (2.47), Isocitrate dehydrogenase (1.55), and the dihydrolipoamide acetyltransferase subunit of the 2-oxoglutarate dehydrogenase complex (2.04) as well as malate dehydrogenase (2.30). The up regulation of these TCA cycle enzymes are in keeping with the increase of acetyl-CoA and pyruvate as previously noted.

 In the pentose phosphate pathway, four enzymes are regulated. While the Transketolase (1.96) and Transaldolase (1.76) are up regulated, the 6-phosphogluconate dehydrogenase (-1.39) and phosphopentomutase (-1.67) are down regulated. The down regulation of 6-phosphogluconate dehydrogenase suggests a diminution of the oxidative phase of the pentose phosphate pathway but an up regulation of the non-oxidative phase as seen by the up regulation of both the transketolase and transaldolase enzymes. These would further add to the pool of fructose-6-phosphate and glyceraldehyde-3-phosphate for glycolysis. The down regulation of phosphopentomutase would also suggest that the products of the pentose phosphate pathway are not channeled to other biosynthetic activities but to increase the flux of the glycolytic pathway towards pyruvate and acetyl-CoA. The increase in acetyl-CoA would not favor the oxidative phase of the pentose phosphate pathway, as acetyl-CoA is a known inhibitor of glucose-6-phosphate dehydrogenase.

There is up regulation of the inorganic pyrophosphatase (4.98), which is needed for biosynthetic activity. Inorganic pyrophosphatase hydrolyses pyrophosphate to phosphate with release of energy that can be utilized to drive forward many biosynthetic reactions. Also, there is up regulation of the alkylphosphonate utilization operon protein (1.81).

However, the up regulation of both the inorganic pyrophosphatase and the alkylphosphonate utilization operon protein may be due to the increased terpene synthetic activity, which generates increased load of inorganic pyrophosphate (Urbina et al., 1999).

Table 6.6 List of differentially regulated proteins involved in central carbon metabolism between N and N+ at 95% confidence level without multiple test correction

Uniprot accession #	ID	fold change
C6ED05	Alpha-galactosidase monomer, subunit of alpha-galactosidase	4.36
C6EAL7	Galactitol-specific enzyme IIB component of PTS (GatB, subunit of galactitol PTS permease)	4.33
C6EAL6	Galactitol-specific enzyme IIA component of PTS (GatA, subunit of galactitol PTS permease)	2.68
C6EIJ6	D-3-phosphoglycerate dehydrogenase	1.91
C6EAS3	3-methyl-2-oxobutanoate hydroxymethyltransferase	1.86
C6EHC1	3-oxo-acyl-[acyl-carrier-protein] reductase	1.77
C6EL61	Phosphoenolpyruvate-protein phosphotransferase (Phosphotransferase system, enzyme I)	1.54
C6EFU4	Enoyl-[acyl-carrier-protein] reductase [NADH]	1.52
C6EJL1	2-oxoglutarate dehydrogenase, E2 subunit, dihydrolipoamide succinyltransferase	1.47
C6EE53	Phosphoenolpyruvate carboxylase	1.40
C6EDZ9	Glucose-6-phosphate isomerase	1.32
C6EII4	Fructose bisphosphate aldolase monomer, subunit of fructose bisphosphate aldolase class II	1.27
C6EEX9	Triosephosphate isomerase	1.14
C6EJP6	Phosphoglucomutase	1.14
C6EJL4	Succinate dehydrogenase flavoprotein subunit	-1.19
C6EC50	Glyceraldehyde 3-phosphate dehydrogenase-A monomer,	-1.25
C6EEY2	6-phosphofructokinase	-1.27
C6EL62	Phosphohistidinoprotein-hexose phosphotransferase component of PTS system (Hpr)	-1.32
C6EAR5	Predicted subunit with GalU (Predicted subunit with GalU, subunit of glucose-1-phosphate uridylyltransferase)	-1.59
C6EI79	Formate acetyltransferase (Pyruvate formate-lyase (Inactive))	-1.61
C6EGK5	Acetyl-CoA carboxylase (Acetyl-CoA carboxylase, biotin carboxyl carrier protein)	-1.69
C5WA02	Glycerol 3-phosphate dehydrogenase, aerobic (sn-glycerol-3-phosphate dehydrogenase, aerobic, FAD/NAD(P)-binding)	-1.82
C6EJL3	Succinate dehydrogenase and fumarate reductase iron-sulfur protein (Succinate dehydrogenase iron-sulfur protein, subunit of succinate dehydrogenase)	-2.27
C6EG51	D-ribose transporter subunit (Periplasmic binding protein/LacI transcriptional regulator) (RbsB, subunit of ribose ABC transporter)	-5.00

Table 6.7 List of differentially regulated proteins involved in central carbon metabolism at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change
C6E9W2	Glycerophosphoryl diester phosphodiesterase (Glycerophosphoryl diester phosphodiesterase, periplasmic)	5.23
C6ECC2	Inorganic pyrophosphatase	4.98
C6EDY9	Extracellular solute-binding protein family 1 (MalE, subunit of maltose ABC transporter)	4.46
C6EEN8	NAD-dependent malic enzyme (NAD-ME)	4.12
C5W9Z4	Phosphorylase	4.11
C6EAU3	Pyruvate dehydrogenase E1 component	2.67
C6EKP9	Malate dehydrogenase (Oxaloacetate-decarboxylating) (NADP(+)), Phosphate acetyltransferase	2.51
C6EJL8	Citrate synthase	2.47
C6EGM8	Acetaldehyde dehydrogenase/alcohol dehydrogenase/PFL-deactivase	2.35
C6EGZ6	Malate dehydrogenase	2.30
C5W1L3	Lac repressor (LacI transcriptional repressor) (Lactose operon repressor)	2.28
C6ELD0	3-oxoacyl-(Acyl carrier protein) synthase (Beta-ketoacyl synthase)	2.11
C6EAU2	Dihydrolipoamide acetyltransferase	2.04
C6ELF7	Acetate kinase (Acetokinase)	1.98
C6EAL5	D-tagatose-1,6-bisphosphate aldolase subunit GatZ	1.97
C6EIH4	Transketolase (Transketolase 1, thiamin-binding)	1.96
C6EJI9	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	1.92
C6EAH8	D-lactate dehydrogenase (D-lactate dehydrogenase, FAD-binding, NADH independent)	1.84
C6ECA9	Alpha,alpha-phosphotrehalase (Trehalose-6-P hydrolase)	1.82
C6EAL9	Alcohol dehydrogenase GroES domain protein (Galactitol-1-phosphate dehydrogenase)	1.81
C6EAU1	Dihydrolipoyl dehydrogenase	1.81
C6EBG1	Transaldolase	1.76
C6ELF6	Phosphate acetyltransferase	1.70
C6EDN5	Mannose-6-phosphate isomerase	1.69
C6ECS9	Pyruvate kinase	1.65
C6EGV5	Isocitrate dehydrogenase [NADP]	1.55
C6EII3	Phosphoglycerate kinase	1.54
C6EAA0	Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha	1.37
C6EG71	ATP synthase subunit beta (ATP synthase F1 sector subunit beta)	-1.33
C6EB56	6-phosphogluconate dehydrogenase, decarboxylating	-1.39
C6EG69	ATP synthase subunit alpha (ATP synthase F1 sector subunit alpha)	-1.41
C6EL60	Crr, subunit of enzyme II [glc], trehalose PTS permease	-1.41
C6EBI8	Phosphopentomutase	-1.67
C6EJ93	Enolase (2-phospho-D-glycerate hydro-lyase) (2-phosphoglycerate dehydratase)	-2.00
C6EKZ0	Adenylate kinase (ATP-AMP transphosphorylase)	-2.38

6.3.3.3 Amino acid metabolism

The amino acid biosynthetic activity in N+ compared to N is seen in Table 6.8. There is up regulation of both 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase (1.40) and phospho-2-dehydro-3-deoxyheptonate aldolase

(2.47). 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase is involved in the biosynthesis of lysine and diaminopimelate which is a precursor for the synthesis of the cell wall peptidoglycan. Phospho-2-dehydro-3-deoxyheptonate aldolase is used in the shikimate pathway for the synthesis of the aromatic amino acids tryptophan, tyrosine and phenylalanine from PEP and Erythrose-4-phosphate.

The up-regulation of glycine dehydrogenase [decarboxylating] (2.86)suggests increase in the metabolism of glycine, threonine and serine. This enzyme is also involved in the lipoylprotein component of the dihydrolipoamide acetyltransferase of the pyruvate dehydrogenase complex. This is in keeping with the increase demand on this enzyme. Additionally, the glycine cleavage system H protein which functions in the glyoxylate metabolism is also up-regulated (1.41). There is also up-regulation of the aspartate ammonia-lyase (1.79), which catabolizes aspartate to Fumarate. Cysteine and methionine metabolism is down regulated as seen by the down regulation of S-ribosylhomocysteine lyase (AI-2 synthesis protein) (-1.67).

Table 6.8 List of differentially regulated proteins involved in amino acid metabolism at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change
C6EIK5	Glycine dehydrogenase [decarboxylating]	2.86
C6EJJ0	Phospho-2-dehydro-3-deoxyheptonate aldolase (3-deoxy-D-arabino-heptulosonate 7-phosphate synthase)	2.47
C6ECY6	Aspartate ammonia-lyase	1.79
C6EIK4	Glycine cleavage system H protein	1.41
C6EAB9	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase	1.40
C6EJT9	S-ribosylhomocysteine lyase (AI-2 synthesis protein) (Autoinducer-2 production protein LuxS)	-1.67
C6EJ54	CSD sulfur transfer protein (Putative uncharacterized protein ygdL) (UBA/THIF-type NAD/FAD binding protein)	-2.5

6.3.3.4 Protein biosynthesis

The complex process of protein biosynthesis shows a mixed picture, with the up-regulation of phenylalanine (2.07), serine (1.67) and Histidine (1.52) --tRNA ligases and down regulation of the translation initiation factor (-1.27) as well as several ribosomal proteins. In terms of protein synthesis, there is no significant increase in synthesis in N+ compared to N, as seen in Table 6.9.

Table 6.9 List of differentially regulated proteins involved in protein biosynthesis at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change
C6ECP3	Phenylalanine--tRNA ligase alpha subunit	2.07
C6EGF8	50S ribosomal protein L5	1.97
C6EGE6	50S ribosomal protein L3	1.77
C6EI89	Serine--tRNA ligase	1.67
C6ECP1	50S ribosomal protein L20	1.53
C6EKJ5	Histidine--tRNA ligase	1.52
C6EGG2	50S ribosomal protein L18	1.45
C6EIL8	Lysine--tRNA ligase	1.29
C6EE39	Elongation factor Tu (EF-Tu)	1.14
C6EE33	50S ribosomal protein L7/L12	-1.18
C6EH57	Translation initiation factor IF-2	-1.27
C6EGF5	30S ribosomal protein S17	-1.28
C6EGF3	50S ribosomal protein L16	-1.37
C6EGG4	50S ribosomal protein L30	-1.43
C6EGE8	50S ribosomal protein L23	-1.45
C6EGG5	50S ribosomal protein L15	-1.45
C6EGE7	50S ribosomal protein L4	-1.47
C6EGC3	30S ribosomal protein S7	-1.47
C6EDF4	50S ribosomal protein L28	-1.52

Uniprot Accession Number	Protein ID	Fold change
C6EGG3	30S ribosomal protein S5	-1.52
C6EGH2	50S ribosomal protein L17	-1.54
C6EJY3	30S ribosomal protein S16	-1.56
C6EHT0	30S ribosomal protein S21	-1.75
C6EGH0	30S ribosomal protein S4	-2.00

6.3.3.4 Cell envelope

There is evidence of active metabolism at the cell envelope as seen by the up-regulation of lipopolysaccharide core heptose(II)-phosphate phosphatase (2.07), which is an important enzyme in the synthesis of the outer membrane of *E. coli* (Table 6.10) This is complemented by the up-regulation of 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase (1.40) which, as stated in 6.3.3.3,is involved in the biosynthesis of cell wall peptidoglycan as well as the glucan biosynthesis G (2.12) which is an enzyme in the pathway for synthesis of osmoregulated periplasmic glucan (OPG). Synthesis of OPGs is stimulated in response to osmotic stress.

Molybdopterin guanine dinucleotide-containing S/N-oxide reductase is up regulated more than three fold. This protein in association with two other members of the *torCAD* operon is responsible for reducing trimethylamine N-oxide (TMAO) to trimethylamine which is used in anaerobic growth conditions as a component of the respiratory chain to generate energy especially on non-fermemntable carbon sources as glycerol (Méjean et al., 1994; Santini, 1998).

Table 6.10 List of differentially regulated proteins involved in cell wall metabolism at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change
C6EDY9	Extracellular solute-binding protein family 1 (MalE, subunit of maltose ABC transporter)	4.46
C6EHL0	Molybdopterin guanine dinucleotide-containing S/N-oxide reductase	3.05
C6EHG4	Glucans biosynthesis protein G	2.12
C6E9U8	Lipopolysaccharide core heptose(II)-phosphate phosphatase	2.07
C6EAB9	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase	1.40
C6EI25	OmpA domain protein transmembrane region-containing protein (Outer membrane protein 3a (II*;G;d))	-1.61
C6EG51	D-ribose transporter subunit (Periplasmic binding protein/LacI transcriptional regulator)	-5.00

Table 6.11 List of differentially regulated proteins involved in terpenoid biosynthesis at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change	Function
C6EIL9	Isopentenyl-diphosphate Delta-isomerase (IPP isomerase)	29.01	Terpenoid biosynthesis
C6EL48	Farnesyl diphosphate synthase/geranyl diphosphate synthase	21.24	Terpenoid biosynthesis

6.3.4 Investigation of *E. coli* cellular adaptation strategies to mevalonate, P450 and ferrodoxin-reductase expression

The enhanced terpenoid biosynthetic pathway comprising of terpene synthase, the cytochrome P450 and the ferrodoxin-reductase with mevalonate pathway was compared to the terpene synthase mutant using the 8-plex iTRAQ. A total of 69 proteins were differentially regulated, as seen in Table 6.12 and an additional 151 proteins without Bonferroni correction. There are fewer enzymes that are differentially regulated here, which is in keeping with the phenotypic clustering but unexpected in the light of the metabolic load. The terpene pathway enzymes isopentenyl-diphosphate Delta-isomerase (16.08) and farnesyl diphosphate synthase/geranyl diphosphate synthase (8.26) are both up regulated but are 1.8 fold and 2.6 fold less than N+ respectively. The lac repressor (LacI) (-1.27) and MalE, subunit of maltose ABC transporter (-1.22) are both down regulated at the 95%

confidence level without Bonferroni correction, but contrary to what was seen earlier. There is also an increase in translation activity, as seen with the up regulation of ribosomes and elongation factors.

Table 6.12 List of proteins differentially regulated between N and T at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change	Function
C6ECQ0	UPF0061 protein YdiU	18.57	Uncharacterised
C6V8W6	Putative uncharacterized protein	16.84	Uncharacterised
C6EIL9	Isopentenyl-diphosphate Delta-isomerase (IPP isomerase)	16.08	Terpenoid backbone biosynthesis
C6EHX3	Extracellular solute-binding protein family 5	14.98	Transporter
C6EL48	Farnesyl diphosphate synthase/geranyl diphosphate synthase	8.26	Terpenoid backbone biosynthesis
C6EHL8	Cold shock protein CspG	6.38	Stress
C6EKZ1	Chaperone protein HtpG (Heat shock protein HtpG)	4.59	Stress
C6ECC2	Inorganic pyrophosphatase	4.40	Oxidative phosphorylation
C6ED05	Alpha-galactosidase monomer, subunit of alpha-galactosidase(Alpha-galactosidase, NAD(P)-binding)	4.20	Galactose metabolism
C6EK14	RNA polymerase sigma factor	3.53	Transcription regulation
C6EGF8	50S ribosomal protein L5	3.04	Translation
C6EB40	Chaperone protein DnaK (HSP70)	2.85	stress
C6EGG2	50S ribosomal protein L18	2.62	Translation
C6EGA1	50S ribosomal protein L34	2.53	Translation
C6EGE6	50S ribosomal protein L3	2.45	Translation
C6ECU7	Superoxide dismutase	2.45	Stress
C6EE78	ATP-dependent protease ATPase subunit HslU	2.42	Stress
C6ECP1	50S ribosomal protein L20	2.34	Translation
C6EEC4	Cold-shock DNA-binding domain protein (CspA transcriptional activator) (Major cold shock protein)	2.27	Stress
C6E9U8	Lipopolysaccharide core heptose(II)-phosphate phosphatase	2.25	Outer membrane biosynthesis
C6ECY2	60 kDa chaperonin (GroEL protein)	2.23	Stress
C6ECY6	Aspartate ammonia-lyase	2.22	Amino acid metabolism
C6ED16	Alkylphosphonate utilization operon protein PhnA	2.17	Phosphonate/phosphinate metabolism
C6EEI7	Oligopeptidase A	2.17	Proteolysis/stress
C6EJX9	Protein GrpE (HSP-70 cofactor)	2.15	Stress
C6EH42	50S ribosomal protein L27	1.93	Translation
C6EJZ8	ATP-dependent chaperone ClpB (ClpB chaperone)	1.82	Stress
C6EBZ8	ProP effector	1.73	Transporter
C6EH63	ATP-dependent RNA helicase (DEAD/DEAH box helicase domain protein)	1.71	Stress/ RNA degradation

Uniprot Accession Number	Protein ID	Fold change	Function
C6EIL8	Lysine--tRNA ligase	1.70	Translation
C6EGG0	30S ribosomal protein S8	1.68	Translation
C6EFU4	Enoyl-[acyl-carrier-protein] reductase [NADH]	1.66	Fatty acid metabolism
C6EGF1	50S ribosomal protein L22	1.65	Translation
C6EAB6	30S ribosomal protein S2	1.63	Translation
C6ECB7	Fructose-1,6-bisphosphatase class 1 (FBPase class 1)	1.62	Glycolysis/gluconeogenesis
C6ECE6	50S ribosomal protein L9	1.59	Translation
C6EJT2	Protein RecA (Recombinase A)	1.59	Homologous recombination/DNA repair
C6EBG1	Transaldolase	1.53	Pentose phosphate pathway
C6EGF7	50S ribosomal protein L24	1.46	Translation
C6EC50	Glyceraldehyde 3-phosphate dehydrogenase-A monomer	1.44	Glycolysis/gluconeogenesis
C6EG30	Transcription termination factor Rho (ATP-dependent helicase Rho)	1.41	Transcription regulation
C6EGC4	Elongation factor G (EF-G)	1.41	Translation
C5W9B3	Polyribonucleotide nucleotidyltransferase	1.34	RNA degradation
C6EGG8	30S ribosomal protein S13	1.32	Translation
C6EJY6	50S ribosomal protein L19	1.29	Translation
C6EKX2	Chaperone and weak protein oxidoreductase	1.29	Stress
C6EGN0	Global DNA-binding transcriptional dual regulator H-NS	1.28	Transcription regulation
C6EGB7	Peptidyl-prolyl cis-trans isomerase	1.27	Stress/protein folding
C6EAB5	Elongation factor Ts (EF-Ts)	1.27	Translation
C6EGF5	30S ribosomal protein S17	1.26	Translation
C6EE39	Elongation factor Tu (EF-Tu)	1.18	Translation
C6EI70	Integration host factor subunit beta (IHF-beta)	0.84	Transcription regulation
C6EKZ0	Adenylate kinase (ATP-AMP transphosphorylase)	0.76	Cell growth/ nucleotide biosynthesis
C6EL60	Crr, subunit of enzyme II [glc], trehalose PTS permease	0.76	PTS
C6EB56	6-phosphogluconate dehydrogenase, decarboxylating	0.76	Pentose phosphate pathway
C6EJ93	Enolase (2-phospho-D-glycerate hydro-lyase)	0.74	Glycolysis/gluconeogenesis
C6ED65	Single-stranded DNA-binding protein	0.73	Homologous recombination/DNA repair
C6EAA7	Chaperone protein skp	0.70	Stress
C6EBX6	Glucose-6-phosphate 1-dehydrogenase	0.70	Pentose phosphate pathway
C6EG71	ATP synthase subunit beta	0.69	Oxidative phoshorylation
C6EF45	Conserved protein (Putative uncharacterized protein yihD)	0.68	Uncharacterised
C5WBA2	Thiol:disulfide interchange protein DsbA	0.66	Stress/protein folding

Uniprot Accession Number	Protein ID	Fold change	Function
C6EG32	Thioredoxin	0.61	Stress
C6EG69	ATP synthase subunit alpha (ATP synthase F1 sector subunit alpha)	0.57	Oxidative phoshorylation
C6EBI8	Phosphopentomutase	0.52	Pentose phosphate pathway
C6EIX1	Cationic amino acid ABC transporter, periplasmic binding protein (GlnH, subunit of glutamine ABC transporter)	0.52	Transporter (glutamine)
C6EIX0	DNA protection during starvation protein	0.45	stress
C6EG68	ATP synthase subunit delta (ATP synthase F(1) sector subunit delta)	0.43	Oxidative phoshorylation
C6EG70	ATP synthase gamma chain (ATP synthase F1 sector gamma subunit)	0.32	Oxidative phoshorylation

6.3.4.1 Central carbon metabolism (CCM)

Inthe central carbon metabolism of N compared to T, the direction of glycolysis/gluconeogenesis was favorable towards gluconeogenesis. As seen in Table 6.13, fructose-1,6-bisphosphatase (1.62) and glyceraldehyde 3-phosphate (1.44) are up regulated and enolase (-1.35) is down regulated. In addition to these, Crr, subunit of enzyme II [glc], PTS permease (-1.32) is down regulated. There is no evidence of any change in activity in the TCA cycle or pyruvate metabolism. In the pentose phosphate pathway, glucose-6-phosphate dehydrogenase (-1.43), 6-phosphogluconate dehydrogenase (-1.32) and phosphopentomutase (-1.92) are all down regulated,while transaldolase (1.53) is up regulated. This does not suggest any significant increase in activity of the pathway. There is, however, an up regulation of alpha-galactosidase, NAD(P)-binding (4.20) which hydrolyses glycerolipids and glycolipids. Enoyl-[acyl-carrier-protein] reductase (1.66) fatty acid biosynthesis is up regulated.

Table 6.13 List of differentially regulated proteins involved central carbon metabolism between N and T at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change
C6ED05	Alpha-galactosidase monomer, subunit of alpha-galactosidase	4.20
C6EFU4	Enoyl-[acyl-carrier-protein] reductase [NADH]	1.66
C6ECB7	Fructose-1,6-bisphosphatase class 1	1.62
C6EBG1	Transaldolase	1.53
C6EC50	Glyceraldehyde 3-phosphate dehydrogenase-A monomer	1.44
C6EL60	Crr, subunit of enzyme II [glc], trehalose PTS permease	-1.32
C6EB56	6-phosphogluconate dehydrogenase, decarboxylating	-1.32
C6EJ93	Enolase (2-phospho-D-glycerate hydro-lyase)	-1.35
C6EBX6	Glucose-6-phosphate 1-dehydrogenase	-1.43
C6EG71	ATP synthase subunit beta	-1.45
C6EG69	ATP synthase subunit alpha (ATP synthase F1 sector subunit alpha)	-1.75
C6EBI8	Phosphopentomutase	-1.92
C6EG68	ATP synthase subunit delta (ATP synthase F(1) sector subunit delta)	-2.33
C6EG70	ATP synthase gamma chain (ATP synthase F1 sector gamma subunit)	-3.12

6.3.4.2 Stress response

The stress response of T relative to N is expected to be significant considering the additional burden of the mevalonate pathway, cytochrome P450 and the electron transport protein ferrodoxin- reductase.As shown in Table 6.14, there is up regulation of both the heat shock and the cold shock proteins, in keeping with the relative increase in metabolic burden. The components of the shock proteins up regulated included Cold shock protein CspG (6.38), Cold-shock DNA-binding domain protein (CspA transcriptional activator) (Major cold shock protein) (2.27), Chaperone protein HtpG (Heat shock protein HtpG) (4.59), Protein GrpE (HSP-70 cofactor) (2.15) and Chaperone protein DnaK (HSP70) (2.85). Otherstress response proteins that are up regulated include 60 kDa chaperonin (GroEL protein) (2.23), ATP-dependent protease ATPase subunit HslU (2.42), ATP-dependent chaperone ClpB (ClpB chaperone) (1.82) and the chaperone and weak protein oxidoreductase (1.29). These heat shock proteins may not be responding to heat shock but to other denaturing stimuli (Gross, 1996). Oxidative stress is evidenced by the up regulation of Superoxide dismutase (2.45) even though there is no evidence of up regulation of the alkylhydroperoxide reductases, but there is down regulation of Thioredoxin (-

1.64). In addition to osmotic stress, there is an indication of perisplamic osmotic stress as seen by the up regulation of Extracellular solute-binding protein family 5 (Predicted transporter subunit) (14.98),ProP effector (1.73)   and the RNA polymerase sigma factor E (rpoE) (3.53)(Kunte et al., 1999). There is also evidence of stress induced transcriptional regulation, as seen with the up regulation of ATP-dependent RNA helicase (DEAD/DEAH box helicase domain protein) (1.71), Transcription termination factor Rho (ATP-dependent helicase Rho) (1.41) and Global DNA-binding transcriptional dual regulator H-NS (H-NS transcriptional dual regulator) (1.28). H-NS is a global transcriptional repressor that causes condensation and supercoiling of DNA (Atlung and Ingmer, 2003; Dame, 2005; Dame et al., 2006; Luijsterburg et al., 2006)

Table 6.14 List of differentially regulated proteins involved stress response and transcription regulation between N and T at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change
C6EHX3	Extracellular solute-binding protein family 5 (Predicted transporter subunit)	14.98
C6EHL8	Cold shock protein CspG	6.38
C6EKZ1	Chaperone protein HtpG (Heat shock protein HtpG)	4.59
C6EK14	RNA polymerase sigma factor	3.53
C6EB40	Chaperone protein DnaK (HSP70)	2.85
C6ECU7	Superoxide dismutase	2.45
C6EE78	ATP-dependent protease ATPase subunit HslU	2.42
C6EEC4	Cold-shock DNA-binding domain protein (CspA transcriptional activator)	2.27
C6E9U8	Lipopolysaccharide core heptose(II)-phosphate phosphatase	2.25
C6ECY2	60 kDa chaperonin (GroEL protein)	2.23
C6EEI7	Oligopeptidase A	2.17
C6EJX9	Protein GrpE (HSP-70 cofactor)	2.15
C6EJZ8	ATP-dependent chaperone ClpB (ClpB chaperone)	1.82
C6EBZ8	ProP effector	1.73
C6EH63	ATP-dependent RNA helicase (DEAD/DEAH box helicase domain protein)	1.71
C6EJT2	Protein RecA (Recombinase A)	1.59
C6EG30	Transcription termination factor Rho (ATP-dependent helicase Rho)	1.41
C5W9B3	Polyribonucleotide nucleotidyltransferase	1.34
C6EKX2	Chaperone and weak protein oxidoreductase	1.29
C6EGN0	Global DNA-binding transcriptional dual regulator H-NS	1.28
C6EGB7	Peptidyl-prolyl cis-trans isomerase	1.27
C6EI70	Integration host factor subunit beta (IHF-beta)	-1.19
C6EKZ0	Adenylate kinase (ATP-AMP transphosphorylase)	-1.32
C6ED65	Single-stranded DNA-binding protein	-1.37
C6EAA7	Chaperone protein skp	-1.43
C5WBA2	Thiol:disulfide interchange protein DsbA	-1.52
C6EG32	Thioredoxin	-1.64
C6EIX1	Cationic amino acid ABC transporter, periplasmic binding protein (GlnH, subunit of glutamine ABC transporter)	-1.92
C6EIX0	DNA protection during starvation protein	-2.22

6.3.5 Investigation of *E. coli* cellular adaptation strategies to cytochrome P450 and ferrodoxin-reductase

To investigate the effect of cytochrome P450 on the expression of terpene synthase in *E. coli,* phenotypes N and P were compared. As seen in Table 6.15, a total of 69 proteins from the iTRAQ experiment showed differences in abundance, with 44 proteins showing increased abundance and 25 with reduced abundance. These proteins were then grouped into functional groups and are discussed according to the major metabolic pathways that they represent.

Table 6.15 List of proteins differentially regulated between N and P at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change	Function
C6EHX3	Extracellular solute-binding protein family 5	25.43	Transporter
C6ECQ0	UPF0061 protein YdiU	22.42	Uncharacterised
C6V8W6	Putative uncharacterized protein	14.44	Uncharacterised
C6EHL8	Cold shock protein CspG	5.63	Stress
C6ED05	Alpha-galactosidase monomer,	4.55	Galactose metabolism
C6EKZ1	Chaperone protein HtpG (Heat shock protein HtpG)	4.25	Stress
C6ECC2	Inorganic pyrophosphatase	4.11	Oxidative phosphorylation
C6EGF8	50S ribosomal protein L5	3.10	Translation
C6EGA1	50S ribosomal protein L34	2.92	Translation
C6ECY6	Aspartate ammonia-lyase	2.70	Amino acid metabolism
C6ECU7	Superoxide dismutase	2.64	Stress
C6EK14	RNA polymerase sigma factor	2.57	Transcription regulation
C6EGG2	50S ribosomal protein L18	2.47	Translation
C6EB40	Chaperone protein DnaK (HSP70)	2.41	Stress
C6EE78	ATP-dependent protease ATPase subunit HslU (Heat shock protein HslU)	2.41	stress
C6EGE6	50S ribosomal protein L3	2.34	Translation
C6E9U8	Lipopolysaccharide core heptose (II)-phosphate phosphatase	2.31	Outer membrane biosynthesis
C6EH42	50S ribosomal protein L27	2.24	Translation
C6EJX9	Protein GrpE (HSP-70 cofactor)	2.19	stress
C6ECP1	50S ribosomal protein L20	2.13	Translation
C6ED16	Alkylphosphonate utilization operon protein PhnA	2.10	Phosphonate/phosphinate metabolism
C6ECY2	60 kDa chaperonin (GroEL protein) (Protein Cpn60)	2.08	Stress
C6EJZ8	ATP-dependent chaperone ClpB (ClpB chaperone)	2.07	Stress
C6EKC3	6,7-dihydropteridine reductase	2.06	Nitroaromatic degradation
C6EEI7	Oligopeptidase A	1.97	Proteolysis/stress
C6EHB1	Endoribonuclease L-PSP (Predicted L-PSP (MRNA) endoribonuclease)	1.95	Inhibitor of protein synthesis
C6ECB7	Fructose-1,6-bisphosphatase class 1 (FBPase class 1)	1.93	Glycolysis/gluconeogenesis
C6EFU4	Enoyl-[acyl-carrier-protein] reductase [NADH]	1.90	Fatty acid metabolism
C6EBG1	Transaldolase	1.88	Pentose phosphate pathway
C6EEC4	Cold-shock DNA-binding domain protein (CspA transcriptional activator) (Major cold shock protein)	1.81	stress
C6EC50	Glyceraldehyde 3-phosphate dehydrogenase-A monomer	1.81	Glycolysis/gluconeogenesis
C6EIL8	Lysine--tRNA ligase	1.71	Translation
C6ECE6	50S ribosomal protein L9	1.61	Translation

Uniprot Accession Number	Protein ID	Fold change	Function
C6EAL5	D-tagatose-1,6-bisphosphate aldolase subunit GatZ	1.59	Glycerol metabolism
C6EGF1	50S ribosomal protein L22	1.58	Translation
C6ECA5	Conserved protein, subunit of YjgF homotrimer (Endoribonuclease L-PSP) (Ketoacid-binding protein)	1.57	Inhibitor of protein synthesis
C6EH63	ATP-dependent RNA helicase (DEAD/DEAH box helicase domain protein)	1.56	Stress/ RNA degradation
C6EJT2	Protein RecA (Recombinase A)	1.52	Homologous recombination/DNA repair
C6ECZ5	Lysine--tRNA ligase	1.51	Translation
C6EAB6	30S ribosomal protein S2	1.50	Translation
C6EGC4	Elongation factor G (EF-G)	1.33	Translation
C6EAB5	Elongation factor Ts (EF-Ts)	1.32	Translation
C6EAU3	Pyruvate dehydrogenase E1 component	1.19	Pyruvate metabolism
C6EE39	Elongation factor Tu (EF-Tu)	1.15	Translation
C6EE18	HU, DNA-binding transcriptional regulator, alpha subunit (Histone family protein DNA-binding protein)	0.87	Transcription regulator
C6EJ93	Enolase (2-phospho-D-glycerate hydro-lyase) (2-phosphoglycerate dehydratase)	0.84	Glycolysis/gluconeogenesis
C6EE33	50S ribosomal protein L7/L12	0.83	Translation
C6EKZ0	Adenylate kinase (ATP-AMP transphosphorylase)	0.83	Cell growth/ nucleotide biosynthesis
C6EDF4	50S ribosomal protein L28	0.80	Translation
C6EC07	Cold-shock DNA-binding domain protein	0.79	Stress /transcription regulation
C6EB56	6-phosphogluconate dehydrogenase, decarboxylating	0.79	Pentose phosphate pathway
C6EGG4	50S ribosomal protein L30	0.79	Translation
C6ED65	Single-stranded DNA-binding protein	0.73	Homologous recombination/DNA repair
C6EAA7	Chaperone protein skp	0.72	stress
C6EG71	ATP synthase subunit beta	0.72	Oxidative phoshorylation
C6E9U5	Bifunctional polymyxin resistance protein ArnA	0.71	Outer membrane biosynthesis
C6EHM5	Phosphotyrosine-protein phosphatase (Protein tyrosine phosphatase)	0.70	Signal transduction/PTM
C6EG32	Thioredoxin	0.65	Stress
C6EBX6	Glucose-6-phosphate 1-dehydrogenase	0.64	Pentose phosphate pathway
C6EBI8	Phosphopentomutase	0.63	Pentose phosphate pathway
C6EG68	ATP synthase subunit delta	0.59	Oxidative phoshorylation
C6EG69	ATP synthase subunit alpha	0.58	Oxidative phoshorylation
C6EC51	Peptide methionine sulfoxide reductase MsrB	0.57	Protein repair/oxidative stress
C6EIX0	DNA protection during starvation protein	0.57	stress
C5WBA2	Thiol:disulfide interchange protein DsbA	0.57	Stress/protein folding
C5W1L3	Lac repressor (LacI transcriptional repressor) (Lactose operon repressor)	0.55	Transcription regulation

Uniprot Accession Number	Protein ID	Fold change	Function
C6EIX1	Cationic amino acid ABC transporter	0.50	Transporter (glutamine)
C6EJL4	Succinate dehydrogenase flavoprotein subunit	-2.04	TCA cycle
C6EG70	ATP synthase gamma chain subunit	0.37	Oxidative phoshorylation

6.3.4.1Central carbon metabolism

A total of twelve proteins involved in the central carbon metabolism were regulated as seen in Table 6.16; four of these belong to the pentose phosphate pathway, with 3 in the glycolysis/gluconeogenesis pathway. In the glycolytic/gluconeogenetic pathway, the up regulation of irreversible gluconeogenetic fructose-1,6-bisphosphatase (1.93) would strongly suggest that there is gluconeogenesis in P compared to N. The up regulation of the bidirectional glyceraldehyde 3-phosphate (1.81) and down regulation of enolase (-1.19) which is another bidirectional enzyme may tend to support this. Also, tagatose-1,6-bisphosphate aldolase (1.59) which converts tagatose-1,6-bisphosphateb to DHAP and G-3-P could be acting in concert with the gluconeogenesis. Also of significant note is the down regulation of the only TCA cycle (succinate dehydrogenase flavoprotein subunit (-2.04)) enzyme. The up regulation of pyruvate dehydrogenase (1.19) in the light of the above may not be significant. In the pentose phosphate pathway there is a down regulation of three out of the four regulated enzymes. Only transaldolase (1.88) was observed to be up-regulated, while glucose-6-phosphate dehydrogenase (-1.56) and 6-phosphogluconate dehydrogenase (-1.27) in the oxidative phase of the pathway were down regulated. This does not suggest any increase in activity in the pathway. Phosphopentomutase (-1.59) which functions in the biosynthetic pathway for purine synthesis was down regulated. There is however an up regulation of Alpha-galactosidase, NAD(P)-binding (4.55).

Table 6.16 List of differentially regulated proteins between N and P involved in central carbon metabolism at the 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change
C6ED05	Alpha-galactosidase monomer, subunit of alpha-galactosidase (Alpha-galactosidase, NAD(P)-binding)	4.55
C6ECB7	Fructose-1,6-bisphosphatase class 1 (FBPase class 1)	1.93
C6EFU4	Enoyl-[acyl-carrier-protein] reductase [NADH]	1.90
C6EBG1	Transaldolase	1.88
C6EC50	Glyceraldehyde 3-phosphate dehydrogenase-A monomer, subunit of glyceraldehyde 3-phosphate dehydrogenase-A complex	1.81
C6EAL5	D-tagatose-1,6-bisphosphate aldolase subunit GatZ	1.59
C6EAU3	Pyruvate dehydrogenase E1 component	1.19
C6EJ93	Enolase (2-phospho-D-glycerate hydro-lyase)	-1.19
C6EB56	6-phosphogluconate dehydrogenase, decarboxylating	-1.27
C6EBX6	Glucose-6-phosphate 1-dehydrogenase	-1.56
C6EBI8	Phosphopentomutase	-1.59
C6EJL4	Succinate dehydrogenase flavoprotein subunit	-2.04

6.3.4.2 Stress response

The stress response (Table 6.17) in this case is due to the expression of the cytochrome P450 and the ferrodoxin-reductase (in plasmids) which is the only difference between N and P. Compared to the stress response between N and N+, the cold shock response here is similar with cold shock protein CspG (5.63) and cold shock protein CspA (1.91) observed to be up regulated and cold shock protein CspC (-1.27) was down regulated. In addition, there are heat shock proteins that are up regulated here but not observed in N vs N+. These include Chaperone protein HtpG (Heat shock protein HtpG) (4.25), chaperone protein DnaK (HSP70) (2.41), ATP-dependent protease ATPase subunit HslU (2.41), Protein GrpE (HSP-70 cofactor) (2.19). Other chaperone proteins include 60 kDa chaperonin (GroEL protein) (2.08) and ATP-dependent chaperone ClpB (ClpB chaperone) (2.07). These additional stress response proteins are also seen in T and would therefore be due to the presence of P450 and the ferrodoxin –reductase. In terms of oxidative stress, superoxide dismutase (2.64) and thioredoxin (-1.54) are regulated in the same

direction while the alkylhydroperoxide reductases are not diferentially regulated. Thiol:disulfide interchange protein (DsbA) (-1.75) as well as the DNA protection during starvation protein (Dps) (-1.75) were down regulated. The periplasmic stress regulator RNA polymerase sigma factor (rpoE) (2.57)was up regulated. One major significant difference is seen in the down regulation of Lac repressor (LacI transcriptional repressor) (-1.82) which is similar to the trend seen in T.

Table 6.17 List of differentially regulated proteins involved in transcriptional control and stress response between N and P at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change
C6EHL8	Cold shock protein CspG (Cold-shock DNA-binding domain protein) (DNA-binding transcriptional regulator)	5.63
C6EKZ1	Chaperone protein HtpG (Heat shock protein HtpG) (High temperature protein G)	4.25
C6ECU7	Superoxide dismutase	2.64
C6EK14	RNA polymerase sigma factor	2.57
C6EB40	Chaperone protein DnaK (HSP70) (Heat shock 70 kDa protein)	2.41
C6EE78	ATP-dependent protease ATPase subunit HslU (Heat shock protein HslU)	2.41
C6EJX9	Protein GrpE (HSP-70 cofactor)	2.19
C6ECY2	60 kDa chaperonin (GroEL protein) (Protein Cpn60)	2.08
C6EJZ8	ATP-dependent chaperone ClpB (ClpB chaperone)	2.07
C6EEI7	Oligopeptidase A	1.97
C6EHB1	Endoribonuclease L-PSP (Predicted L-PSP (MRNA) endoribonuclease)	1.95
C6EEC4	Cold-shock DNA-binding domain protein (CspA transcriptional activator) (Major cold shock protein)	1.81
C6EH63	ATP-dependent RNA helicase (DEAD/DEAH box helicase domain protein) (DeaD, DEAD-box RNA helicase)	1.56
C6EJT2	Protein RecA (Recombinase A)	1.52
C6EE18	HU, DNA-binding transcriptional regulator, alpha subunit (Histone family protein DNA-binding protein) (Transcriptional dual regulator HU-alpha (HU-2), subunit of HU transcriptional dual regulator)	0.87
C6EC07	Cold-shock DNA-binding domain protein (Stress protein, member of the CspA family; predicted DNA-binding transcriptional regulator) (Stress protein, member of the CspA-family)	0.79
C6ED65	Single-stranded DNA-binding protein	0.73
C6EAA7	Chaperone protein skp	0.72
C6EG32	Thioredoxin	-1.54
C6EIX0	DNA protection during starvation protein	-1.75
C5WBA2	Thiol:disulfide interchange protein DsbA	-1.75
C5W1L3	Lac repressor (LacI transcriptional repressor) (Lactose operon repressor)	-1.82

6.3.3.4 Protein biosynthesis

There is general up regulation of protein biosynthetic machinery as observed by the up regulation of the ribosomal proteins and the translation elongation factors (Table 6.18). At the same time, there is an up regulation of Endoribonuclease L-PSP (Predicted L-PSP (MRNA) endoribonuclease) (1.95) Endoribonuclease L-PSP ((Ketoacid-binding protein (1.57) and ATP-dependent RNA helicase (DeaD, DEAD-box RNA helicase) (1.56) which are known to inhibit protein synthesis and to degrade RNA (Lüking et al., 1998; Toone et al., 1991).

Table 6.18 List of differentially regulated proteins involved in protein biosynthesis between N and P at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change
C6EGF8	50S ribosomal protein L5	3.10
C6EGA1	50S ribosomal protein L34	2.92
C6ECY6	Aspartate ammonia-lyase	2.70
C6EGG2	50S ribosomal protein L18	2.47
C6EGE6	50S ribosomal protein L3	2.34
C6EH42	50S ribosomal protein L27	2.24
C6ECP1	50S ribosomal protein L20	2.13
C6EHB1	Endoribonuclease L-PSP (Predicted L-PSP (MRNA) endoribonuclease)	1.95
C6EIL8	Lysine--tRNA ligase	1.71
C6ECE6	50S ribosomal protein L9	1.61
C6EGF1	50S ribosomal protein L22	1.58
C6ECA5	Conserved protein, subunit of YjgF homotrimer (Endoribonuclease L-PSP) (Ketoacid-binding protein)	1.57
C6EH63	ATP-dependent RNA helicase (DEAD/DEAH box helicase domain protein) (DeaD, DEAD-box RNA helicase)	1.56
C6ECZ5	Lysine--tRNA ligase	1.51
C6EAB6	30S ribosomal protein S2	1.50
C6EGC4	Elongation factor G (EF-G)	1.33
C6EAB5	Elongation factor Ts (EF-Ts)	1.32
C6EE39	Elongation factor Tu (EF-Tu)	1.15
C6EE33	50S ribosomal protein L7/L12	0.83
C6EDF4	50S ribosomal protein L28	0.80
C6EGG4	50S ribosomal protein L30	0.79

6.3.3.4 Cell envelope

There is some evidence of active metabolism at the cell envelope in terms of cell wall biosynthetic activities and membrane transport as seen in Table 6.19. The lipopolysaccharide core heptose(II)-phosphate phosphatase (2.31) was up regulated while the bifunctional polymyxin resistance protein (-1.41) which is responsible producing lipid A of the cell wall lipopolysaccharide backbone was down regulated. The extracellular solute-binding protein family 5 (25.43) was up regulated. The extracellular solute-binding protein family 5 is perisplasmic dipeptide binding and import protein that responds to osmotic shock (Olson et al., 1991; Tam and Saier, 1993). Also down regulated was the phosphotyrosine-protein phosphatase (-1.43) which is responsible for regulation of many cell processes such as growth and differentiation, cell cycle and transcription control amongst others through post-translational modification (Zhang et al., 1995).

Table 6.19 List of differentially regulated proteins involved in cell envelope metabolism between N and P at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change
C6EHX3	Extracellular solute-binding protein family 5 (Predicted transporter subunit) (Predicted transporter subunit: periplasmic-binding component of ABC superfamily)	25.43
C6E9U8	Lipopolysaccharide core heptose (II)-phosphate phosphatase	2.31
C6E9U5	Bifunctional polymyxin resistance protein ArnA	-1.41
C6EHM5	Phosphotyrosine-protein phosphatase (Protein tyrosine phosphatase)	-1.43
C6EIX1	Cationic amino acid ABC transporter, periplasmic binding protein (GlnH, subunit of glutamine ABC transporter)(Glutamine ABC transporter periplasmic protein)	-2.00

6.3.4 Investigation of *E. coli* cellular adaptation strategies between N+ and T

To study the effect of cytochrome P450 and ferrodoxin-reductase on a mevalonate and terpene synthase expressing *E. coli* mutant, N+ was compared to T. a direct

comparison between these two mutants would indicate metabolic effect of the P450 and the ferrodoxin-reductase on the system. It would also give more insight on the reason for the difference in terpene yield between the two. As seen in Table 6.20 a total of 127 proteins were differentially regulated at the 95% confidence with Bonferroni correction. Figure 6.10 shows the pathways that the differentially regulated proteins have been mapped into, which shows at a glance, that pathways involved in the central carbon metabolism (glycolysis/gluconeogenesis, pyruvate metabolism, the TCA cycle, and the pentose phosphate pathway) are mostly down regulated. Also down regulated are the enzymes involved in terpene synthesis. There was general up regulation of the ribosomes alongside the other translational enzymes such as elongation factor G (1.32), translation initiation factor IF-2 (1.53) and translation initiation factor IF-1 (1.72). There was also up regulation of both RNA polymerase and of RNA degradation enzymes. However, there was down regulation of aminoacyl-tRNA biosynthesis and enzymes of cell division.

6.3.4.1 Central carbon metabolism (CCM)

There is an overall down regulation of the central carbon metabolism. In the glycolysis/gluconeogenesis pathway,phosphoglycerate kinase (-1.41), phosphoglycerate mutase (-1.89) and phosphoglucomutase (-1.36) were down regulated while fructose-1,6-bisphosphatase (1.60), glyceraldehyde 3-phosphate dehydrogenase (1.80) and enolase (1.46) were up regulated. The up regulation of the unidirectional fructose-1,6-bisphosphatase shows evidence of gluconeogenesis in T compared to N+. In pyruvate metabolism, 7 enzymes were down regulated namely, pyruvate dehydrogenase E1 component (-2.46), dihydrolipoamide acetyltransferase (-2.24), dihydrolipoyl dehydrogenase (-1.95), acetaldehyde dehydrogenase/alcohol dehydrogenase/PFL-deactivase (-2.13), D-lactate dehydrogenase (-2.05), malate dehydrogenase (Oxaloacetate-decarboxylating) (-2.36) and phosphate acetyltransferase (-1.71) and one up regulated protein pyruvate formate-lyase (Inactive) (1.67). The TCA cycle is seen to be repressed as all the regulated enzymes were down regulated. The 6 TCA cycle enzymes down regulated include 2-oxoglutarate dehydrogenase, E2 subunit, dihydrolipoamide

succinyltransferase (-1.36), (succinyl-CoA ligase [ADP-forming] subunit alpha (-1.39), (succinyl-CoA ligase [ADP-forming] subunit beta (-1.41), Isocitrate dehydrogenase (-1.62), Citrate synthase (-2.28) and Malate dehydrogenase (-3.03). The general trend of metabolic repression of the central carbon metabolism is also seen in the pentose phosphate pathway with phosphoglucomutase (-1.36) and Glucose-6-phosphate 1-dehydrogenase (-1.48) down regulated. In addition, other sources of carbon utilisation also follow the same trend. D-tagatose-1,6-bisphosphate aldolase subunit GatZ (-1.63) and Galactitol-specific enzyme IIA component of PTS(GatA, subunit of galactitol PTS permease) (-2.24)involved in galactitol metabolism, MalE, subunit of maltose ABC transporter (-5.49) and maltose/maltodextrin phosphorylase, malP (-3.68) involved in import and metabolism of maltose/maltodextrin were also down regulated. Other down regulated carbon metabolic enzymes include trehalose-6-P hydrolase (-2.01) and Glycerophosphoryl diester phosphodiesterase (-4.21).

6.3.4.2 Stress response and transcription regulation

There is a mixed picture of stress response by T with up regulation of heat shock and cold shock proteins and chaperones while the oxidative and osmotic stress proteins are down regulated as seen by the down regulation of AhpF component, subunit of alkylhydroperoxide reductase (-1.40), superoxide dismutase (-1.26), RNA polymerase sigma factor (rpoE) (-1.32) and glucans biosynthesis protein G (-1.66). There was general up regulation of transcription regulators with HU, DNA-binding transcriptional regulator, alpha subunit (1.68), Global DNA-binding transcriptional dual regulator H-NS (1.43) and NusA antitermination factor (1.21) all up regulated.DNA-directed RNA polymerase (1.23) and DNA-directed RNA polymerase subunit beta (1.16)were also up regulated while Lac repressor (LacI transcriptional repressor) (-3.00) was down regulated.

Table 6.20 List of proteins differentially regulated between N+ and T at 95% confidence level with multiple test correction

Uniprot Accession Number	Protein ID	Fold change	Function
C6ECQ0	UPF0061 protein YdiU	20.50	Uncharacterised
C6V8W6	Putative uncharacterized protein	19.73	Uncharacterised
C6EHX3	Extracellular solute-binding protein family 5 (Predicted transporter subunit)	12.69	Transporter
C6EG51	D-ribose transporter subunit (Periplasmic binding protein/LacI transcriptional regulator	4.27	Transporter
C6EB40	Chaperone protein DnaK (HSP70)	3.52	Stress
C6EKZ1	Chaperone protein HtpG (Heat shock protein HtpG)	3.45	Stress
C6EE78	ATP-dependent protease ATPase subunit HslU	2.79	Stress
C6EFG0	Fe/S biogenesis protein NfuA	2.66	Stress/protein folding
C6EGH0	30S ribosomal protein S4	2.63	Translation
C6ECE7	30S ribosomal protein S18	2.62	Translation
C6EK62	UPF0250 protein YbeD	2.49	Uncharacterised
C6ECY2	60 kDa chaperonin (GroEL protein)	2.26	Stress
C6EJX9	Protein GrpE (HSP-70 cofactor)	2.21	Stress
C6EGF0	30S ribosomal protein S19)	2.13	Translation
C6ECE9	30S ribosomal protein S6	2.10	Translation
C6EAV7	Protein translocase subunit SecA	2.09	transporter
C6EH02	30S ribosomal protein S9	2.09	Translation
C6EGG0	30S ribosomal protein S8	1.99	Translation
C6EJZ8	ATP-dependent chaperone ClpB	1.96	Stress
C6EB30	30S ribosomal protein S20	1.95	Translation
C6EHT0	30S ribosomal protein S21	1.88	Translation
C6EGG9	30S ribosomal protein S11	1.86	Translation
C6EKZ0	Adenylate kinase	1.83	Cell growth/ nucleotide biosynthesis
C6EGG2	50S ribosomal protein L18	1.80	Translation
C6EH63	ATP-dependent RNA helicase (DEAD/DEAH box helicase domain protein)	1.80	Stress/ RNA degradation
C6EC50	Glyceraldehyde 3-phosphate dehydrogenase-A monomer	1.80	Glycolysis/gluconeogenesis
C6EGG8	30S ribosomal protein S13	1.79	Translation
C6EGK5	Acetyl-CoA carboxylase (Biotin carboxyl carrier protein)	1.74	Fatty acid biosynthesis
C6EJY3	30S ribosomal protein S16	1.73	Translation
C6EI98	Translation initiation factor IF-1	1.72	Translation
C6EI52	Asparagine--tRNA ligase	1.72	Translation
C6EGF2	30S ribosomal protein S3	1.71	Translation
C6EAB6	30S ribosomal protein S2	1.71	Translation
C6EE18	HU, DNA-binding transcriptional regulator, alpha subunit	1.68	Transcription regulation
C6EGH2	50S ribosomal protein L17	1.67	Translation

Uniprot Accession Number	Protein ID	Fold change	Function
C6EI79	Formate acetyltransferase(Pyruvate formate-lyase (Inactive))	1.67	Pyruvate metabolism
C6EH58	Ribosome-binding factor A	1.66	Translation
C6EGF5	30S ribosomal protein S17	1.61	Translation
C6ECB7	Fructose-1,6-bisphosphatase class 1	1.60	Glycolysis/gluconeogenesis
C6EGC3	30S ribosomal protein S7	1.59	Translation
C6EC07	Cold-shock DNA-binding domain protein (Stress protein, member of the CspA family; predicted DNA-binding transcriptional regulator)	1.58	Stress /transcription regulation
C6EGE7	50S ribosomal protein L4	1.58	Translation
C6EGF8	50S ribosomal protein L5	1.54	Translation
C6EJY6	50S ribosomal protein L19	1.54	Translation
C6EGF1	50S ribosomal protein L22	1.53	Translation
C6EH57	Translation initiation factor IF-2	1.53	Translation
C6EFW4	Predicted protein (Putative uncharacterized protein yciN)	1.53	Uncharacterised
C6ECP1	50S ribosomal protein L20	1.52	Translation
C6ELG8	N utilization substance protein B homolog (Protein NusB)	1.48	Transcription regulation
C6EEI7	Oligopeptidase A (EC 3.4.24.70)	1.47	Proteolysis/stress
C6EGG5	50S ribosomal protein L15	1.47	Translation
C6ECA5	Conserved protein, subunit of YjgFhomotrimer	1.47	Inhibitor of protein synthesis
C6EJ93	Enolase	1.46	Glycolysis/gluconeogenesis
C6EGG3	30S ribosomal protein S5	1.45	Translation
C6EGE8	50S ribosomal protein L23	1.44	Translation
C6EH01	50S ribosomal protein L13	1.44	Translation
C6EGE9	50S ribosomal protein L2	1.44	Translation
C6EGN0	Global DNA-binding transcriptional dual regulator H-NS	1.43	Transcription regulation
C6ECE6	50S ribosomal protein L9	1.43	Translation
C6EKX2	Chaperone and weak protein oxidoreductase	1.42	stress
C6EGK4	AccC, subunit of biotin carboxylase and acetyl-CoA carboxylase	1.41	Fatty acid biosynthesis
C6EDF4	50S ribosomal protein L28	1.40	Translation
C6EGE6	50S ribosomal protein L3	1.38	Translation
C6EGE5	30S ribosomal protein S10	1.36	Translation
C6EA12	50S ribosomal protein L25	1.33	Translation
C6EGC4	Elongation factor G (EF-G)	1.32	Translation
C6EGG4	50S ribosomal protein L30	1.30	Translation
C6EI71	30S ribosomal protein S1	1.30	Translation
C6EGG1	50S ribosomal protein L6	1.29	Translation
C6ECY6	Aspartate ammonia-lyase	1.24	Amino acid metabolism
C6EE31	DNA-directed RNA polymerase	1.23	Transcription regulation
C6EH56	NusA antitermination factor	1.21	Transcription regulation
C6EE33	50S ribosomal protein L7/L12	1.20	Translation

Uniprot Accession Number	Protein ID	Fold change	Function
C6EEC4	Cold-shock DNA-binding domain protein (CspA transcriptional activator)	1.19	Stress
C6EE32	DNA-directed RNA polymerase subunit beta	1.16	Transcription regulation
C6EL29	Trigger factor (TF)	1.11	stress
C6EAW0	Cell division protein FtsZ	-1.19	Cell division
C6ECU7	Superoxide dismutase	-1.26	stress
C6EBI7	Purine nucleoside phosphorylase DeoD-type	-1.26	Nucleotide metabolism
C6EA50	Aminoacyl-histidine dipeptidase (Peptidase D)	-1.27	proteolysis
C6EAA0	Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha	-1.27	Pyruvate metabolism
C5W9B3	Polyribonucleotide nucleotidyltransferase	-1.29	RNA degradation
C6EJP6	Phosphoglucomutase	-1.36	Pentose phosphate pathway
C6EJK9	Succinyl-CoA ligase [ADP-forming] subunit alpha	-1.39	TCA
C5W0U2	AhpF component, subunit of alkylhydroperoxide reductase	-1.40	Stress
C6EII3	Phosphoglycerate kinase	-1.41	Glycolysis/gluconeogenesis
C6EJL0	Succinyl-CoA ligase [ADP-forming] subunit beta	-1.41	TCA
C6ECJ2	Nitroreductase	-1.46	Nitroaromatic degradation
C6EBX6	Glucose-6-phosphate 1-dehydrogenase	-1.48	Pentose phosphate pathway
C6EJT4	Alanine--tRNA ligase	-1.50	Translation
C6EI75	Phosphoserine aminotransferase	-1.57	Amino acid biosynthesis
C6EAB9	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase (Tetrahydrodipicolinate N-succinyltransferase)	-1.57	Amino acid/peptidoglycan biosynthesis
C6EGV5	Isocitrate dehydrogenase [NADP]	-1.62	TCA
C6EAL5	D-tagatose-1,6-bisphosphate aldolase subunit GatZ	-1.63	Glycerol metabolism
C6EAA7	Chaperone protein skp	-1.65	Stress
C6EHG4	Glucans biosynthesis protein G	-1.66	OsmoregulatedPerisplasmicglucan biosynthesis
C6EGL0	Cell shape determining protein, MreB/Mrl family	-1.69	Peptidoglycan synthesis/cell wall
C6ELF6	Phosphate acetyltransferase	-1.71	Pyruvate metabolism

Uniprot Accession Number	Protein ID	Fold change	Function
C6EIK5	Glycine dehydrogenase [decarboxylating]	-1.71	Amino acid metabolism
C6EKJ5	Histidine--tRNA ligase	-1.74	Translation
C6EIK4	Glycine cleavage system H protein	-1.78	Amino acid metabolism
C6EIL9	Isopentenyl-diphosphate Delta-isomerase (IPP isomerase)	-1.80	C6EIL9
C6EJI9	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	-1.89	Glycolysis/gluconeogenesis
C6EAL9	Alcohol dehydrogenase GroES domain protein (Galactitol-1-phosphate dehydrogenase)	-1.93	Galactose metabolism
C6EAU1	Dihydrolipoyl dehydrogenase	-1.95	Pyruvate metabolism
C6ECP3	Phenylalanine--tRNA ligase alpha subunit	-1.99	Translation
C6ECA9	Alpha,alpha-phosphotrehalase (Trehalose-6-P hydrolase)	-2.01	Stress protection
C6ELD0	3-oxoacyl-(Acyl carrier protein) synthase	-2.01	Fatty acid biosynthesis
C6EAH8	D-lactate dehydrogenase (D-lactate dehydrogenase, FAD-binding, NADH independent)	-2.05	Pyruvate metabolism
C6EGM8	Acetaldehyde dehydrogenase / alcohol dehydrogenase / PFL-deactivase	-2.13	Pyruvate metabolism
C6EB00	Chaperone SurA (Peptidyl-prolyl cis-trans isomerase SurA)	-2.18	Stress
C6EAL6	Galactitol-specific enzyme IIA component of PTS	-2.24	PTS/ Galactose metabolism
C6EAU2	Dihydrolipoamide acetyltransferase	-2.24	Pyruvate metabolism
C6EJL8	Citrate synthase	-2.28	TCA
C6EJJ0	Phospho-2-dehydro-3-deoxyheptonate aldolase	-2.29	Amino acid biosynthesis/shikimate pathway
C6EKP9	Malate dehydrogenase (Oxaloacetate-decarboxylating) (NADP(+)	-2.36	Pyruvate metabolism
C6EAU3	Pyruvate dehydrogenase E1 component	-2.46	Pyruvate metabolism
C6EL48	Farnesyl diphosphate synthase / geranyl diphosphate synthase	-2.57	C6EL48
C5W1L3	Lac repressor (LacI transcriptional repressor) (Lactose operon repressor)	-3.00	Transcription regulation
C6EGZ6	Malate dehydrogenase	-3.03	TCA
C5W9Z4	Phosphorylase	-3.68	Starch and glycogen metabolism
C6EGB7	Peptidyl-prolyl cis-trans isomerase	-3.88	Stress/protein folding
C6E9W2	Glycerophosphoryl diester phosphodiesterase	-4.21	Glycerol metabolism

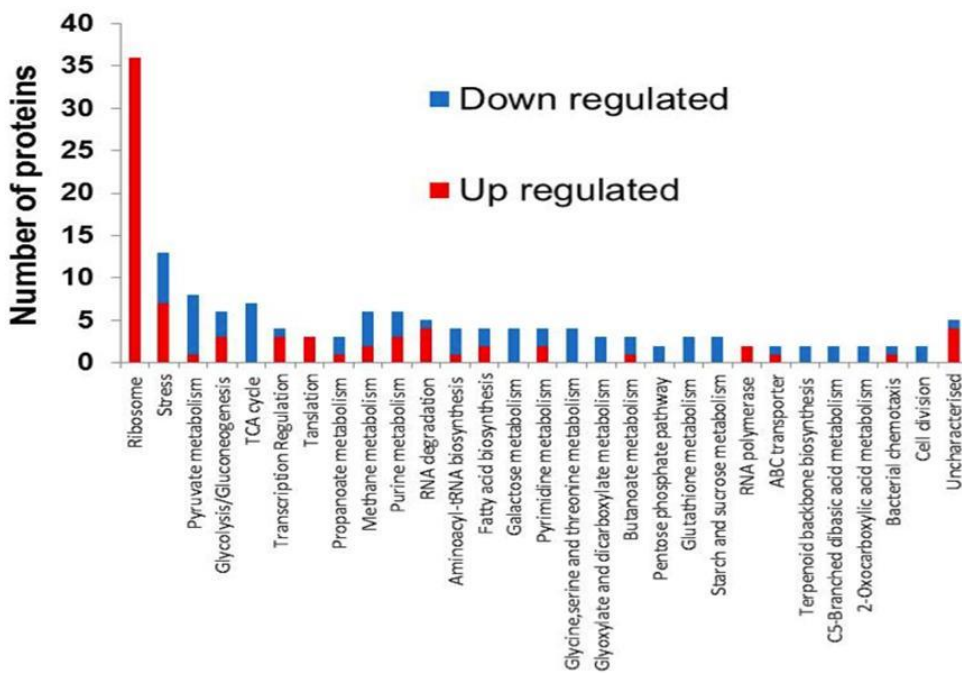| Uniprot Accession Number	Protein ID	Fold change	Function
C6EDY9	Extracellular solute-binding protein family 1 (MalE, subunit of maltose ABC transporter)	-5.49	Transporter
C6EL58	Predicted protein (Putative uncharacterized protein yfeK)	-8.75	Uncharacterised

Figure 6.11 Functional annotation of regulated proteins between N+ and Twas performed using KAAS. For annotation, the bi-directional best hit (BBH) method was used

6.4 Discussion

As a follow up to chapter 5, attempts were made to increase the number of identifiedproteinsby instituting more optimization steps such as digestion and fractionation methodology. Two-dimensional chromatographic separation reduces the complexity of peptides presented to the mass spectrometer for fragmentation

and identification at any given time with significant improvements in protein numbers and especially for low abundance proteins. An efficient offline first dimensional HPLC separation of peptides significantly enriches the peptide yield.Furthermore, as reported by Ow et al (2011b), up to 20% improvement in accuracy of iTRAQ results can be achieved by using high resolution fractionation methodologies. As tested in this chapter, the use of high resolution HILIC fractionation significantly increased the separation and resolution efficiencies and subsequent increased the number of identified and quantified proteins as seen in this chapter.

There was increase in the glycolysis, pyruvate metabolism and TCA cycle enzymes (Figure 6.12 and Figure 6.13) to yield acetyl-CoA for the synthesis of DMAPP and IPP. The increase in the glycolysis potentially leads to anincrease of pyruvate, which is partly converted to lactate by lactate dehydrogenase as well as to acetate by a combination of acetate kinase and phosphate acetyl transferase. The increase in the glycolytic pathway was from several sources: maltodextrin phosphorylase (malP) hydrolyzes maltodextrin to glucose-1-phosphate and enters the glycolytic pathway. The up-regulation of the maltodextrin active transport system (malE) ensures supply of maltohexose for the glycolytic pathway. Additionally, mannose-6-phosphate isomerase (manA) converts mannose-6-phosphate to F6P and tagatose-1,6-bisphosphate aldolase (gatZ) converts tagatose-1,6-bisphosphate to DHAP and G3P. Trehalose hydrolase up-regulation ensures the conversion of extracellular trehalose-6-phosphate to glucose-6-phosphate. Crr, the Glucose-specific PTS system enzyme IIA component was down regulated possibly by CRP mediated catabolite repression of glucose transport.

To get a global picture of the metabolic changes and to deduce the possible metabolic rational for the observed phenotypic differences, specific elements were examined.

To examine the elements responsible for the increased abundance of both IPP isomerase and FPP synthase which are necessary for terpene synthesis in the mutants, it was necessary to re-examine the iTRAQ conditions in which these proteins were observed to have increased in abundance (N+ and T).

Figure 6.12 Metabolic changes seen in glycolytic/gluconeogenetic pathway *E. coli* mevalonate mutant as compared to control (N) using iTRAQ based quantitation at 30 h post induction. Red and blue arrows show increased and decreased protein abundance respectively

In so doing, it was important to examine these (T vs N+) in the light of (N vs N+) and (N vs T) comparison Table 6.21 shows a list of regulated proteins that are commonly regulated in both N+ and T.

There was a comparative over expression of FPP synthase (2.57 fold) and IPP isomerase (1.8 fold) in N+ compared to that of T.Increased abundance of these proteins would require an up regulation of the terpene synthase immediately downstream to increase the turnover of the terpene Since both strains contain the FPP overproducing plasmid, but there was a significant difference in expression of FPP pathway enzymes, it is reasoned that the greater expression of these enzymes in N+ compared to T, would probably be due to over expression or under expression of protein that are unique to N+ or proteins that act to reduce expression in T. Since IPP isomerase and FPP synthase were up regulated in both conditions, the shared or common proteins between them would potentially represent the minimum set of proteins required for the successful expression of terpene synthase in an FPP expressing mutant under the same growth conditions. The proteins unique to each condition would hypothetically represent the reason for the differences in the level of expression of the terpene synthase (represented by the level of expression of both IPP isomerase and FPP synthase. It is important to bear in mind that the only genetic differences in this two strains is in the P450 and the ferrodoxin reductase in T but lacking in N+.

These proteins were almost all differentially regulated in the same direction except for the lac repressor protein LacI which was up regulated (2.28) in N+ and down regulated (-1.27) in T. The major metabolic functional group of these proteins includes stress response proteins, protein synthesis, transcriptional regulators and central carbon metabolism.
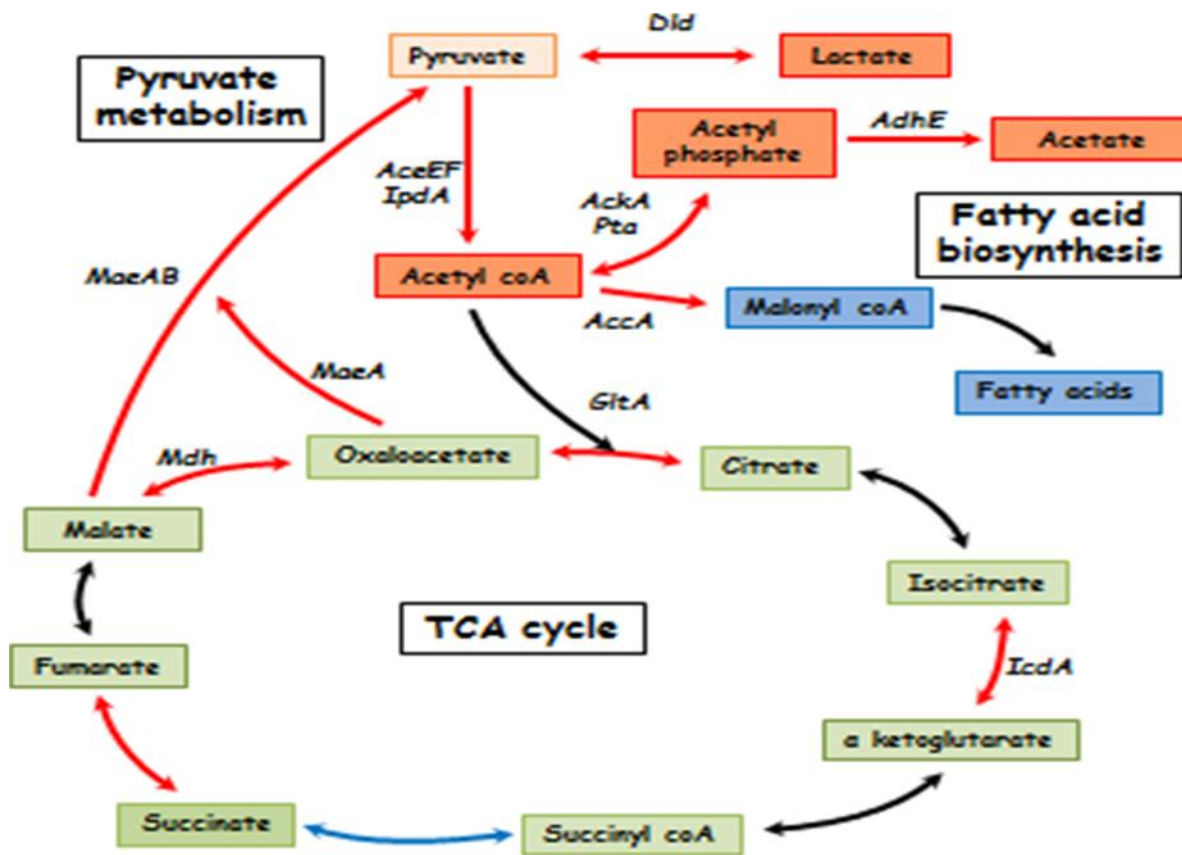
Figure 6.13 Metabolic changes seen in pyruvate metabolism and TCA cycle of *E. coli* mevalonate mutant as compared to control (N) using iTRAQ based quantitation at 30 h post induction. Red and blue arrows show increased and decreased protein abundance respectively

Table 6.21List of proteins that seen to be regulated in T and N+ compared to the control (N) at 95% confidence level with Bonferroni correction

Uniprot Accession Number	Protein ID	N vs N+	N vs T	*N* vs N + / N vs T
C6EDY9	Extracellular solute-binding protein family 1 (MalE)	4.46	0.82	5.44
C6EGB7	Peptidyl-prolyl cis-trans isomerase	4.94	1.27	3.89
C6EGZ6	Malate dehydrogenase	2.3	0.79	2.91
C5W1L3	Lac repressor (LacI transcriptional repressor)	2.28	0.79	2.89
C6EL48	Farnesyl diphosphate synthase / geranyl diphosphate synthase	21.24	8.26	2.57
C6EAU3	Pyruvate dehydrogenase E1 component	2.67	1.08	2.47
C6EEN8	NAD-dependent malic enzyme (NAD-ME)	4.12	1.73	2.38
C6EJL1	2-oxoglutarate dehydrogenase, E2 subunit	1.47	0.78	1.88
C6EIL9	Isopentenyl-diphosphate Delta-isomerase (IPP isomerase)	29.01	16.08	1.80
C6EGL0	Cell shape determining protein, MreB/Mrl family	1.38	0.88	1.57
C6EIH4	Transketolase (Transketolase 1, thiamin-binding)	1.96	1.25	1.57
C6EEX9	Triosephosphate isomerase	1.14	0.75	1.52
C6EAJ8	Antiporter inner membrane protein (Mrp protein)	2.24	1.51	1.48
C6EKC3	6,7-dihydropteridine reductase	2.12	1.44	1.47
C6EF32	NtrB (Sensory histidine kinase in two-component regulatory system with GlnG)	0.27	0.2	1.35
C6EIX1	Cationic amino acid ABC transporter	0.71	0.53	1.34
C6ECS9	Pyruvate kinase	1.65	1.24	1.33
C6EK14	RNA polymerase sigma factor	4.65	3.53	1.32
C6EEW7	Pyruvate-flavodoxin oxidoreductase	0.46	0.35	1.31
C6EJP6	Phosphoglucomutase	1.14	0.88	1.30
C5W9B3	Polyribonucleotide nucleotidyltransferase	1.73	1.34	1.29

Uniprot Accession Number	Protein ID	N vs N+	N vs T	N vs N + ――――― N vs T
C6ELF7	Acetate kinase	1.98	1.58	1.25
C6EBG1	Transaldolase	1.76	1.42	1.24
C6EJL4	Succinate dehydrogenase flavoprotein subunit	0.84	0.68	1.24
C6EI80	Pyruvate formate lyase activating enzyme 1	0.81	0.7	1.16
C6EBI8	Phosphopentomutase (Phosphodeoxyribomutase)	0.6	0.52	1.15
C6ECC2	Inorganic pyrophosphatase	4.98	4.4	1.13
C6EEY2	6-phosphofructokinase	0.79	0.73	1.08
C6EL62	Phosphohistidinoprotein-hexose phosphotransferase component of PTS system (Hpr)	0.76	0.74	1.03
C6ED05	Alpha-galactosidase monomer	4.36	4.34	1.00
C6EE39	Elongation factor Tu (EF-Tu)	1.14	1.18	-1.04
C6EL60	Crr, subunit of enzyme II [glc] (Glucose-specific PTS)	0.71	0.75	-1.06
C6EB56	6-phosphogluconate dehydrogenase, decarboxylating	0.72	0.77	-1.07
C6E9U8	Lipopolysaccharide core heptose (II)-phosphate phosphatase	2.07	2.25	-1.09
C6EG30	Transcription termination factor Rho	1.27	1.42	-1.12
C6EE32	DNA-directed RNA polymerase subunit beta	0.76	0.88	-1.16
C6EEC4	Cold-shock DNA-binding domain protein (CspA transcriptional activator)	1.91	2.27	-1.19
C6EBZ8	ProP effector	1.45	1.73	-1.19
C6ED16	Alkylphosphonate utilization operon protein PhnA	1.81	2.17	-1.20
C6EA47	Sigma factor-binding protein Crl	0.51	0.63	-1.24
C6ECY6	Aspartate ammonia-lyase	1.79	2.22	-1.24
C6ED65	Single-stranded DNA-binding protein	0.57	0.73	-1.28
C6EJT2	Protein RecA (Recombinase A)	1.24	1.59	-1.28

Uniprot Accession Number	Protein ID	N vs N+	N vs T	$\dfrac{N \text{ vs } N+}{N \text{ vs } T}$
C6EI70	Integration host factor subunit beta (IHF-beta)	0.63	0.84	-1.33
C6EIX0	DNA protection during starvation protein	0.33	0.45	-1.36
C6EJL3	Succinate dehydrogenase and fumarate reductase	0.44	0.6	-1.36
C6EJ93	Enolase	0.5	0.74	-1.48
C6EJV6	DNA binding protein, nucleoid-associated (H-NS-like)	2.12	3.14	-1.48
C6EH57	Translation initiation factor IF-2	0.79	1.22	-1.54
C6EE18	HU, DNA-binding transcriptional regulator, alpha subunit	0.53	0.92	-1.74
C6EC50	Glyceraldehyde 3-phosphate dehydrogenase-A monomer	0.8	1.39	-1.74
C6EKZ0	Adenylate kinase (AK)	0.42	0.76	-1.81
C6ECP5	Integration host factor subunit alpha (IHF-alpha)	0.43	0.78	-1.81

The peptidyl-prolyl cis/trans isomerase (FkpA) chaperone was significantly more (3.89 fold) up regulated in N+ than in T. Two other peptidyl-prolyl cis/trans isomerases (SurA and Skp) were observed to show similar trend. These proteins act as chaperones to ensure proper folding of proteins. Their activity help to reduce the RpoE induced stress response to misfolded proteins and reduce formation of protein aggregates (Liu and Walsh, 1990; Ruiz-Perez et al., 2010). However, these proteins were observed to be down regulated in the P mutant when compared to N. Proteins involved in transcriptional regulation by causing folding and compaction of DNA and in so doing cause a global or specific reduction in transcription include integration host factors (alpha and beta), H-NS, HU (Dame, 2005b). These proteins even though differentially regulated in the same direction but were all observed to be down regulated inN+.compared to T. Similarly, the DNA protection during starvation protein (Dps) which binds to and protects DNA from toxic substances during starvation or stationary phaseand the single-stranded DNA-binding proteinwere observed to be relatively less in abundance in N+ compared to T. Taken together, this could mean that T is more starved with more reduction in the rate of transcription.

In central carbon metabolism, several proteins were observed to be relatively more up regulated in N+ compared to T. These include malate dehydrogenase, pyruvate dehydrogenase E1 component, NAD-dependent malic enzyme, 2-oxoglutarate dehydrogenase, E2 subunit, transketolase, triosephosphate isomerase, extracellular solute-binding protein family 1 (MalE), pyruvate kinase, transaldolase, acetate kinase and phosphoglucomutase. This would indicate more or better energy source utilization in N+ strain.

Lac repressor (LacI transcriptional repressor) levels are seen to be up regulated in N+ (2.28) and down regulated in T (-1.27). The *lac* repressor serves to regulate the *lac* operon and is involved in the efficient and selective use of carbon sources by *E. coli* and as such LacI will ensure the preferential use of glucose if available by binding to *lacO* and preventing the transcription of the *lac* operon. An up regulation of LacI will therefore signify availability and dependency of N+ on glucose and the repression of the utilization of other carbon sources such as glycerol, other hexoses/pentoses and polysaccharides such as maltodextrin. This process is usually achieved by the cAMP-CRP mediated catabolite repression. As glucose levels fall, cAMP is up regulated and this subsequently binds and activates CRP, and in association induces *lac* operon. CRP is up regulated in N+ in the presence of up regulated LacI. It is known that the presence of LacI dictates the preference of glucose over lactose and glycerol in that order but increasing the level of LacI would change the preference to glucose/glycerol/lactose (Blommel et al., 2007). However, elevated levels of CRP would favor the utilization of sources other than glucoseand the maltose/maltodextrin utilization operon is induced by cAMP-CRP(Boos and Shuman, 1998; Dippel and Boos, 2005). The glucose specific PTS system crr is known to negatively regulate cAMP-CRP levels (Boos and Shuman, 1998) and so a down regulation of crr would support increase in cAMP-CRP levels. This would therefore be responsible for the up regulation of the maltose/maltodextrin utilization operon and glycerol and other carbon sources such as mannose, trehalose, and galactitol in order to meet the increase demand for acetyl-CoA. The enhance level of cAMP-CRP is responsible for increased utilization of carbon sources through the activation of the *rpoS* regulon. Some of the enzymes in this regulon up regulated include, Acetate kinase (Acka), Phosphate

acetyltransferase (pta), Aldehyde Dehydrogenase (AldA), ArgT, subunit of lysine/arginine/ornithine ABC Transporter (Cationic amino acid ABC transporter, periplasmic binding protein), ATP synthase subunit alpha (AtpA), dipeptide transporter (DppA), D-tagatose-1,6-bisphosphate aldolase subunit GatY (gatY), glycerol-3-phosphate transporter (UgpB), Maltose-binding periplasmic protein (MalE), D-galactose-binding periplasmic protein (MglB), Extracellular solute-binding protein family 1(Han and Lee, 2006; Xu and Johnson, 1995). These proteins are potentially responsible for the significant more expression of the terpene synthase precursors in N.

The elements that are responsible for the low abundance of both IPP isomerase and FPP synthase and hence low terpene production are examined next. The P and T strains showed relatively low abundance of these proteins when compared to N and N+ respectively. These elements would be due to the presence of P450 and ferrodoxin reductase in both P and T. The most prominent common theme in these comparisons that were not observed in the N+ strain was the presence of the heat shock proteins which were all observed to be up regulated, while there was a relative down regulation of the cold shock proteins when compared to N+. of significance is the observation of two uncharacterized proteins that are significantly upregulated in these two strains, UPF0061 protein YdiU with 18.57 fold in T and 22.42 fold increase in P compared to N and putative uncharacterized protein (C6V8W6) with 16.84 and 14.44 fold increase in T and P respectively.

6.5 Conclusions

This chapter, the metabolic profile of a high yield terpene producing *E. coli* strain has been examined and compared to the baseline terpene strain. The cytochrome P450/ferrodoxin reductase strains with or without the FPP-overproducing pathway have also been examined and comparisons made.

There is a proteomic evidence of an increase in glycolysis and pyruvate metabolism to yield acetyl-CoA for the synthesis of DMAPP and IPP in the N+ strain. Thelikely

215

increase in the glycolysis leads to a flux of pyruvate which was partly converted to lactate by lactate dehydrogenase as well as to acetate by a combination of acetate kinase and phosphate acetyl transferase. The increase in the glycolytic pathway is from several sources: maltodextrin phosphorylase (malP) hydrolysis maltodextrin to glucose-1-phosphate and enters the glycolytic pathway. The up-regulation of the maltodextrin active transport system (malE) ensures supply of maltohexose for the glycolytic pathway. Additionally, mannose-6-phosphate isomerase (manA) converts mannose-6-phosphate to F6P and tagatose-1,6-bisphosphate aldolase (gatZ) converts tagatose-1,6-bisphosphate to DHAP and G3P. Trehalose hydrolase up-regulation ensures the conversion of extracellular trehalose-6-phosphate to glucose-6-phosphate. Crr, the Glucose-specific PTS system enzyme IIA component is down regulated possibly by CRP mediated catabolite repression of glucose transport. The presence of cytochrome P450 is thought to have led to the drastic reduction of the terpene synthetic pathway.enzymes. The over abundance of heat shock proteins, reduction of cold shock proteins and the Peptidyl-prolyl cis-trans isomerases are an indication of protein aggregation and the CCM is depressed as a metabolic consequence. This would be the probable cause of reduce terpene synthase and hence terpene in P and T.

In the next chapter, these findings would be put into perspective with the rest of the results of this thesis.

CHAPTER7

Conclusions and future work

7.1 Introduction

The aim of this thesis was to understand the metabolism of an *E. coli* strain engineered to produce a sesquiterpene from *Nostoc sp.*PCC7120 and the effect of other additional genes such as cytochrome P450 and a mevalonate pathway on sesquiterpene production. To achieve this goal, molecular biological techniques as well as bio-analytical and chemical analytical tools were employed. The major findings of this thesis are summarised in this chapter and the outlook for future work is also discussed.

7.2 Quantification of terpene

In order to improve any system, it is important to first ensure that the system is established and functionally tested. Once the productive efficiency of this system is tested subsequent refinement and improvement is measured. The sesquiterpene from *Nostoc sp.* PCC7120 was functionally expressed in *E. coli* and the terpene product confirmed to be the anticancer compound β-elemene using GC-MS. Using the endogenous DXP pathway to supply IPP/DMAPP precursors, this single plasmid carrying strain produced a baseline value of 38 mg/L of the terpene.Qualitative assay of this metabolite produced from an *E. coli* host have been performed, but there is no known report of a quantitative assay. In view of adding an FPP-overproducing plasmid with or without the cytochrome P450/ferrodoxin reductase plasmids to this system, production was tested with empty versions of these two plasmids. The product titres from the three plamid system showed more than a 7 fold reduction in the baseline titres of β-elemene from the single plasmid strain. This significant reduction is ascribed to the extra metabolic burden imposed by adding two additional plasmids. Since the IPP/DMAPP precursors are essential for cell wall metabolism, the shunting of these to terpene synthetic pathway had a negative effect on cell growth and survival. The addition of a codon optimized FPP- overproducing plasmid increased the terpene titres by more than 34 fold to 178 mg/L and the specific productivity per gram of cells increased by more than 40 fold. As a first pass yield without any optimisation, this

result compares well with other studies using same or similar FPP-overproducing system with a sesquiterpene synthase. Some notable examples include 500 mg/L of amorphadiene (Redding-Johanson et al., 2011a), 900 mg/L of bisabolene (Peralta-Yahya et al., 2011a), 136 mg/L of farnesol (Wang et al., 2010) and 380 mg/L of α-farnesene (Wang et al., 2011). Using a codon optimised terpene synthase as employed in these studies, would possibly give yields within the same range. The cytochrome P450 was very poorly expressed but nonetheless, its effect on the terpene production was negative and profound, reducing the terpene output by almost 90 fold.

7.3 Metabolism

There is always some metabolic consequences accompanying the production of any foreign protein in *E. coli* which is evident in this thesis. The metabolic changes are usuallystress induced and the effect of the stress is the observed phenotypic changes. The stress in this case comes from the terpene synthase and the plasmid, and the effect is manifested in the rate of growth and changes in the relative abundance of proteins in terpene synthase carrying straincompared to the empty plasmid carrying *E. coli* strain. The growth retardation observed (as measured from OD values) in the terpene producing strain (N) could be attributed to reduction in the synthesis of the cell wall peptidoglycan. This is because the IPP/DMAPP precursors that are used for the synthesis of undecaprenyl pyrophosphate of lipid II are hijacked to make terpene by the terpene synthase to the detriment of the cell. The relative decrease in abundance of Tetrahydrodipicolinate N-succinyltransferase and Glutamine-fructose-6-phosphate aminotransferase are in support of this assertion as shown by the growth retardation. The apparent demand for both pyruvate and glyceraldehyde-3-phophate needed for IPP/DMAPP synthesis is likely to have triggered an increase channeling of resources by the organism to meet this demand. This adaptation strategy is likely to be through a combination of reduction of cell wall synthesis, catabolic degradation of the cell wall to release more IPP/DMAPP through colicin M as well as increase catabolism of tryptophan, cysteine and serine by tryptophanase to yield pyruvate. The increase in catabolism of tryptophan also produces the toxic indole which may have contributed to growth

inhibition. Even though the central carbon metabolism seemed to have been activated to provide pyruvate and G3P, mobilization of all possible carbon sources available to the cell in this case did not seem to have been utilized when compared to the N+ strain.

The addition of an exogenous precursor pathway appeared to have relieved the catabolism of the cell wall peptidoglycan, increased the efficiency of the carbon source utilization with up regulation of the central carbon metabolism, which led to increased titres of terpene.The effect of the FPP-overproducing plasmid led to a significant increase in the mobilization of several carbon sources with an increase in the activity of the glycolytic pathway, the pyruvate metabolism as well as the citric acid cycle which potentially increased the flux of acetyl-CoA needed for the synthesis of IPP/DMAPP sufficiently to meet the demand for both cell biosynthesis and as well as synthesis of the terpene. This increase in the CCM enzymes possibly caused an over-abundance of acetyl-CoA which was then shunted to produce acetic acid and lactic acid which are themselves not growth friendly (Zhou et al., 2003). So, while, the FPP-overproducing strain relieved the catabolic effect on the cell wall, it had its own negative consequence by inducing more significant acid stress. There is also the possibility that there was an additional membrane stress brought about by excessive production of HMG-CoA as seen by the increase in malonyl-CoA and other fatty acid biosynthetic enzymes. HMG-CoA accumulation is known to be toxic to *E. coli* cells as reported by Kizer et al. (2008). These (acidosis and HMG-CoA) would be the probable explanations for the growth retardation seen in N+ compared to P.

The FPP overproducing pathway theoretically increases the efficiency of carbon usage and mobilization through up regulation of cAMP-CRPincluding utilization of glycerol. The relative abundance of level of cAMP-CRP is thus likely responsible for the increased utilization of carbon sources through the activation of the rpoS regulon. These proteins are potentially responsible for the significant over expression of the terpene synthase precursors in N+(D. S.-W. Ow et al., 2009).The carbon scavenging seen in N+ which is not seen in N is attributable to growth on glycerol, as growth on a low carbon source such as glycerol is also known to induce this response (Martínez-Gómez et al., 2012). However, since both cells were grown

on exactly the same media, the exaggerated response in N+ is due primarily to the FPP pathway is which responsible for the much higher titres of terpene seen here. This is most likely mediated through cAMP-CRP catabolite de-repression. This cAMP-CRP mediated process was not seen in any of the mutants carrying the cytochrome P450 insert. There was also a significant increase in expression of heat shock proteins in cytochrome P450 containing strains (P and T), which was absent in the high yield N+ strain. This heat shock response is hypothetically incriminated in both the reduction of P450 synthesis and as well as terpene production (as seen in terpene titres of P and T). With the addition of P450 to the N strain (three plasmid system), titres for P reduced 5 fold. While the addition of P450 to N+ strain, titres for T reduced 90 fold. Cold shock response as seen in N+ seems to be favourable for terpene synthesis.The poor expression of cytochrome P450, is hypothesized to be the effect of increase in heat shock proteins.and a diminution or lack of the cAMP-CRP mediated carbon source scavenging capabilities of N+ As an effect, the P450 may be miss-folded or forming aggregates which could lead to over-expression of the heat shock proteins and proteases needed for degradation of the protein or be due to a translational block (Kusano et al., 1999).

7.4 Future work

The work in this thesis has highlighted some areas that would require further studies for validation purposes and/or metabolic engineering or optimisation.

7.4.1 Cytochrome P450

The poor expression of cytochrome P450 was seen to have a huge negative effect on terpene production based on the hypothesis that the cytochrome P450 suffers post-translational degradation or forms inclusion bodies. To investigate this would require the analysis of the mRNA transcript in those strains haboring the cytochrome P450 and a targeted proteomic method such as MRM analysis will show significant disparity between transcript level and expressed protein. Western blotting method would be used to probe the cell to see if the protein is forming inclusion body aggregates. It is also possible that the

poor expression would be at the level of transcription, in which case the mRNA transcript is expected to be very low as well. In which case, a codon optimised insert would be expected to address thisand lowering the temperature of cultivation even further would also help to improve protein expression and proper folding. Using a bacterial cytochrome P450 such as P450 BM-3 from *Bacillus megaterium* could also be another solution. P450 BM-3 is the fastest known P450 enzyme with a very good expression in *E. coli and* is self-contained with both P450 and reductase domains that can be engineered through site directed mutagenesis to use other short chain substrates such as monoterpenes and sesquiterpenes (Appel et al., 2001; Bernhardt, 2006).Knock out studies of the P450 gene through random mutagenesis using error-prone PCR (Park et al., 2010)/site directed mutagenesis which will also create a library of mutants for activity screening. This has the potential of changing the regio/sterio-selectivities of the P450 protein and producing mutants with specific or degenerate reactivities. More efficient and productive pathways could then be constructed by selecting appropriate mutants based on target product or substrate.

7.4.2 Terpene

As significant increases in terpene titres were achieved with the exogenous FPP over-producing pathway, further increases may be achievable through the following steps:

Use of a codon optimised terpene synthase could potentially increase the efficiency of transcription.

Reduction of the metabolic burden of multiple plasmids on the host cell by reducing the number of plasmids to two would be helpful. This can be done by cloning the terpene synthase, P450 and the ferrodoxin reductase into a single plasmid or using a mutant of the self-contained P450 BM-3. Alternatively, a genomic integration of the genes into the host genome could bypass this impediment (Csörgő and Pósfai, 2007; Datsenko and Wanner, 2000; Peti and Page, 2007)

Deletingthe genes that shunt acetyl-CoA from the mevalonate pathway into lactate, acetate and ethanol(Merlin et al., 2002) could increase the acetyl-CoA available to make precursors

and with a codon optimised terpene synthase; a even higher titres of the terpene may be achieved.

Seeing that CRP is associated with an increase in carbon scavenging and with increase in the activity of the CCM, the gene of this protein could serve as a metabolic engineering prospect. Indeed efforts at expressing cAMP-independent CRP mutant strains have shown positive result in carbon utilisation (Clomburg and Gonzalez, 2010; Karimova et al., 2004; Khankal et al., 2009)

7.4.3 Assay validation

Despite the use of an exogenous pathway to supply IPP/DMAPP precursors for terpene synthesis, there was still some growth retardation. This growth retardation cannot be wholly attributed to the plasmid burden or protein production. The relative abundance of fatty acid biosynthetic proteins could be due to overabundance HMG-CoA, as reported by Kizer et al. (2008). To assess this, metabolite assay of the culture for malonyl-CoA and HMG-CoA should be carried out. Additionally, fatty acid methyl ester (FAME)analysis should(Csörgő and Pósfai, 2007; Datsenko and Wanner, 2000; Peti and Page, 2007) be done to assess cell wall fatty acid profile.

7.4.4 Uncharacterised proteins

The four uncharacterised proteins that showed very high fold changes could possibly have some significance in the expression or otherwise of the terpene synthase or the P450. These genes can be explored using gene deletion methodologies (Merlin et al., 2002).

References

Aaron, J.A., Lin, X., Cane, D.E., Christianson, D.W., 2010. Structure of epi-isozizaene synthase from Streptomyces coelicolor A3 (2), a platform for new terpenoid cyclization templates. Biochemistry 49, 1787–1797.

Aebersold, R., Mann, M., 2003. Mass spectrometry-based proteomics. Nature 422, 198–207.

Aggarwal, K., Choe, L.H., Lee, K.H., 2006. Shotgun proteomics using the iTRAQ isobaric tags. Briefings in functional genomics & proteomics 5, 112–120.

Agger, S.A., Lopez-Gallego, F., Hoye, T.R., Schmidt-Dannert, C., 2008b. Identification of sesquiterpene synthases from Nostoc punctiforme PCC 73102 and Nostoc sp. strain PCC 7120. Journal of bacteriology 190, 6084–6096.

Aharoni, A., Jongsma, M.A., Bouwmeester, H.J., 2005. Volatile science? Metabolic engineering of terpenoids in plants. Trends in plant science 10, 594–602.

Ajikumar, P.K., Xiao, W.-H., Tyo, K.E., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T.H., Pfeifer, B., Stephanopoulos, G., 2010. Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. Science 330, 70–74.

Al-Soud, W.A., Rådström, P., 1998. Capacity of Nine Thermostable DNA Polymerases To Mediate DNA Amplification in the Presence of PCR-Inhibiting Samples. Appl. Environ. Microbiol. 64, 3748–3753.

Alder, A., Bigler, P., Werck-Reichhart, D., Al-Babili, S., 2009. In vitro characterization of Synechocystis CYP120A1 revealed the first nonanimal retinoic acid hydroxylase. FEBS Journal 276, 5416–5431.

Alper, H., Miyaoku, K., Stephanopoulos, G., 2005. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. Nature biotechnology 23, 612–616.

Alpert, A.J., 1990. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. Journal of Chromatography A 499, 177–196.

Alpert, A.J., Andrews, P.C., 1988. Cation-exchange chromatography of peptides on poly (2-sulfoethyl aspartamide)-silica. Journal of Chromatography A 443, 85–96.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25, 3389–3402.

Alvares, A.P., Schilling, G., Levin, W., Kuntzman, R., 1967. Studies on the induction of CO-binding pigments in liver microsomes by phenobarbital and 3-methylcholanthrene. Biochemical and biophysical research communications 29, 521–526.

Andrews, G.L., Simons, B.L., Young, J.B., Hawkridge, A.M., Muddiman, D.C., 2011. Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). Analytical chemistry 83, 5442–5446.

Ansari, M.Z., Yadav, G., Gokhale, R.S., Mohanty, D., 2004. NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. Nucleic acids research 32, W405–W413.

Apfel, C.M., Takács, B., Fountoulakis, M., Stieger, M., Keck, W., 1999. Use of Genomics To Identify Bacterial Undecaprenyl Pyrophosphate Synthetase: Cloning, Expression, and Characterization of the Essential uppS Gene. J. Bacteriol. 181, 483–492.

Appel, D., Lutz-Wahl, S., Fischer, P., Schwaneberg, U., Schmid, R.D., 2001. A P450 BM-3 mutant hydroxylates alkanes, cycloalkanes, arenes and heteroarenes. Journal of biotechnology 88, 167–171.

Atlung, T., Ingmer, H., 2003. H-NS: a modulator of environmentally regulated gene expression. Molecular microbiology 24, 7–17.

Atsumi, S., Liao, J.C., 2008. Metabolic engineering for advanced biofuels production from Escherichia coli. Current opinion in biotechnology 19, 414–419.

Bach, T.J., 1995. Some new aspects of isoprenoid biosynthesis in plants—a review. Lipids 30, 191–202.

Badet-Denisot, M.-A., Leriche, C., Massière, F., Badet, B., 1995. Nitrogen transfer in E. coli glucosamine-6P synthase. Investigations using substrate and bisubstrate analogs. Bioorganic & Medicinal Chemistry Letters 5, 815–820.

Bailey, J.E., 1991. Toward a science of metabolic engineering. Science 252, 1668–1675.

Bailey, J.E., Sburlati, A., Hatzimanikatis, V., Lee, K., Renner, W.A., Tsai, P.S., 1996. Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. Biotechnology and bioengineering 52, 109–121.

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B., 2007. Quantitative mass spectrometry in proteomics: a critical review. Analytical and bioanalytical chemistry 389, 1017–1031.

Barkovich, R., Liao, J.C., 2001a. Review: metabolic engineering of isoprenoids. Metabolic Engineering 3, 27–39.

Barkovich, R., Liao, J.C., 2001b. Review: metabolic engineering of isoprenoids. Metabolic Engineering 3, 27–39.

Barreteau, H., Kovac, A., Boniface, A., Sova, M., Gobec, S., Blanot, D., 2008. Cytoplasmic steps of peptidoglycan biosynthesis. FEMS microbiology reviews 32, 168–207.

Barrios-Llerena, M.E., Burja, A.M., Wright, P.C., 2007. Genetic analysis of polyketide synthase and peptide synthetase genes in cyanobacteria as a mining tool for secondary metabolites. Journal of industrial microbiology & biotechnology 34, 443–456.

Bauer, C.E., Elsen, S., Bird, T.H., 1999. Mechanisms for Redox Control of Gene Expression. Annual Review of Microbiology 53, 495–523.

Bernhardt, R., 2006. Cytochromes P450 as versatile biocatalysts. Journal of biotechnology 124, 128–145.

Bertea, C.M., Freije, J.R., Van der Woude, H., Verstappen, F.W.A., Perk, L., Marquez, V., De Kraker, J.-W., Posthumus, M.A., Jansen, B.J.M., De Groot, A.E., 2005. Identification of intermediates and enzymes involved in the early steps of artemisinin biosynthesis in Artemisia annua. Planta medica 71, 40–47.

Betenbaugh, M., Bentley, W., 2008. Metabolic engineering in the 21st century: meeting global challenges of sustainability and health. Current opinion in biotechnology 19, 411.

Bird, D.A., Facchini, P.J., 2001. Berberine bridge enzyme, a key branch-point enzyme in benzylisoquinoline alkaloid biosynthesis, contains a vacuolar sorting determinant. Planta 213, 888–897.

Blommel, P.G., Becker, K.J., Duvnjak, P., Fox, B.G., 2007. Enhanced Bacterial Protein Expression During Auto-Induction Obtained by Alteration of Lac Repressor Dosage and Medium Composition. Biotechnology progress 23, 585–598.

Bologna, F.P., Andreo, C.S., Drincovich, M.F., 2007. Escherichia coli Malic Enzymes: Two Isoforms with Substantial Differences in Kinetic Properties, Metabolic Regulation, and Structure. J. Bacteriol. 189, 5937–5946.

Bolwell, G.P., Bozak, K., Zimmerlin, A., 1994. Plant cytochrome P450. Phytochemistry 37, 1491–1506.

Boos, W., Ehmann, U., Bremer, E., Middendorf, A., Postma, P., 1987. Trehalase of Escherichia coli. Mapping and cloning of its structural gene and identification of the

enzyme as a periplasmic protein induced under high osmolarity growth conditions. Journal of Biological Chemistry 262, 13212–13218.

Boos, W., Shuman, H., 1998. Maltose/maltodextrin system of Escherichia coli: transport, metabolism, and regulation. Microbiology and Molecular Biology Reviews 62, 204–229.

Born, T.L., Blanchard, J.S., 1999. Structure/function studies on enzymes in the diaminopimelate pathway of bacterial cell wall biosynthesis. Current Opinion in Chemical Biology 3, 607–613.

Boucher, Y., Doolittle, W.F., 2000. The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. Molecular Microbiology 37, 703–716.

Bouhss, A., Trunkfield, A.E., Bugg, T.D.H., Mengin-Lecreulx, D., 2008. The biosynthesis of peptidoglycan lipid-linked intermediates. FEMS Microbiology Reviews 32, 208–233.

Bouvier, F., Rahier, A., Camara, B., 2005. Biogenesis, molecular regulation and function of plant isoprenoids. Progress in lipid research 44, 357–429.

Bozak, K.R., Yu, H., Sirevag, R., Christoffersen, R.E., 1990. Sequence analysis of ripening-related cytochrome P-450 cDNAs from avocado fruit. Proceedings of the National Academy of Sciences 87, 3904–3908.

Brahmachari, G., 2004. Neem—an omnipotent plant: a retrospection. Chembiochem 5, 408–421.

Buckler, D.R., Anand, G.S., Stock, A.M., 2000. Response-regulator phosphorylation and activation: a two-way street? Trends in microbiology 8, 153.

Burja, A.M., Banaigs, B., Abou-Mansour, E., Grant Burgess, J., Wright, P.C., 2001. Marine cyanobacteria—a prolific source of natural products. Tetrahedron 57, 9347–9377.

Butler, M.S., 2005. Natural products to drugs: natural product derived compounds in clinical trials. Natural product reports 22, 162–195.

Cane, D.E., Oliver, J.S., Harrison, P.H., Abell, C., Hubbard, B.R., Kane, C.T., Lattman, R., 1990. Biosynthesis of pentalenene and pentalenolactone. Journal of the American Chemical Society 112, 4513–4524.

Cane, D.E., Xue, Q., Fitzsimons, B.C., 1996. Trichodiene synthase. Probing the role of the highly conserved aspartate-rich region by site-directed mutagenesis. Biochemistry 35, 12369–12376.

Capon, R.J., 1995. Marine sesquiterpene/quinones, in: Atta-ur-Rahman (Ed.), Studies in Natural Products Chemistry. Elsevier, pp. 289–326.

Carmody, M., Murphy, B., Byrne, B., Power, P., Rai, D., Rawlings, B., Caffrey, P., 2005. Biosynthesis of amphotericin derivatives lacking exocyclic carboxyl groups. Journal of Biological Chemistry 280, 34420–34426.

Carter, O.A., Peters, R.J., Croteau, R., 2003. Monoterpene biosynthesis pathway construction in Escherichia coli. Phytochemistry 64, 425–433.

Castrillo, J.I., Zeef, L.A., Hoyle, D.C., Zhang, N., Hayes, A., Gardner, D.C., Cornell, M.J., Petty, J., Hakes, L., Wardleworth, L., 2007. Growth control of the eukaryote cell: a systems biology study in yeast. Journal of Biology 6, 4.

Chang, M.C., Keasling, J.D., 2006. Production of isoprenoid pharmaceuticals by engineered microbes. Nature chemical biology 2, 674–681.

Chappell, J., 1995. The biochemistry and molecular biology of isoprenoid metabolism. Plant physiology 107, 1.

Chau, M., Jennewein, S., Walker, K., Croteau, R., 2004. Taxol biosynthesis: molecular cloning and characterization of a cytochrome P450 taxoid 7β-hydroxylase. Chemistry & biology 11, 663–672.

Chemler, J.A., Koffas, M.A., 2008. Metabolic engineering for plant natural product biosynthesis in microbes. Current opinion in biotechnology 19, 597–605.

Chen, W., Lu, Y., Wu, J., Gao, M., Wang, A., Xu, B., 2011. Beta-elemene inhibits melanoma growth and metastasis via suppressing vascular endothelial growth factor-mediated angiogenesis. Cancer chemotherapy and pharmacology 67, 799–808.

Choe, S., Dilkes, B.P., Fujioka, S., Takatsuto, S., Sakurai, A., Feldmann, K.A., 1998. The DWF4 gene of Arabidopsis encodes a cytochrome P450 that mediates multiple 22α-hydroxylation steps in brassinosteroid biosynthesis. The Plant Cell Online 10, 231–243.

Choi, J.-W., Lee, J., Nishi, K., Kim, Y.-S., Jung, C.-H., Kim, J.-S., 2008. Crystal Structure of a Minimal Nitroreductase, ydjA, from Escherichia coli K12 with and without FMN Cofactor. Journal of molecular biology 377, 258–267.

Chou, W.-M., Kutchan, T.M., 1998. Enzymatic oxidations in the biosynthesis of complex alkaloids. The Plant Journal 15, 289–300.

Christianson, D.W., 2006. Structural Biology and Chemistry of the Terpenoid Cyclases. Chem. Rev. 106, 3412–3442.

Christman, M.F., Morgan, R.W., Jacobson, F.S., Ames, B.N., 1985. Positive control of a regulon for defenses against oxidative stress and some heat-shock proteins in Salmonella typhimurium. Cell 41, 753–762.

Clardy, J., Walsh, C., 2004. Lessons from natural molecules. Nature 432, 829–837.

Clomburg, J.M., Gonzalez, R., 2010. Biofuel production in Escherichia coli: the role of metabolic engineering and synthetic biology. Applied Microbiology and Biotechnology 86, 419–434.

Clouse, S.D., Langford, M., McMorris, T.C., 1996. A brassinosteroid-insensitive mutant in Arabidopsis thaliana exhibits multiple defects in growth and development. Plant Physiology 111, 671–678.

Connolly, D.M., Winkler, M.E., 1989. Genetic and physiological relationships among the miaA gene, 2-methylthio-N6-(delta 2-isopentenyl)-adenosine tRNA modification, and spontaneous mutagenesis in Escherichia coli K-12. Journal of bacteriology 171, 3233–3246.

Connolly, J.D., Hill, R.A., 1991. Dictionary of terpenoids. 1. Mono-and sesquiterpenoids. CRC Press.

Coon, M.J., Vaz, A.D., Bestervelt, L.L., 1996. Cytochrome P450 2: peroxidative reactions of diversozymes. The FASEB journal 10, 428–434.

Cortassa, S., 2002. An introduction to metabolic and cellular engineering. World Scientific Press.

Cox, J., Mann, M., 2008. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. Nature biotechnology 26, 1367–1372.

Cox, J., Mann, M., 2011. Quantitative, high-resolution proteomics for data-driven systems biology. Annual review of biochemistry 80, 273–299.

Cripps, R.E., Eley, K., Leak, D.J., Rudd, B., Taylor, M., Todd, M., Boakes, S., Martin, S., Atkinson, T., 2009. Metabolic engineering of Geobacillus thermoglucosidasius for high yield ethanol production. Metabolic engineering 11, 398–408.

Croteau, R.B., Davis, E.M., Ringer, K.L., Wildung, M.R., 2005. (−)-Menthol biosynthesis and molecular genetics. Naturwissenschaften 92, 562–577.

Csörgő, B., Pósfai, G., 2007. Directed homologous recombination for genome engineering in Escherichia coli. Acta Biologica Hungarica 58, 1–10.

Cunningham, F.X., Lafond, T.P., Gantt, E., 2000. Evidence of a role for LytB in the nonmevalonate pathway of isoprenoid biosynthesis. Journal of bacteriology 182, 5841–5848.

Cupp-Vickery, J.R., Poulos, T.L., 1995. Structure of cytochrome P450eryF involved in erythromycin biosynthesis. Nature Structural & Molecular Biology 2, 144–153.

Daly, J.W., 1998. Thirty years of discovering arthropod alkaloids in amphibian skin. Journal of natural products 61, 162–172.

Daly, J.W., Spande, T.F., Garraffo, H.M., 2005. Alkaloids from amphibian skin: a tabulation of over eight-hundred compounds. Journal of Natural Products 68, 1556–1575.

Dame, R.T., 2005. The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. Molecular microbiology 56, 858–870.

Dame, R.T., Noom, M.C., Wuite, G.J., 2006. Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. Nature 444, 387–390.

Dardel, F., 1988. Computer simulation of DNA ligation: determination of initial DNA concentrations favouring the formation of recombinant molecules. Nucleic acids research 16, 1767–1778.

Datsenko, K.A., Wanner, B.L., 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proceedings of the National Academy of Sciences 97, 6640–6645.

Davalieva, K., Efremov, G.D., 2010. Influence of salts and PCR inhibitors on the amplification capacity of three thermostable DNA polymerases.

Davis, E.M., Croteau, R., 2000. Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes, in: Biosynthesis. Springer, pp. 53–95.

Dawson, J.H., Sono, M., 1987. Cytochrome P-450 and chloroperoxidase: thiolate-ligated heme enzymes. Spectroscopic determination of their active-site structures and mechanistic implications of thiolate ligation. Chemical Reviews 87, 1255–1276.

De Luca, V., St Pierre, B., 2000. The cell and developmental biology of alkaloid biosynthesis. Trends in plant science 5, 168–173.

Degtyarenko, K.N., Archakov, A.I., 1993. Molecular evolution of P450 superfamily and P450-containing monooxygenase systems. FEBS letters 332, 1–8.

DeJong, J.M., Liu, Y., Bollon, A.P., Long, R.M., Jennewein, S., Williams, D., Croteau, R.B., 2006. Genetic engineering of taxol biosynthetic genes in Saccharomyces cerevisiae. Biotechnology and bioengineering 93, 212–224.

Denisov, I.G., Makris, T.M., Sligar, S.G., Schlichting, I., 2005. Structure and chemistry of cytochrome P450. Chemical reviews 105, 2253–2278.

Diaz Ricci, J.C., Hernández, M.E., 2000. Plasmid effects on Escherichia coli metabolism. Critical reviews in biotechnology 20, 79–108.

Dietrich, J.A., Yoshikuni, Y., Fisher, K.J., Woolard, F.X., Ockey, D., McPhee, D.J., Renninger, N.S., Chang, M.C., Baker, D., Keasling, J.D., 2009. A novel semi-biosynthetic route for artemisinin production using engineered substrate-promiscuous P450BM3. ACS chemical biology 4, 261–267.

Dippel, R., Boos, W., 2005. The maltodextrin system of Escherichia coli: metabolism and transport. Journal of bacteriology 187, 8322–8331.

Donadio, S., McAlpine, J.B., Sheldon, P.J., Jackson, M., Katz, L., 1993. An erythromycin analog produced by reprogramming of polyketide synthesis. Proceedings of the National Academy of Sciences 90, 7119–7123.

Durst, F., Nelson, D.R., 1995. Diversity and evolution of plant P450 and P450-reductases. Drug metabolism and drug interactions 12, 189–206.

Edwards, D.J., Marquez, B.L., Nogle, L.M., McPhail, K., Goeger, D.E., Roberts, M.A., Gerwick, W.H., 2004. Structure and Biosynthesis of the Jamaicamides, New Mixed Polyketide-Peptide Neurotoxins from the Marine Cyanobacterium< i> Lyngbya majuscula</i>. Chemistry & Biology 11, 817–833.

Edwards, J.S., Palsson, B.O., 2000. The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proceedings of the National Academy of Sciences 97, 5528–5533.

Edwards, P.A., Ericsson, J., 1999. Sterols and isoprenoids: signaling molecules derived from the cholesterol biosynthetic pathway. Annual review of biochemistry 68, 157–185.

Eguchi, T., Morita, M., Kakinuma, K., 1998. Multigram Synthesis of Mevalonolactone-d 9 and Its Application to Stereochemical Analysis by 1H NMR of the Saturation Reaction in the Biosynthesis of the 2, 3-Di-O-phytanyl-sn-glycerol Core of the Archaeal Membrane Lipid. Journal of the American Chemical Society 120, 5427–5433.

Eilert, K.D., Foran, D.R., 2009. Polymerase resistance to polymerase chain reaction inhibitors in bone*. Journal of forensic sciences 54, 1001–1007.

Eisenreich, W., Bacher, A., Arigoni, D., Rohdich, F., 2004. Biosynthesis of isoprenoids via the non-mevalonate pathway. Cellular and Molecular Life Sciences CMLS 61, 1401–1426.

El-Mansi, E.M.T., Holms, W.H., 1989. Control of Carbon Flux to Acetate Excretion During Growth of Escherichia coli in Batch and Continuous Cultures. J Gen Microbiol 135, 2875–2883.

Enfissi, E., Fraser, P.D., Lois, L.-M., Boronat, A., Schuch, W., Bramley, P.M., 2005. Metabolic engineering of the mevalonate and non-mevalonate isopentenyl diphosphate-forming pathways for the production of health-promoting isoprenoids in tomato. Plant Biotechnology Journal 3, 17–27.

Engels, B., Dahm, P., Jennewein, S., 2008. Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. Metabolic engineering 10, 201–206.

Estabrook, R.W., 2003. A passion for P450s (remembrances of the early history of research on cytochrome P450). Drug metabolism and disposition 31, 1461–1473.

Estabrook, R.W., Cooper, D.Y., Rosenthal, O., 1963. The light reversible carbon monoxide inhibition of the steroid C21-hydroxylase system of the adrenal cortex. Biochemische Zeitschrift 338, 741.

Evans, C., Noirel, J., Ow, S.Y., Salim, M., Pereira-Medrano, A.G., Couto, N., Pandhal, J., Smith, D., Pham, T.K., Karunakaran, E., Zou, X., Biggs, C.A., Wright, P.C., 2012. An insight into iTRAQ: where do we stand now? Anal Bioanal Chem 404, 1011–1027.

Faulkner, D.J., 1988. Marine natural products. Natural Product Reports 5, 613–663.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M., 1989. Electrospray ionization for mass spectrometry of large biomolecules. Science 246, 64–71.

Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.N., Willmitzer, L., 2000. Metabolite profiling for plant functional genomics. Nature biotechnology 18, 1157–1161.

Fu, P., Panke, S., 2009. Systems Biology and Synthetic Biology,. John Wiley & Sons.

Fuszard, M.A., Ow, S.Y., Gan, C.S., Noirel, J., Ternan, N.G., McMullan, G., Biggs, C.A., Reardon, K.F., Wright, P.C., 2013. The quantitative proteomic response of Synechocystis sp. PCC6803 to phosphate acclimation. Aquatic Biosystems 9, 5.

Gan, C.S., Reardon, K.F., Wright, P.C., 2005. Comparison of protein and peptide prefractionation methods for the shotgun proteomic analysis of Synechocystis sp. PCC 6803. Proteomics 5, 2468–2478.

Garfinkel, D., 1958. Studies on pig liver microsomes. I. Enzymic and pigment composition of different microsomal fractions. Archives of biochemistry and biophysics 77, 493–509.

Geron, C., Rasmussen, R., R. Arnts, R., Guenther, A., 2000. A review and synthesis of monoterpene speciation from forests in the United States. Atmospheric Environment 34, 1761–1781.

Gershenzon, J., Croteau, R., 1990. Regulation of monoterpene biosynthesis in higher plants, in: Biochemistry of the Mevalonic Acid Pathway to Terpenoids. Springer, pp. 99–160.

Gilar, M., Olivova, P., Daly, A.E., Gebler, J.C., 2005. Orthogonality of separation in two-dimensional liquid chromatography. Analytical chemistry 77, 6426–6434.

Glen, A., Gan, C.S., Hamdy, F.C., Eaton, C.L., Cross, S.S., Catto, J.W., Wright, P.C., Rehman, I., 2008. iTRAQ-facilitated proteomic analysis of human prostate cancer cells identifies proteins associated with progression. Journal of proteome research 7, 897–907.

Glick, B.R., 1995. Metabolic load and heterologous gene expression. Biotechnology Advances 13, 247–261.

Goldstein, J., Pollitt, N.S., Inouye, M., 1990. Major cold shock protein of Escherichia coli. Proceedings of the National Academy of Sciences 87, 283–287.

Gómez-Galera, S., Pelacho, A.M., Gené, A., Capell, T., Christou, P., 2007. The genetic manipulation of medicinal and aromatic plants. Plant Cell Reports 26, 1689–1715.

Gonzalez, F.J., Nebert, D.W., 1990. Evolution of the P450 gene superfamily:: animal-plant "warfare", molecular drive and human genetic differences in drug oxidation. Trends in Genetics 6, 182–186.

Goodell, E.W., Higgins, C.F., 1987. Uptake of cell wall peptides by Salmonella typhimurium and Escherichia coli. J. Bacteriol. 169, 3861–3865.

Gotoh, O., 1992. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. Journal of Biological Chemistry 267, 83–90.

Goudie, A.S., 2013. The Human Impact on the Natural Environment: Past, Present, and Future. John Wiley & Sons.

Gräwert, T., Groll, M., Rohdich, F., Bacher, A., Eisenreich, W., 2011. Biochemistry of the non-mevalonate isoprenoid pathway. Cellular and Molecular Life Sciences 68, 3797–3814.

Griffin, B.W., 1991. Chloroperoxidase: a review. Peroxidases in chemistry and biology 2, 85–137.

Grimm, C.C., Champagne, E.T., Ohtsubo, K., 2001. SPME/GC/MS. Flavor, Fragrance, and Odor Analysis 115, 229.

Gross, C.A., 1996. Function and regulation of the heat shock proteins. Escherichia coli and Salmonella: cellular and molecular biology, 2nd ed. ASM Press, Washington, DC 1382–1399.

Groves, J.T., 1997. Artificial enzymes: The importance of being selective. Nature 389, 329–330.

Grundmann, A., Kuznetsova, T., Afiyatullov, S.S., Li, S.-M., 2008. FtmPT2, an N-Prenyltransferase from Aspergillus fumigatus, Catalyses the Last Step in the Biosynthesis of Fumitremorgin B. Chembiochem 9, 2059–2063.

Gutiérrez, R.M.P., Flores, A.M., Solís, R.V., Jimenez, J.C., 2008. Two new antibacterial norabietane diterpenoids from cyanobacteria, Microcoleous lacustris. J Nat Med 62, 328–331.

Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y., Aebersold, R., 2000. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. PNAS 97, 9390–9395.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R., 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nature biotechnology 17, 994–999.

Hahn, F.M., Hurlburt, A.P., Poulter, C.D., 1999. Escherichia coli open reading frame 696 is idi, a nonessential gene encoding isopentenyl diphosphate isomerase. Journal of bacteriology 181, 4499–4504.

Hamdane, D., Zhang, H., Hollenberg, P., 2008. Oxygen activation by cytochrome P450 monooxygenase. Photosynthesis research 98, 657–666.

Han, M.-J., Lee, S.Y., 2006. The Escherichia coli proteome: past, present, and future prospects. Microbiology and molecular biology reviews 70, 362–439.

Hannemann, F., Bichet, A., Ewen, K.M., Bernhardt, R., 2007. Cytochrome P450 systems—biological variations of electron transport chains. Biochimica et Biophysica Acta (BBA)-General Subjects 1770, 330–344.

Hannig, G., Makrides, S.C., 1998. Strategies for optimizing heterologous protein expression in Escherichia coli. Trends in Biotechnology 16, 54–60.

Hanson, J.R., 2003. Natural Products: The Secondary Metabolites. Royal Society of Chemistry.

Hanson, J.R., 2007. The classes of natural product and their isolation, in: Natural Products.

Harada, H., Shindo, K., Iki, K., Teraoka, A., Okamoto, S., Yu, F., Hattan, J., Utsumi, R., Misawa, N., 2011. Efficient functional analysis system for cyanobacterial or plant cytochromes P450 involved in sesquiterpene biosynthesis. Applied microbiology and biotechnology 90, 467–476.

Harada, H., Yu, F., Okamoto, S., Kuzuyama, T., Utsumi, R., Misawa, N., 2009. Efficient synthesis of functional isoprenoids from acetoacetate through metabolic pathway-engineered Escherichia coli. Applied microbiology and biotechnology 81, 915–925.

Harel, Y., Ohad, I., Kaplan, A., 2004. Activation of photosynthesis and resistance to photoinhibition in cyanobacteria within biological desert crust. Plant physiology 136, 3070–3079.

Harrison, D.M., 1990. The biosynthesis of triterpenoids, steroids, and carotenoids. Nat. Prod. Rep. 7, 459–484.

Hartzell, H., 1991. The yew tree: a thousand whispers: biography of a species. Hulogosi Eugene, OR.

Hassan, H.M., Fridovich, I., 1977. Enzymatic defenses against the toxicity of oxygen and of streptonigrin in Escherichia coli. Journal of bacteriology 129, 1574–1583.

Hayaishi, O., 2005. Fifty years of oxygen activation. Journal of Biological Inorganic Chemistry 10, 1–2.

Hayaishi, O., Stanier, R.Y., 1951. The bacterial oxidation of tryptophan III.: Enzymatic Activities of Cell-Free Extracts from Bacteria Employing the Aromatic Pathway1. Journal of bacteriology 62, 691.

Hefner, J., Rubenstein, S.M., Ketchum, R.E., Gibson, D.M., Williams, R.M., Croteau, R., 1996. Cytochrome P450-catalyzed hydroxylation of taxa-4 (5), 11 (12)-diene to taxa-4 (20), 11 (12)-dien-5a-o1: the first oxygenation step in taxol biosynthesis. Chemistry & biology 3, 479–489.

Herrero, A., Flores, E., 2008. The Cyanobacteria: Molecular Biology, Genomics, and Evolution. Horizon Scientific Press.

Herz, S., Wungsintaweekul, J., Schuhr, C.A., Hecht, S., Lüttgen, H., Sagner, S., Fellermeier, M., Eisenreich, W., Zenk, M.H., Bacher, A., 2000. Biosynthesis of terpenoids: YgbB protein converts 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate to 2C-methyl-D-erythritol 2, 4-cyclodiphosphate. Proceedings of the National Academy of Sciences 97, 2486–2490.

Higgins, C.F., 1992. ABC Transporters: From Microorganisms to Man. Annual Review of Cell Biology 8, 67–113.

Hirschberg, J., Chamovitz, D., 2004. Carotenoids in cyanobacteria. The molecular biology of cyanobacteria 559–579.

Hoffmann, D., Hevel, J.M., Moore, R.E., Moore, B.S., 2003. Sequence analysis and biochemical characterization of the nostopeptolide A biosynthetic gene cluster from Nostoc sp. GSV224. Gene 311, 171.

Hoffmann, E. de, Stroobant, V., 2013. Mass Spectrometry: Principles and Applications. John Wiley & Sons.

Hoffmann, F., Rinas, U., 2004. Stress induced by recombinant protein production in Escherichia coli, in: Physiological Stress Responses in Bioprocesses. Springer, pp. 73–92.

Hopwood, D.A., 2007. Streptomyces in nature and medicine. Oxford University Press.

Horbach, S., Sahm, H., Welle, R., 1993. Isoprenoid biosynthesis in bacteria: two different pathways? FEMS microbiology letters 111, 135–140.

Horwitz, S.B., 1994. How to make taxol from scratch. Nature 367, 593.

Hsu, E., 2006. The history of qing hao in the Chinese materia medica. Transactions of the Royal Society of Tropical Medicine and hygiene 100, 505–508.

Scorei, R.I., Popa, R., 2010. Boron-containing compounds as preventive and chemotherapeutic agents for cancer. Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents) 10, 346–351.

Ikawa, M., Sasner, J.J., Haney, J.F., 2001. Activity of cyanobacterial and algal odor compounds found in lake waters on green alga Chlorella pyrenoidosa growth. Hydrobiologia 443, 19–22.

Inoue, A., Horikoshi, K., 1991. Estimation of solvent-tolerance of bacteria by the solvent parameter log P. Journal of Fermentation and Bioengineering 71, 194–196.

Ioannides, C., 1996. cytochrome P450. CRC Press.

Isman, M.B., Koul, O., Luczynski, A., Kaminski, J., 1990. Insecticidal and antifeedant bioactivities of neem oils and their relationship to azadirachtin content. Journal of agricultural and food chemistry 38, 1406–1411.

Istvan, E.S., Deisenhofer, J., 2001. Structural mechanism for statin inhibition of HMG-CoA reductase. Science 292, 1160–1164.

Ito, M., 2008. Studies on perilla relating to its essential oil and taxonomy. Phytochemistry research progress. Nova Science Publishers, New York 13–30.

Jaki, B., Heilmann, J., Sticher, O., 2000a. New antibacterial metabolites from the cyanobacterium Nostoc commune(EAWAG 122b). J. Nat. Prod. 63, 1283–1285.

Jaki, B., Orjala, J., Heilmann, J., Linden, A., Vogler, B., Sticher, O., 2000b. Novel extracellular diterpenoids with biological activity from the cyanobacterium Nostoc commune. J. Nat. Prod. 63, 339–343.

Jennewein, S., Croteau, R., 2001. Taxol: biosynthesis, molecular genetics, and biotechnological applications. Applied Microbiology and Biotechnology 57, 13–19.

Jiang, W., Hou, Y., Inouye, M., 1997. CspA, the major cold-shock protein of Escherichia coli, is an RNA chaperone. Journal of Biological Chemistry 272, 196–202.

Jones, A.C., Gu, L., Sorrels, C.M., Sherman, D.H., Gerwick, W.H., 2009. New tricks from ancient algae: natural products biosynthesis in marine cyanobacteria. Current opinion in chemical biology 13, 216–223.

Jones, P.G., Mitta, M., Kim, Y., Jiang, W., Inouye, M., 1996. Cold shock induces a major ribosomal-associated protein that unwinds double-stranded RNA in Escherichia coli. Proceedings of the National Academy of Sciences 93, 76–80.

Kakinuma, K., Yamagishi, M., Fujimoto, Y., Oshima, T., Ikekawa, N., 1988. Stereochemistry of the biosynthesis of sn-2, 3-O-diphytanyl glycerol, membrane lipid of archaebacteria Halobacterium halobium. Journal of the American Chemical Society 110, 4861–4863.

Karimova, G., Ladant, D., Ullmann, A., 2004. Relief of catabolite repression in a cAMP-independent catabolite gene activator mutant of Escherichia coli. Research in microbiology 155, 76–79.

Ke, N., Baudry, J., Makris, T.M., Schuler, M.A., Sligar, S.G., 2005. A retinoic acid binding cytochrome P450: CYP120A1 from Synechocystis sp. PCC 6803. Archives of biochemistry and biophysics 436, 110–120.

Keasling, J.D., 2010. Microbial production of isoprenoids, in: Handbook of Hydrocarbon and Lipid Microbiology. Springer, pp. 2951–2966.

Kelly, S.L., Lamb, D.C., Cannieux, M., Greetham, D., Jackson, C.J., Marczylo, T., Ugochukwu, C., Kelly, D.E., 2001. An old activity in the cytochrome P450 superfamily (CYP5I) and a new story of drugs and resistance. Biochemical Society Transactions 29, 122–127.

Kenney, L.J., 1997. Kinase Activity of EnvZ, an Osmoregulatory Signal Transducing Protein of Escherichia coli. Archives of biochemistry and biophysics 346, 303–311.

Khankal, R., Chin, J.W., Ghosh, D., Cirino, P.C., 2009. Transcriptional effects of CRP* expression in Escherichia coli. Journal of biological engineering 3, 13.

Kholodenko, B.N., Westerhoff, H.V., 2004. Metabolic engineering in the post genomic era. Taylor & Francis.

Kim, S.O., Merchant, K., Nudelman, R., Beyer Jr, W.F., Keng, T., DeAngelo, J., Hausladen, A., Stamler, J.S., 2002. OxyR: a molecular code for redox-related signaling. Cell 109, 383–396.

Kingston, D.G., Newman, D.J., 2007. Taxoids: cancer-fighting compounds from nature. Current opinion in drug discovery & development 10, 130–144.

Kirby, J., Keasling, J.D., 2009. Biosynthesis of plant isoprenoids: perspectives for microbial engineering. Annual review of plant biology 60, 335–355.

Kizer, L., Pitera, D.J., Pfleger, B.F., Keasling, J.D., 2008. Application of Functional Genomics to Pathway Optimization for Increased Isoprenoid Production. Appl. Environ. Microbiol. 74, 3229–3241.

Klingenberg, M., 2003. Pigments of rat liver microsomes. Archives of biochemistry and biophysics 409, 2–6.

Köcher, T., Swart, R., Mechtler, K., 2011. Ultra-High-Pressure RPLC Hyphenated to an LTQ-Orbitrap Velos Reveals a Linear Relation between Peak Capacity and Number of Identified Peptides. Anal. Chem. 83, 2699–2704.

Koehn, F.E., 2008. High impact technologies for natural products screening, in: Natural Compounds as Drugs Volume I. Springer, pp. 175–210.

Koehn, F.E., Carter, G.T., 2005. The evolving role of natural products in drug discovery. Nature Reviews Drug Discovery 4, 206–220.

Koppisch, A.T., Fox, D.T., Blagg, B.S., Poulter, C.D., 2002. E. coli MEP synthase: steady-state kinetic analysis and substrate binding. Biochemistry 41, 236–243.

Kovach, M.E., Elzer, P.H., Steven Hill, D., Robertson, G.T., Farris, M.A., Roop II, R.M., Peterson, K.M., 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene 166, 175–176.

Kühnel, K., Ke, N., Cryle, M.J., Sligar, S.G., Schuler, M.A., Schlichting, I., 2008. Crystal Structures of Substrate-Free and Retinoic Acid-Bound Cyanobacterial Cytochrome P450 CYP120A1†‡. Biochemistry 47, 6552–6559.

Kunte, H.J., Crane, R.A., Culham, D.E., Richmond, D., Wood, J.M., 1999. Protein ProQ Influences Osmotic Activation of Compatible Solute Transporter ProP in Escherichia coli K-12. Journal of Bacteriology 181, 1537.

Kusano, K., Waterman, M.R., Sakaguchi, M., Omura, T., Kagawa, N., 1999. Protein Synthesis Inhibitors and Ethanol Selectively Enhance Heterologous Expression of P450s and Related Proteins in Escherichia coli. Archives of biochemistry and biophysics 367, 129–136.

Kuzuyama, T., Shimizu, T., Takahashi, S., Seto, H., 1998. Fosmidomycin, a specific inhibitor of 1-deoxy-D-xylulose 5-phosphate reductoisomerase in the nonmevalonate pathway for terpenoid biosynthesis. Tetrahedron Letters 39, 7913–7916.

Lamb, D.C., Guengerich, F.P., Kelly, S.L., Waterman, M.R., 2006. Exploiting Streptomyces coelicolor A3 (2) P450s as a model for application in drug discovery.

Lamb, D.C., Ikeda, H., Nelson, D.R., Ishikawa, J., Skaug, T., Jackson, C., Omura, S., Waterman, M.R., Kelly, S.L., 2003. Cytochrome P450 complement (CYPome) of the avermectin-producer Streptomyces avermitilis and comparison to that of Streptomyces coelicolor A3 (2). Biochemical and biophysical research communications 307, 610–619.

Larsen, T.O., 1997. Identification of cheese-associated fungi using selected ion monitoring of volatile terpenes. Letters in applied microbiology 24, 463–466.

Laybourn-Parry, J., Pearce, D.A., 2007. The biodiversity and ecology of Antarctic lakes: models for evolution. Philosophical Transactions of the Royal Society B: Biological Sciences 362, 2273–2289.

Lee, T.S., Krupa, R.A., Zhang, F., Hajimorad, M., Holtz, W.J., Prasad, N., Lee, S.K., Keasling, J.D., 2011. BglBrick vectors and datasheets: A synthetic biology platform for gene expression. Journal of Biological Engineering 5, 12.

Leeper, F.J., Vederas, J.C., 2000. Biosynthesis: Aromatic Polyketides, Isoprenoids, Alkaloids. Springer.

Leitner, A., Lindner, W., 2006. Chemistry meets proteomics: The use of chemical tagging reactions for MS-based proteomics. Proteomics 6, 5418–5434.

Lemieux, M.J., Huang, Y., Wang, D.-N., 2004. Glycerol-3-phosphate transporter of< i> Escherichia coli: Structure, function and regulation. Research in microbiology 155, 623–629.

Lepesheva, G.I., Waterman, M.R., 2007. Sterol 14α-demethylase cytochrome P450 (CYP51), a P450 in all biological kingdoms. Biochimica et Biophysica Acta (BBA)-General Subjects 1770, 467–477.

Lesburg, C.A., Zhai, G., Cane, D.E., Christianson, D.W., 1997. Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. Science 277, 1820–1824.

Leung, J., Giraudat, J., 1998. Abscisic acid signal transduction. Annual review of plant biology 49, 199–222.

Levine, D.P., 2006. Vancomycin: a history. Clinical Infectious Diseases 42, S5–S12.

Lewis, D.F., Pratt, J.M., 1998. The P450 catalytic cycle and oxygenation mechanism. Drug metabolism reviews 30, 739–786.

Lewis, D.F.V., 2001. Guide to Cytochromes P450: Structure and Function. CRC Press.

Li, Q.Q., Wang, G., Huang, F., Banda, M., Reed, E., 2010. Antineoplastic effect of β-elemene on prostate cancer cells and other types of solid tumour cells. Journal of Pharmacy and Pharmacology 62, 1018–1027.

Li, X., Wang, G., Zhao, J., Ding, H., Cunningham, C., Chen, F., Flynn, D.C., Reed, E., Li, Q.Q., 2005. Antiproliferative effect of β-elemene in chemoresistant ovarian carcinoma cells is mediated through arrest of the cell cycle at the G2-M phase. Cellular and molecular life sciences 62, 894–904.

Li, Z., Adams, R.M., Chourey, K., Hurst, G.B., Hettich, R.L., Pan, C., 2012. Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. Journal of proteome research 11, 1582–1590.

Lichtenthaler, H.K., Rohmer, M., Schwender, J., 1997. Two independent biochemical pathways for isopentenyl diphosphate and isoprenoid biosynthesis in higher plants. Physiologia Plantarum 101, 643–652.

Liebler, D.C., 2002. Introduction to Proteomics: Tools for the New Biology. Springer.

Lindberg, P., Park, S., Melis, A., 2010. Engineering a platform for photosynthetic isoprene production in cyanobacteria, using Synechocystis as the model organism. Metabolic engineering 12, 70–79.

Liu, C., Zhao, Y., Wang, Y., 2006. Artemisinin: current state and perspectives for biotechnological production of an antimalarial drug. Applied microbiology and biotechnology 72, 11–20.

Liu, J., Walsh, C.T., 1990. Peptidyl-prolyl cis-trans-isomerase from Escherichia coli: a periplasmic homolog of cyclophilin that is not inhibited by cyclosporin A. PNAS 87, 4028–4032.

Liu, X., Ferenci, T., 2001. An analysis of multifactorial influences on the transcriptional control of ompF and ompC porin expression under nutrient limitation. Microbiology 147, 2981–2989.

Liu, Y., Wang, L., Jung, J.H., Zhang, S., 2007. Sesterterpenoids. Nat Prod Rep 24, 1401–1429.

Longworth, J., Noirel, J., Pandhal, J., Wright, P.C., Vaidyanathan, S., 2012. HILIC- and SCX-Based Quantitative Proteomics of Chlamydomonas reinhardtii during Nitrogen Starvation Induced Lipid and Carbohydrate Accumulation. J. Proteome Res. 11, 5959–5971.

Lorne Burke, T.W., Mant, C.T., Black, J.A., Hodges, R.S., 1989. Strong cation-exchange high-performance liquid chromatography of peptides: Effect of non-specific hydrophobic interactions and linearization of peptide retention behaviour. Journal of Chromatography A 476, 377–389.

Luijsterburg, M.S., Noom, M.C., Wuite, G.J., Dame, R.T., 2006. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. Journal of structural biology 156, 262.

Lüking, A., Stahl, U., Schmidt, U., 1998. The protein family of RNA helicases. Critical reviews in biochemistry and molecular biology 33, 259–296.

Lüttgen, H., Rohdich, F., Herz, S., Wungsintaweekul, J., Hecht, S., Schuhr, C.A., Fellermeier, M., Sagner, S., Zenk, M.H., Bacher, A., 2000. Biosynthesis of terpenoids: YchB protein of Escherichia coli phosphorylates the 2-hydroxy group of 4-diphosphocytidyl-2C-methyl-D-erythritol. Proceedings of the National Academy of Sciences 97, 1062–1067.

Lynch, A.S., 1996. Responses to molecular oxygen. Escherichia coli and Salmonella 1526–1538.

Mahmoud, S.S., Croteau, R.B., 2002. Strategies for transgenic manipulation of monoterpene biosynthesis in plants. Trends in Plant Science 7, 366–373.

Maia, B.H., Paula, J.R. de, Sant'Ana, J., Silva, M., Fernandes, J.B., Vieira, P.C., Costa, M. do S., Ohashi, O.S., Silva, J.N.M., 2000. Essential oils of Toona and Cedrela species (Meliaceae): taxonomic and ecological implications. Journal of the Brazilian Chemical Society 11, 629–639.

Mandel, M.A., Feldmann, K.A., Herrera-Estrella, L., Rocha-Sosa, M., León, P., 1996. CLA1, a novel gene required for chloroplast development, is highly conserved in evolution. The Plant Journal 9, 649–658.

Mann, J., 1987. Secondary metabolism. Oxford University Press.

Mann, M., Kulak, N.A., Nagaraj, N., Cox, J., 2013. The Coming Age of Complete, Accurate, and Ubiquitous Proteomes. Molecular Cell 49, 583–590.

Marguilis, L., Sagan, D., Eldredge, N., 2000. What is life? Univ of California press.

Martin, V.J., Yoshikuni, Y., Keasling, J.D., 2001. The in vivo synthesis of plant sesquiterpenes by Escherichia coli. Biotechnology and bioengineering 75, 497–503.

Martin, V.J.J., Pitera, D.J., Withers, S.T., Newman, J.D., Keasling, J.D., 2003. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. Nat Biotech 21, 796–802.

Martínez-Gómez, K., Flores, N., Castañeda, H.M., Martínez-Batallar, G., Hernández-Chávez, G., Ramírez, O.T., Gosset, G., Encarnación, S., Bolivar, F., 2012. New insights into Escherichia coli metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. Microb Cell Fact 11, 46.

Mashego, M.R., Rumbold, K., De Mey, M., Vandamme, E., Soetaert, W., Heijnen, J.J., 2007. Microbial metabolomics: past, present and future methodologies. Biotechnology letters 29, 1–16.

Mau, C.J., Croteau, R., 2006. Cytochrome P450 oxygenases of monoterpene metabolism. Phytochemistry Reviews 5, 373–383.

Maurel, C., Reizer, J., Schroeder, J.I., Chrispeels, M.J., Saier, M.H., 1994. Functional characterization of the Escherichia coli glycerol facilitator, GlpF, in Xenopus oocytes. Journal of Biological Chemistry 269, 11869–11872.

McLean, M.A., Maves, S.A., Weiss, K.E., Krepich, S., Sligar, S.G., 1998. Characterization of a Cytochrome P450 from the Acidothermophilic ArchaeaSulfolobus solfataricus. Biochemical and biophysical research communications 252, 166–172.

Méjean, V., Lobbi-Nivol, C., Lepelletier, M., Giordano, G., Chippaux, M., Pascal, M.C., 1994. TMAO anaerobic respiration in Escherichia coli: involvement of the tor operon. Molecular Microbiology 11, 1169–1179.

Merfort, I., 2002. Review of the analytical techniques for sesquiterpenes and sesquiterpene lactones. Journal of Chromatography A 967, 115–130.

Merlin, C., McAteer, S., Masters, M., 2002. Tools for Characterization of Escherichia coli Genes of Unknown Function. J. Bacteriol. 184, 4573–4581.

Meshnick, S.R., Taylor, T.E., Kamchonwongpaisan, S., 1996. Artemisinin and the antimalarial endoperoxides: from herbal remedy to targeted chemotherapy. Microbiol. Rev. 60, 301–315.

Michalski, A., Damoc, E., Hauschild, J.-P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., Horning, S., 2011. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Molecular & Cellular Proteomics 10.

Misawa, N., 2011. Pathway engineering for functional isoprenoids. Current Opinion in Biotechnology 22, 627–633.

Miyagi, M., Rao, K.C., 2007. Proteolytic 18O-labeling strategies for quantitative proteomics. Mass spectrometry reviews 26, 121–136.

Mo, S., Krunic, A., Pegan, S.D., Franzblau, S.G., Orjala, J., 2009. An Antimicrobial Guanidine-Bearing Sesterterpene from the Cultured Cyanobacterium Scytonema sp. J. Nat. Prod. 72, 2043–2045.

Moldoveanu, N., Kates, M., 1988. Biosynthetic studies of the polar lipids of< i> Halobacterium cutirubrum</i>. Formation of isoprenyl ether intermediates. Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism 960, 164–182.

Montellano, P.R.O. de, 1995. Cytochrome P450: Structure, Mechanism, and Biochemistry. Springer.

Montellano, P.R.O. de, 2005. Cytochrome P450: Structure, Mechanism, and Biochemistry. Springer.

Morant, M., Jørgensen, K., Schaller, H., Pinot, F., Møller, B.L., Werck-Reichhart, D., Bak, S., 2007. CYP703 is an ancient cytochrome P450 in land plants catalyzing in-chain

hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. The Plant Cell Online 19, 1473–1487.

Motoyama, A., Yates III, J.R., 2008. Multidimensional LC separations in shotgun proteomics. Analytical chemistry 80, 7187–7193.

Mukherjee, J., Ow, S.Y., Noirel, J., Biggs, C.A., 2011. Quantitative protein expression and cell surface characteristics of Escherichia coli MG1655 biofilms. PROTEOMICS 11, 339–351.

Murby, M., Uhlén, M., Stahl, S., 1996. Upstream strategies to minimize proteolytic degradation upon recombinant production in Escherichia coli. Protein expression and purification 7, 129–136.

Nagarajan, M., Maruthanayagam, V., Sundararaman, M., 2012. SAR analysis and bioactive potentials of freshwater and terrestrial cyanobacterial compounds: a review. Journal of Applied Toxicology.

Nagata, N., Suzuki, M., Yoshida, S., Muranaka, T., 2002. Mevalonic acid partially restores chloroplast and etioplast development in Arabidopsis lacking the non-mevalonate pathway. Planta 216, 345–350.

Naidong, W., 2003. Bioanalytical liquid chromatography tandem mass spectrometry methods on underivatized silica columns with aqueous/organic mobile phases. Journal of Chromatography B 796, 209–224.

Nakashima, K., Kanamaru, K., Mizuno, T., Horikoshi, K., 1996. A novel member of the cspA family of genes that is induced by cold shock in Escherichia coli. Journal of bacteriology 178, 2994–2997.

Nebert, D.W., Nelson, D.R., Feyereisen, R., 1989. Evolution of the cytochrome P450 genes. Xenobiotica 19, 1149–1160.

Neilan, B.A., Dittmann, E., Rouhiainen, L., Bass, R.A., Schaub, V., Sivonen, K., Börner, T., 1999. Nonribosomal peptide synthesis and toxigenicity of cyanobacteria. Journal of bacteriology 181, 4089–4097.

Nelson, D.R., 2006. Plant cytochrome P450s from moss to poplar. Phytochemistry Reviews 5, 193–204.

Nelson, D.R., Kamataki, T., Waxman, D.J., Guengerich, F.P., Estabrook, R.W., Feyereisen, R., Gonzalez, F.J., Coon, M.J., Gunsalus, I.C., Gotoh, O., 1993. The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. DNA and cell biology 12, 1–51.

Nelson, D.R., Ming, R., Alam, M., Schuler, M.A., 2008. Comparison of cytochrome P450 genes from six plant genomes. Tropical Plant Biology 1, 216–235.

Nelson, D.R., Strobel, H.W., 1988. On the membrane topology of vertebrate cytochrome P-450 proteins. Journal of Biological Chemistry 263, 6038–6050.

Newman, D.J., Cragg, G.M., 2012. Natural products as sources of new drugs over the 30 years from 1981 to 2010. Journal of natural products 75, 311–335.

Newman, D.J., Cragg, G.M., Snader, K.M., 2000. The influence of natural products upon drug discovery. Natural product reports 17, 215–234.

Newman, J.D., Marshall, J., Chang, M., Nowroozi, F., Paradise, E., Pitera, D., Newman, K.L., Keasling, J.D., 2006. High-level production of amorpha-4,11-diene in a two-phase partitioning bioreactor of metabolically engineered Escherichia coli. Biotechnology and Bioengineering 95, 684–691.

Nguyen, H.T.T., Nevoigt, E., 2009. Engineering of Saccharomyces cerevisiae for the production of dihydroxyacetone (DHA) from sugars: A proof of concept. Metabolic engineering 11, 335–346.

Nilsson, T., Larsen, T.O., Montanarella, L., Madsen, J.Ø., 1996. Application of head-space solid-phase microextraction for the analysis of volatile metabolites emitted by Penicillium species. Journal of Microbiological Methods 25, 245–255.

Noor, E., Eden, E., Milo, R., Alon, U., 2010. Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. Molecular cell 39, 809–820.

Ogura, K., Koyama, T., 1998. Enzymatic aspects of isoprenoid chain elongation. Chemical reviews 98, 1263–1276.

Olson, E.R., Dunyak, D.S., Jurss, L.M., Poorman, R.A., 1991. Identification and characterization of dppA, an Escherichia coli gene encoding a periplasmic dipeptide transport protein. Journal of bacteriology 173, 234–244.

Omura, T., 1999. Forty years of cytochrome P450. Biochemical and biophysical research communications 266, 690–698.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., Mann, M., 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Molecular & cellular proteomics 1, 376–386.

Orrenius, S., Dallner, G., Ernster, L., 1964. Inhibition of the TPNH-linked lipid peroxidation of liver microsomes by drugs undergoing oxidative demethylation. Biochemical and biophysical research communications 14, 329–334.

Ow, D.S.-W., Lee, D.-Y., Tung, H.-H., Lin-Chao, S., 2009. Plasmid Regulation and Systems-Level Effects on Escherichia coli Metabolism, in: Systems Biology and Biotechnology of Escherichia Coli. Springer, pp. 273–294.

Ow, S.Y., Cardona, T., Taton, A., Magnuson, A., Lindblad, P., Stensjö, K., Wright, P.C., 2008. Quantitative shotgun proteomics of enriched heterocysts from Nostoc sp. PCC 7120 using 8-plex isobaric peptide tags. J. Proteome Res. 7, 1615–1628.

Ow, S.Y., Salim, M., Noirel, J., Evans, C., Rehman, I., Wright, P.C., 2009. iTRAQ underestimation in simple and complex mixtures:"the good, the bad and the ugly". Journal of proteome research 8, 5347–5355.

Ow, S.Y., Salim, M., Noirel, J., Evans, C., Wright, P., 2011. Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation. Proteomics 11, 2341–2346.

Ow, S.Y., Wright, P.C., 2009. Current trends in high throughput proteomics in cyanobacteria. FEBS Lett. 583, 1744–1752.

Page, J.E., Hause, G., Raschke, M., Gao, W., Schmidt, J., Zenk, M.H., Kutchan, T.M., 2004. Functional analysis of the final steps of the 1-deoxy-D-xylulose 5-phosphate (DXP) pathway to isoprenoids in plants using virus-induced gene silencing. Plant physiology 134, 1401–1413.

Pandhal, J., Ow, S.Y., Noirel, J., Wright, P.C., 2011. Improving N-glycosylation efficiency in Escherichia coli using shotgun proteomics, metabolic network analysis, and selective reaction monitoring. Biotechnology and Bioengineering 108, 902–912.

Pandhal, J., Woodruff, L., Jaffe, S., Desai, P., Ow, S.Y., Noirel, J., Gill, R.T., Wright, P.C., 2013. Inverse metabolic engineering to improve Escherichia coli as an N-glycosylation host. Biotechnology and bioengineering.

Papagianni, M., 2012. Recent advances in engineering the central carbon metabolism of industrially important bacteria. Microb Cell Fact 11, 50.

Pauli, H.H., Kutchan, T.M., 1998. Molecular cloning and functional heterologous expression of two alleles encoding (S)-N-methylcoclaurine 3′-hydroxylase (CYP80B1), a new methyl jasmonate-inducible cytochrome P-450-dependent mono-oxygenase of benzylisoquinoline alkaloid biosynthesis. The Plant Journal 13, 793–801.

Payne, A.H., Hales, D.B., 2004. Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. Endocrine reviews 25, 947–970.

Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J., Gygi, S.P., 2003. Evaluation of Multidimensional Chromatography Coupled with Tandem Mass Spectrometry (LC/LC−MS/MS) for Large-Scale Protein Analysis: The Yeast Proteome. J. Proteome Res. 2, 43–50.

Peng, J., Gygi, S.P., 2001. Proteomics: the move to mixtures. Journal of Mass Spectrometry 36, 1083–1091.

Peralta-Yahya, P.P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J.D., Lee, T.S., 2011. Identification and microbial production of a terpene-based advanced biofuel. Nature Communications 2, 483.

Peti, W., Page, R., 2007. Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost. Protein Expression and Purification 51, 1–10.

Pham, A.T., Ichiba, T., Yoshida, W.Y., Scheuer, P.J., Uchida, T., Tanaka, J., Higa, T., 1991. Two marine sesquiterpene thiocyanates. Tetrahedron letters 32, 4843–4846.

Pham, T.K., Roy, S., Noirel, J., Douglas, I., Wright, P.C., Stafford, G.P., 2010. A quantitative proteomic analysis of biofilm adaptation by the periodontal pathogen Tannerella forsythia. Proteomics 10, 3130–3141.

Pichler, P., Köcher, T., Holzmann, J., Mazanek, M., Taus, T., Ammerer, G., Mechtler, K., 2010. Peptide Labeling with Isobaric Tags Yields Higher Identification Rates Using iTRAQ 4-Plex Compared to TMT 6-Plex and iTRAQ 8-Plex on LTQ Orbitrap. Anal. Chem. 82, 6549–6558.

Picotti, P., Bodenmiller, B., Mueller, L.N., Domon, B., Aebersold, R., 2009. Full Dynamic Range Proteome Analysis of S. cerevisiae by Targeted Proteomics. Cell 138, 795–806.

Pietra, F., 1990. A secret world: natural products of marine life. Birkhäuser Verlag.

Pikuleva, I.A., 2006. Cholesterol-metabolizing cytochromes P450. Drug metabolism and disposition 34, 513–520.

Pitera, D.J., Paddon, C.J., Newman, J.D., Keasling, J.D., 2007. Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. Metabolic Engineering 9, 193–207.

Poulos, T.L., 2005. Structural biology of heme monooxygenases. Biochemical and biophysical research communications 338, 337–345.

Poulos, T.L., Finzel, B.C., Howard, A.J., 1987. High-resolution crystal structure of cytochrome P450cam. Journal of molecular biology 195, 687–700.

Pylypenko, O., Schlichting, I., 2004. Structural aspects of ligand binding to and electron transfer in bacterial and fungal P450s. Annual review of biochemistry 73, 991–1018.

Ramaswamy, A.V., Sorrels, C.M., Gerwick, W.H., 2007. Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium Lyngbya majuscula. Journal of natural products 70, 1977–1986.

Redding-Johanson, A.M., Batth, T.S., Chan, R., Krupa, R., Szmidt, H.L., Adams, P.D., Keasling, J.D., Soon Lee, T., Mukhopadhyay, A., Petzold, C.J., 2011. Targeted proteomics for metabolic pathway optimization: application to terpene production. Metabolic engineering 13, 194–203.

Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440, 940–943.

Robert, F.O., Pandhal, J., Wright, P.C., 2010. Exploiting cyanobacterial P450 pathways. Current opinion in microbiology 13, 301–306.

Roberts, S.C., 2007. Production and engineering of terpenoids in plant cell culture. Nature chemical biology 3, 387–395.

Rohdich, F., Hecht, S., Gärtner, K., Adam, P., Krieger, C., Amslinger, S., Arigoni, D., Bacher, A., Eisenreich, W., 2002. Studies on the nonmevalonate terpene biosynthetic pathway: metabolic role of IspH (LytB) protein. Proceedings of the National Academy of Sciences 99, 1158–1163.

Rohmer, M., Knani, M.H., Simonin, P., Sutter, B., Sahm, H., 1993. Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. Biochem. j 295, 517–524.

Rohmer, M., Seemann, M., Horbach, S., Bringer-Meyer, S., Sahm, H., 1996. Glyceraldehyde 3-phosphate and pyruvate as precursors of isoprenic units in an alternative non-mevalonate pathway for terpenoid biosynthesis. Journal of the American Chemical Society 118, 2564–2566.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., 2004. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Molecular & cellular proteomics 3, 1154–1169.

Roussel, F., Khan, K.K., Halpert, J.R., 2000. The importance of SRS-1 residues in catalytic specificity of human cytochrome P450 3A4. Archives of biochemistry and biophysics 374, 269–278.

Rozkov, A., Avignone-Rossa, C. a., Ertl, P. f., Jones, P., O'Kennedy, R. d., Smith, J. j., Dale, J. w., Bushell, M. e., 2004. Characterization of the metabolic burden on Escherichia coli DH1 cells imposed by the presence of a plasmid containing a gene therapy sequence. Biotechnology and Bioengineering 88, 909–915.

Ruiz-Perez, F., Henderson, I.R., Nataro, J.P., 2010. Interaction of FkpA, a peptidyl-prolyl cis/trans isomerase with EspP autotransporter protein. Gut microbes 1, 339–344.

Ruzicka, L., 1953. The isoprene rule and the biogenesis of terpenic compounds. Experientia 9, 357–367.

Rynkiewicz, M.J., Cane, D.E., Christianson, D.W., 2001. Structure of trichodiene synthase from Fusarium sporotrichioides provides mechanistic inferences on the terpene cyclization cascade. Proceedings of the National Academy of Sciences 98, 13543–13548.

S.D Clouse, Sasse, J.., 1998. Brassinosteroids: essential regulators of plant growth and development. Annual Review of Plant Biology 49, 427–451.

Sano, T., 1990. Expression of a Cloned Streptavidin Gene in Escherichia coli. Proceedings of the National Academy of Sciences 87, 142–146.

Santini, C.-L., 1998. A novel Sec-independent periplasmic protein translocation pathway in Escherichia coli. The EMBO Journal 17, 101–112.

Sardessai, Y., Bhosle, S., 2002. Tolerance of bacteria to organic solvents. Research in Microbiology 153, 263–268.

Schenkman, J.B., 1992. Steroid metabolism by constitutive cytochromes P450. The Journal of steroid biochemistry and molecular biology 43, 1023–1030.

Schilmiller, A.L., Schauvinhold, I., Larson, M., Xu, R., Charbonneau, A.L., Schmidt, A., Wilkerson, C., Last, R.L., Pichersky, E., 2009. Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. Proceedings of the National Academy of Sciences 106, 10865–10870.

Schmidt, A., Kellermann, J, Lottspeich, F., 2005. A novel strategy for quantitative proteomics using isotope-coded protein labels. Proteomics 5, 4–15.

Schröder, G., Unterbusch, E., Kaltenbach, M., Schmidt, J., Strack, D., De Luca, V., Schröder, J., 1999. Light-induced cytochrome P450-dependent enzyme in indole alkaloid biosynthesis: tabersonine 16-hydroxylase. FEBS letters 458, 97–102.

Schuler, M.A., Werck-Reichhart, D., 2003. Functional genomics of P450s. Annual review of plant biology 54, 629–667.

Schwab, W., 2003. Metabolome diversity: too few genes, too many metabolites? Phytochemistry 62, 837–849.

Schwarzer, D., Finking, R., Marahiel, M.A., 2003. Nonribosomal peptides: from genes to products. Natural product reports 20, 275–287.

Shishova, E.Y., Di Costanzo, L., Cane, D.E., Christianson, D.W., 2007. X-ray crystal structure of aristolochene synthase from Aspergillus terreus and evolution of templates for the cyclization of farnesyl diphosphate. Biochemistry 46, 1941–1951.

Simmons, T.L., Andrianasolo, E., McPhail, K., Flatt, P., Gerwick, W.H., 2005. Marine natural products as anticancer drugs. Molecular Cancer Therapeutics 4, 333–342.

Smith, C.D., Zhang, X., Mooberry, S.L., Patterson, G.M., Moore, R.E., 1994. Cryptophycin: a new antimicrotubule agent active against drug-resistant cells. Cancer research 54, 3779–3784.

Smolke, C.D., 2009. The Metabolic Pathway Engineering Handbook. CRC Press.

Sørensen, H.P., Mortensen, K.K., 2005. Advanced genetic strategies for recombinant protein expression in Escherichia coli. Journal of Biotechnology 115, 113–128.

Spatzenegger, M., Jaeger, W., 1995. Clinical importance of hepatic cytochrome P450 in drug metabolism. Drug metabolism reviews 27, 397–417.

Stanier, R.Y., Bazine, G.C., 1977. Phototrophic prokaryotes: the cyanobacteria. Annual Reviews in Microbiology 31, 225–274.

Starks, C.M., Back, K., Chappell, J., Noel, J.P., 1997. Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. Science 277, 1815–1820.

Stephanopoulos, G., 1999. Metabolic fluxes and metabolic engineering. Metabolic engineering 1, 1–11.

Steunou, A.-S., Bhaya, D., Bateson, M.M., Melendrez, M.C., Ward, D.M., Brecht, E., Peters, J.W., Kühl, M., Grossman, A.R., 2006. In situ analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats. Proceedings of the National Academy of Sciences of the United States of America 103, 2398–2403.

Stoilov, I., Jansson, I., Sarfarazi, M., Schenkman, J.B., 2001. Roles of cytochrome p450 in development. Drug metabolism and drug interactions 18, 33–56.

Strom, A.R., Kaasen, I., 1993. Trehalose metabolism in Escherichia coli: stress protection and stress regulation of gene expression. Molecular Microbiology 8, 205–210.

Studier, F.W., 2005. Protein production by auto-induction in high-density shaking cultures. Protein expression and purification 41, 207–234.

Stuehr, D.J., Ikeda-Saito, M., 1992. Spectral characterization of brain and macrophage nitric oxide synthases. Cytochrome P-450-like hemeproteins that contain a flavin semiquinone radical. Journal of Biological Chemistry 267, 20547–20550.

Janikowski,T., Velicogna, D., Punt, M., Daugulis, A., 2002. Use of a two-phase partitioning bioreactor for degrading polycyclic aromatic hydrocarbons by a Sphingomonas sp. Applied Microbiology and Biotechnology 59, 368–376.

Takahashi, S., Kuzuyama, T., Watanabe, H., Seto, H., 1998. A 1-deoxy-D-xylulose 5-phosphate reductoisomerase catalyzing the formation of 2-C-methyl-D-erythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis. Proceedings of the National Academy of Sciences 95, 9879–9884.

Tam, R., Saier, M.H., 1993. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. Microbiological reviews 57, 320–346.

Tan, L.T., 2007. Bioactive natural products from marine cyanobacteria for drug discovery. Phytochemistry 68, 954–979.

Tan, W., Lu, J., Huang, M., Li, Y., Chen, M., Wu, G., Gong, J., Zhong, Z., Xu, Z., Dang, Y., 2011. Anti-cancer natural products isolated from chinese medicinal herbs. Chin Med 6, 27.

Tao, L., Zhou, L., Zheng, L., Yao, M., 2006. Elemene displays anti-cancer ability on laryngeal cancer cells in vitro and in vivo. Cancer chemotherapy and pharmacology 58, 24–34.

Ter Kuile, B.H., Westerhoff, H.V., 2001. Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. FEBS letters 500, 169–171.

Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Hamon, C., 2003. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Analytical chemistry 75, 1895–1904.

Toone, W.M., Rudd, K.E., Friesen, J.D., 1991. deaD, a new Escherichia coli gene encoding a presumed ATP-dependent RNA helicase, can suppress a mutation in rpsB, the gene encoding ribosomal protein S2. Journal of bacteriology 173, 3291–3302.

Topcu, Z., 2000. An optimized recipe for cloning of the polymerase chain reaction-amplified DNA inserts into plasmid vectors. ACTA BIOCHIMICA POLONICA-ENGLISH EDITION- 47, 841–846.

Trantas, E., Panopoulos, N., Ververidis, F., 2009. Metabolic engineering of the complete pathway leading to heterologous biosynthesis of various flavonoids and stilbenoids in Saccharomyces cerevisiae. Metabolic engineering 11, 355–366.

Trapp, S.C., Croteau, R.B., 2001. Genomic organization of plant terpene synthases and molecular evolutionary implications. Genetics 158, 811–832.

Trimurtulu, G., Ohtani, I., Patterson, G.M., Moore, R.E., Corbett, T.H., Valeriote, F.A., Demchik, L., 1994. Total structures of cryptophycins, potent antitumor depsipeptides from the blue-green alga Nostoc sp. strain GSV 224. Journal of the American Chemical Society 116, 4729–4737.

Tsai, C.-C., Emau, P., Jiang, Y., Agy, M.B., Shattock, R.J., Schmidt, A., Morton, W.R., Gustafson, K.R., Boyd, M.R., 2004. Cyanovirin-N inhibits AIDS virus infections in vaginal transmission models. AIDS research and human retroviruses 20, 11–18.

Tyo, K.E., Alper, H.S., Stephanopoulos, G.N., 2007. Expanding the metabolic engineering toolbox: more options to engineer cells. Trends in biotechnology 25, 132–137.

Udeshi, N.D., Mani, D.R., Eisenhaure, T., Mertins, P., Jaffe, J.D., Clauser, K.R., Hacohen, N., Carr, S.A., 2012. Methods for quantification of in vivo changes in protein ubiquitination following proteasome and deubiquitinase inhibition. Molecular & Cellular Proteomics 11, 148–159.

Unterlinner, B., Lenz, R., Kutchan, T.M., 1999. Molecular cloning and functional expression of codeinone reductase: the penultimate enzyme in morphine biosynthesis in the opium poppy Papaver somniferum. The Plant Journal 18, 465–475.

Urbina, J.A., Moreno, B., Vierkotter, S., Oldfield, E., Payares, G., Sanoja, C., Bailey, B.N., Yan, W., Scott, D.A., Moreno, S.N.J., Docampo, R., 1999. Trypanosoma cruzi Contains Major Pyrophosphate Stores, and Its Growth in Vitro and in Vivo Is Blocked by Pyrophosphate Analogs. J. Biol. Chem. 274, 33609–33615.

Van der Werf, M.J., Jellema, R.H., Hankemeier, T., 2005. Microbial metabolomics: replacing trial-and-error by the unbiased selection and ranking of targets. Journal of Industrial Microbiology and Biotechnology 32, 234–252.

Voegele, R.T., Sweet, G.D., Boos, W., 1993. Glycerol kinase of Escherichia coli is activated by interaction with the glycerol facilitator. Journal of bacteriology 175, 1087–1094.

Vollmer, W., Blanot, D., De Pedro, M.A., 2008. Peptidoglycan structure and architecture. FEMS microbiology reviews 32, 149–167.

Walker, K., Croteau, R., 2001. Taxol biosynthetic genes. Phytochemistry 58, 1–7.

Walsh, C.T., 2004. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. Science 303, 1805–1810.

Wang, C., Yoon, S.-H., Jang, H.-J., Chung, Y.-R., Kim, J.-Y., Choi, E.-S., Kim, S.-W., 2011. Metabolic engineering of Escherichia coli for α-farnesene production. Metabolic engineering 13, 648–655.

Wang, C., Yoon, S.-H., Shah, A.A., Chung, Y.-R., Kim, J.-Y., Choi, E.-S., Keasling, J.D., Kim, S.-W., 2010. Farnesol production from Escherichia coli by harnessing the exogenous mevalonate pathway. Biotechnology and bioengineering 107, 421–429.

Wang, G., Li, X., Huang, F., Zhao, J., Ding, H., Cunningham, C., Coad, J.E., Flynn, D.C., Reed, E., Li, Q.Q., 2005. Antitumor effect of β-elemene in non-small-cell lung cancer cells is mediated via induction of cell cycle arrest and apoptotic cell death. Cellular and molecular life sciences 62, 881–893.

Ward, D.M., Ferris, M.J., Nold, S.C., Bateson, M.M., 1998. A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiology and Molecular Biology Reviews 62, 1353–1370.

Wase, N.V., Wright, P.C., 2008. Systems biology of cyanobacterial secondary metabolite production and its role in drug discovery.

Washburn, M.P., Wolters, D., Yates, J.R., 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nature biotechnology 19, 242–247.

Waterman, M.R., 1992. Cytochrome P450: cellular distribution and structural considerations. Current Opinion in Structural Biology 2, 384–387.

Watts, K.T., Lee, P.C., Schmidt-Dannert, C., 2004. Exploring recombinant flavonoid biosynthesis in metabolically engineered Escherichia coli. Chembiochem 5, 500–507.

Welker, M., Von Döhren, H., 2006. Cyanobacterial peptides–nature's own combinatorial biosynthesis. FEMS microbiology reviews 30, 530–563.

Wessjohann, L., 1994. The first total syntheses of taxol. Angewandte Chemie International Edition in English 33, 959–961.

White, N.J., 2008. Qinghaosu (Artemisinin): The Price of Success. Science 320, 330–334.

Wildung, M.R., Croteau, R., 1996. A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis. Journal of Biological Chemistry 271, 9201–9204.

Wilkin, D.J., Kutsunai, S.Y., Edwards, P.A., 1990. Isolation and sequence of the human farnesyl pyrophosphate synthetase cDNA. Coordinate regulation of the mRNAs for farnesyl pyrophosphate synthetase, 3-hydroxy-3-methylglutaryl coenzyme A reductase, and 3-hydroxy-3-methylglutaryl coenzyme A synthase by phorbol ester. J. Biol. Chem. 265, 4607–4614.

Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., 1996. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and arnino acid analysis. Nature Biotechnology 14, 61–65.

Williams, J.A., Martin, F.L., Muir, G.H., Hewer, A., Grover, P.L., Phillips, D.H., 2000. Metabolic activation of carcinogens and expression of various cytochromes P450 in human prostate tissue. Carcinogenesis 21, 1683–1689.

Withers, S.T., Keasling, J.D., 2007. Biosynthesis and engineering of isoprenoid small molecules. Applied microbiology and biotechnology 73, 980–990.

Woese, C.R., 1987. Bacterial evolution. Microbiological reviews 51, 221.

Xiang, S., Usunow, G., Lange, G., Busch, M., Tong, L., 2013. 1-Deoxy-d-Xylulose 5-Phosphate Synthase (DXS), a Crucial Enzyme for Isoprenoids Biosynthesis, in: Isoprenoid Synthesis in Plants and Microorganisms. Springer, pp. 17–28.

Xu, J., Johnson, R.C., 1995. aldB, an RpoS-dependent gene in Escherichia coli encoding an aldehyde dehydrogenase that is repressed by Fis and activated by Crp. Journal of bacteriology 177, 3166–3175.

Xu, L.-H., Fushinobu, S., Takamatsu, S., Wakagi, T., Ikeda, H., Shoun, H., 2010. Regio-and stereospecificity of filipin hydroxylation sites revealed by crystal structures of cytochrome P450 105P1 and 105D6 from Streptomyces avermitilis. Journal of Biological Chemistry 285, 16844–16853.

Yamamoto, H., Miyake, C., Dietz, K.-J., Tomizawa, K.-I., Murata, N., Yokota, A., 1999. Thioredoxin peroxidase in the Cyanobacterium Synechocystis sp. PCC 6803. FEBS letters 447, 269–273.

Yu, F., Utsumi, R., 2009. Diversity, regulation, and genetic manipulation of plant mono- and sesquiterpenoid biosynthesis. Cellular and Molecular Life Sciences 66, 3043–3052.

Yu, Z., Wang, R., Xu, L., Xie, S., Dong, J., Jing, Y., 2011. β-Elemene piperazine derivatives induce apoptosis in human leukemia cells through downregulation of c-FLIP and generation of ROS. PloS one 6, e15843.

Zambonin, C.G., Quinto, M., De Vietro, N., Palmisano, F., 2004. Solid-phase microextraction – gas chromatography mass spectrometry: A fast and simple screening method for the assessment of organophosphorus pesticides residues in wine and fruit juices. Food Chemistry 86, 269–274.

Zerbe, K., Pylypenko, O., Vitali, F., Zhang, W., Rouset, S., Heck, M., Vrijbloed, J.W., Bischoff, D., Bister, B., Süssmuth, R.D., 2002. Crystal structure of OxyB, a cytochrome P450 implicated in an oxidative phenol coupling reaction during vancomycin biosynthesis. Journal of Biological Chemistry 277, 47476–47485.

Zhang, Z.-Y., Zhou, G., Denu, J.M., Wu, L., Tang, X., Mondesert, O., Russell, P., Butch, E., Guan, K.-L., 1995. Purification and characterization of the low molecular weight protein tyrosine phosphatase, Stp1, from the fission yeast Schizosaccharomyces pombe. Biochemistry 34, 10560–10568.

Zhou, S., Causey, T.B., Hasona, A., Shanmugam, K.T., Ingram, L.O., 2003. Production of Optically Pure d-Lactic Acid in Mineral Salts Medium by Metabolically Engineered Escherichia coli W3110. Appl. Environ. Microbiol. 69, 399–407.

Zhu, B.-Y., Mant, C.T., Hodges, R.S., 1991. Hydrophilic-interaction chromatography of peptides on hydrophilic and strong cation-exchange columns. Journal of Chromatography A 548, 13–24.
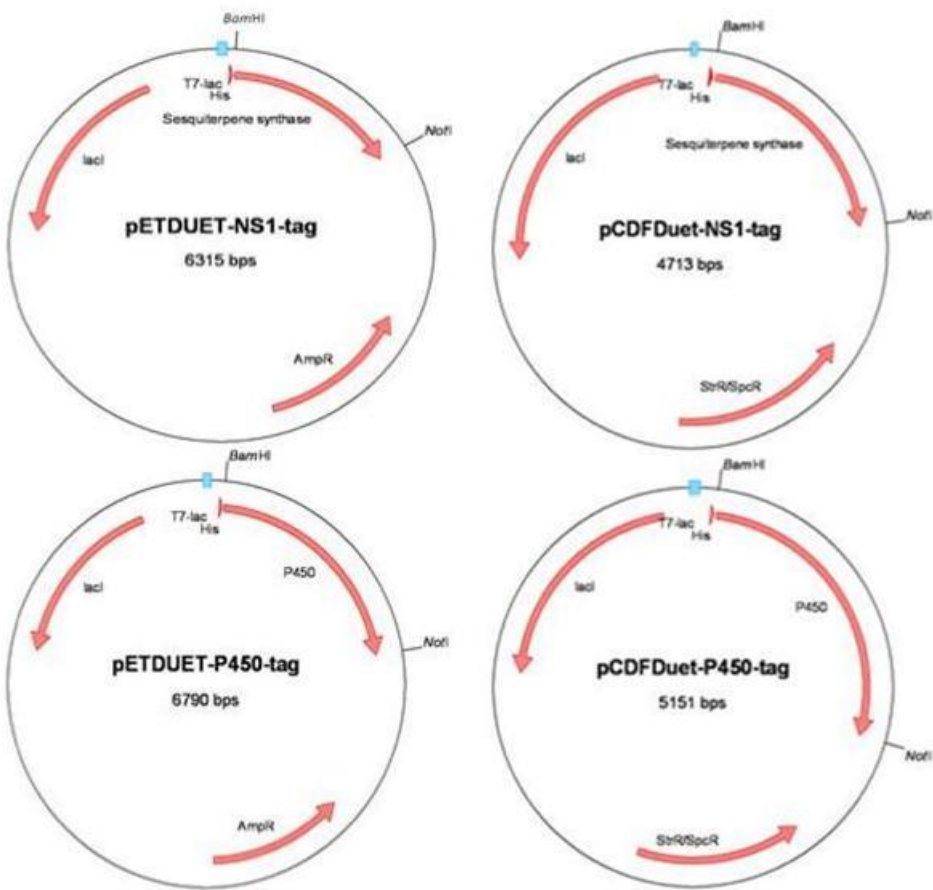
Appendices

Fig A1. Annotated final plasmid maps of pCDFDuet and pETDuet with sesquiterpene synthase and cytochrome P450 genes in-frame with the N-terminal hexa-histidne sequence
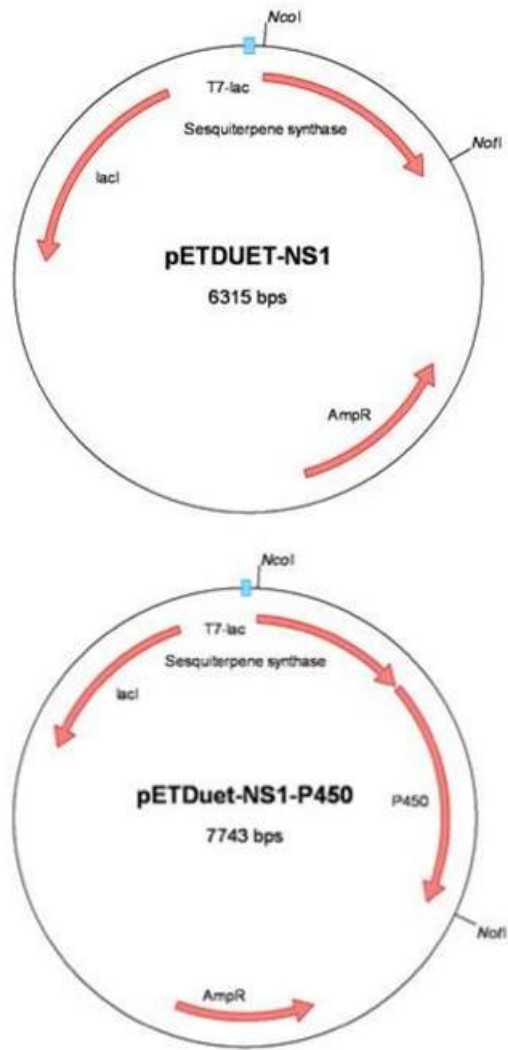
Fig A2. Annotated final plasmid maps of pETDuet containing insert (NS1 and NS1-P450) and cloned into *NcoI* and *NotI* of MCS1

Fig A3. DNA sequence chromatogram of pCDFDuet-NStag showing the N-terminal 6-histidine sequence tag and the ATG start site of the gene.
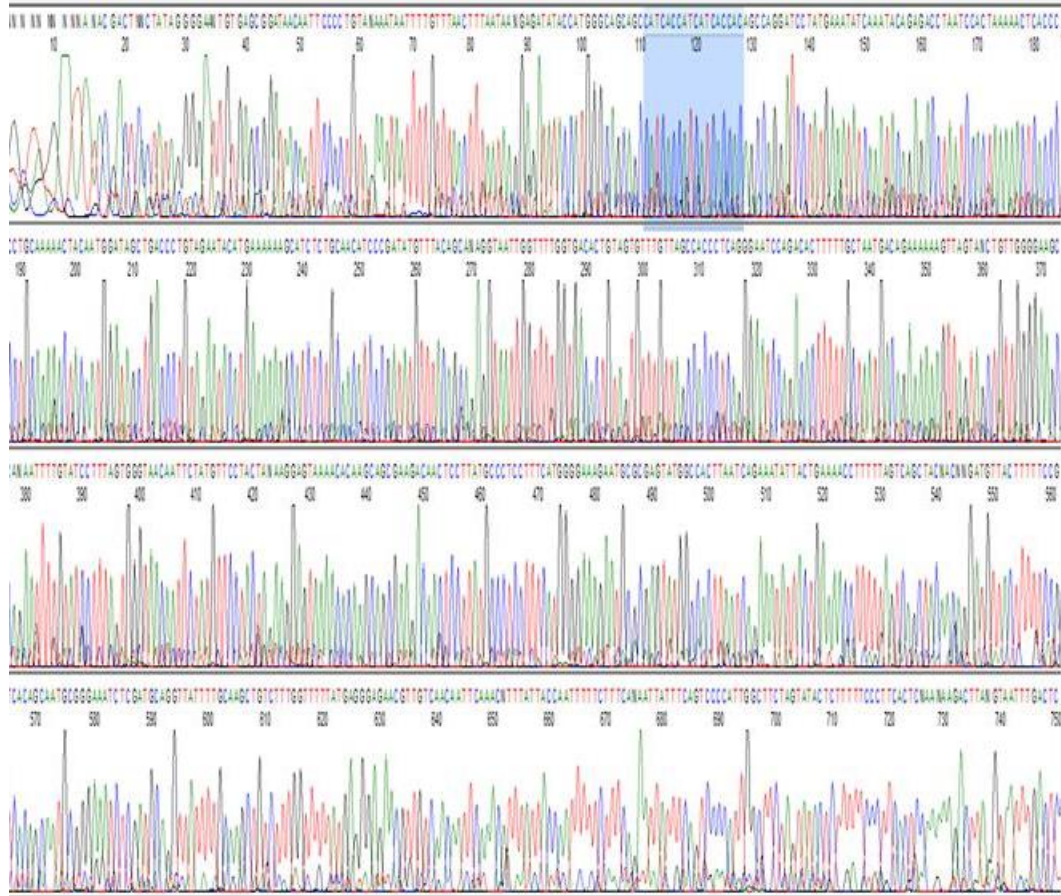
Fig A4. DNA sequence chromatogram of pCDFDuet-P450tag showing the N-terminal 6-histidine sequence tag and the ATG start site of the gene.