Investigation of recombinant protein production by *Escherichia coli*: Expression of Green fluorescent protein and a co-factor dependent flavinated enzyme.

A PhD project in collaboration with GSK

By Paul Hudman

I, Paul Andrew Hudman confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed_____

Abstract

This thesis summarises work done on the *Escherichia coli* strain MG1655 expressing a Green Fluorescent Protein (GFP) and the flavo-protein N-methyl-L-tryptophan oxidase (MTOX) product and examining the effect foreign protein production has on cell growth parameters. It also uses molecular modelling tools to generate data relating to FAD flux and MTOX production, comparable to that seen in *E.coli* fermentations.

The MG1655 strain was chosen as it was the focus of the first K-12 complete sequencing project and closely related to the strain W3110, a second K strain that had been used to develop a number of deletion mutants which were central to the study.

It presents data from shake flask and stirred tank reactor fermentations on minimal, carbon limited and complex media. Samples from these growth experiments were then analysed concentrating on biomass concentration, protein assays (both chemical and fluorimetric), high performance liquid chromatography and calculation of yield parameters. From this a baseline of growth was established with which to compare changes in growth after a shift in protein product from GFP to MTOX. An assay was also developed to measure the amount of active and inactive MTOX enzyme produced and this data compared with the level of FAD available to the cell at specific time points throughout growth.

Finally modelling work is presented and *in silico* values compared with those generated *in vitro*. A discussion of the entire study concludes the work.

Table of Contents

Table of Figures

Abbreviations

ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
BHK21	Baby hamster kidney cells
BSA	Bovine Serum albumin
BVMO	Baeyer – Villiger monooxygenase
CER	Carbon Dioxide evolution rate
CHO	Chinese Hamster Overy
d.f.	Degrees of freedom
DCW	Dry cell weight
DNA	Deoxyribonucleic acid
%CV	Coefficient of variation
DOT	Dissolved oxygen tension
E.coli	*Escherischia coli*
EBA	Energy balance analysis
FAD	Flavin adenine dinucleotide
FBA	Flux balance analysis
FDA	U.S. Food and Drug administration
FMN	Flavin mononucleotide

GFP	Green Fluorescent protein
GSK	GlaxoSmithKline
HPLC	High-performance liquid chromatography
IPTG	Isopropyl-beta-D-thiogalactopyranoside
kD	kilo Daltons
kHz	kilo Hertz
LB	Luria-Bertani (broth)
MCB	Master cell bank
MCS	Multiple cloning site
MSOX	Monomeric sarcosine oxidase
MTOX	N-methyl-L-tryptophan oxidase
N.d.	Not determined
OD	Optical density
OUR	Oxygen uptake rate
PCR	Polymerase chain reaction
PPG	Polypropylene glycol
QPCR	Quantitative polymerase chain reaction
R^2	Coefficient of determination
RFU	Relative fluorescence units
RO	Reverse osmosis
rpm	revolutions per minute
RQ	Respiratory quotient
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SF	Shake flask
μMAX	Maximum growth rate
STR	Stirred tank reactor
vvm	volume per volume per minute
WCB	Working cell bank
WCW	Wet cell weight
WT	Wild type
Xmax	Maximum biomass concentration
Y ac/x	Yield of acetate over biomass
Y x/glu	Yield of biomass over glucose

Acknowledgements

I would like to thank Dr Frank Baganz and Prof. John Ward for their input and support through the course of this work.

I would also like to express thanks to GlaxoSmithKline Research and Development, Stevenage, UK and GlaxoSmithKline Global Manufacturing and Supply, Ulverston, UK for support of this work via a fully funded CASE studentship.

Finally, special thanks to Lindsay and my family for their continuous support and encouragement throughout the research and writing of this work.

Paul Hudman

1. Introduction

1.1. Project scope

Escherichia coli is one of the most important hosts in modern day recombinant protein production. Throughout academia and industry its uses are widespread and with sequence data available for some of the most common strains of the bacteria it has been a favourite organism for many metabolic engineering and metabolic modelling projects in the past (Berry, 1996; Cameron, et al., 1997; Koffas, et al., 1999). The aim of this work was to investigate cofactor supply in an *E.coli* expression system producing the recombinant flavo – protein, N-methyl tryptophan oxidase (MTOX) and understand how strain alterations may affect the supply of FAD required for the activity of the protein product. A number of objectives were set throughout the course of the work which are stated in section 1.9 and discussed throughout the thesis. A discussion of all findings is presented in Chapter 7.

Although there are drawbacks to using a prokaryotic expression system, such as the propensity for expressed protein to form insoluble inclusion bodies (González-Montalbán, 2005), and even these inclusion bodies to cause cytotoxic effects in the host, dramatically reducing growth rate, (Carrio, 2006) they are still widely in use. In some cases the presence of these inclusion bodies can aid the downstream process by being far simpler to remove from the final cell harvest than proteins soluble in the cell. This, the ease of cultivation and the availability of well-established genetic manipulation techniques are often seen to outweigh the disadvantages.

The following work therefore outlines the construction and characterisation of a model system for the investigation of recombinant protein production in *E.coli*. It initially uses a plasmid construct expressing Green Fluorescent Protein (GFP) to act as a recombinant protein product with the benefits of being easy to visualise and measure as described in Chapter 2 (section 2.4.4) and also being designed to be inducible with IPTG.

Therefore the first results chapter describes the characterisation of this strain/plasmid choice focusing on medium composition and comparisons between different minimal and complex media based on growth parameters including carbon uptake and acetate production in addition to the production of the recombinant protein, GFP. These parameters are investigated with the aim of establishing a baseline for bacterial growth and protein production that can be used in comparative studies with mutant *E.coli* strains or novel medium/regime changes.

The second results chapter looks to build on the first but with the introduction of N-Methyltryptophan oxidase (MTOX), a 42 kDa flavinated protein found in a range of organisms (Job et al., 2002) but with, as yet, no known function in *E.coli*. This decision was directed by the potential industrial applications, (discussed in section 1.6) of flavoproteins compared to GFP.

An assay method was developed to measure the levels of both active and inactive enzyme (Hassan – Abdallah, 2005; Koyama, 1996) and applied to measure MTOX activity in samples taken at various points during culture growth. Also, an HPLC method was used

to measure levels of FAD throughout the fermentation (Barile, 1997) in an attempt to relate that information to measured MTOX activity.

The final results chapter concentrates on the use and investigation of a metabolic model consisting of a reaction network developed originally by Edwards and Palsson (2000). The construction of metabolic networks and their use in metabolic engineering projects has been the focus of many research projects in recent years. Here data generated concerning FAD and MTOX flux and biomass production is compared to data generated from *in vivo* experiments. Similarities are seen, and explanations put forward where the model fails to accurately predict the experimental data.

As a whole the experimental work sought to establish a model system for the study of recombinant protein production in, *E.coli* strains, evaluated the production of a flavoprotein and looked at the opportunities that allowed for rational metabolic engineering and also examined the possibility of *in silico* work to guide future investigations and understand the limitations of constraint- based metabolic modeling.

1.2. Recombinant protein expression

Over recent years there has been an explosion of interest in the use of micro-organisms and other cell-based systems for the production of biological products and small molecules. The list of recombinant proteins now routinely used in medicine is large and

seemingly ever growing. With more complex post-translational modifications compounds such as Kogenate® , a recombinant Factor VIII produced in BHK21 cells (Arnold, 1995), and Epogen®, an erythropoietin used for the treatment of anaemia caused by chronic kidney disease, produced in CHO cells (Yoon, 2001) are two of the many recombinant proteins produced in mammalian cell culture. Engerix B and Gene-Vac B (both recombinant Hepatitis B vaccines produced by GlaxoSmithKline (GSK) and the Institute of India respectively) and also a number of recombinant follicle stimulating hormones are produced in yeasts, (Shivananda, 2006).

However, some of the most commonly used recombinant proteins, Humulin® and Humatrope® (Eli Lilly and Co.), forms of human insulin for the treatment of diabetics (FDA, 2003) and growth hormone (IGF-1), (Chen et al., 1995 and Cui et al., 2001) are produced in *E. coli.*. In fact *E.coli* is one of the most important production hosts in diabetic medicine with Pfizer producing an inhaled form of recombinant insulin known as "Exubera" and Sanofi Avensis producing a rapid acting insulin analoge using *E.coli* known as "Lantus".

Other examples can be cited such as the work done by Báez-Viveros et. al., (2004) using bacteria as a tool for the production of aromatic amino acids such as phenylalanine but these will not be discussed here.

Some examples of this type of work often cite low yields, poor scalability, or simply insufficient commercial viability as problems facing the technology, (Frederick, et al.

2004; Hatfield, 2007); this can be seen in some ways as the root of this investigation. A better understanding of the relationships between the production of recombinant proteins, the host cell metabolism and related co-factors used for the activation of a protein product was a key aim throughout the entire project.

Initial plans were to look at the strain MG1655 as it was also chosen for sequencing by Blattner's group (Blattner et al., 1997) and is closely related to W3110, another K strain that has been used to develop deletion mutants in the *E. coli* genome, which were available for research (Mori et al., 2000). Certain mutants from this related strain could then be used to investigate how specific gene deletions altered carbon flux and other metabolic changes such as redox metabolism, to ultimately influence the production of active recombinant protein in the host.

Understandably the field of recombinant protein production using many different hosts has been, and continues to be well researched. There are many avenues research has taken in recent years, which can be broadly grouped into 3 areas.

Firstly much work has been done with very high – throughput or scale down methods, often based in 96 well plates. These have concentrated on vector/host interactions measuring growth parameters and protein production in a number of media, with numerous host organisms and differing recombinant protein vectors (e.g. Hewitt & McDonnell, 2004 and Braud, et al., 2005). Some others have set up protein arrays, investigating a smaller number of host/medium combinations but over a much larger

group of possible recombinant proteins (Benita, et al., 2006). Scale down has been used to look more specifically at certain recombinant proteins and has attempted to understand methods for the optimization of protein production often in lab- scale continuous cultures, (e.g. Chae, et al., 2000 and Mahadevan & Doyle III 2003). Based on this work much research looks to scale up successful small – scale cultures often using modelling techniques to determine the behaviour of a culture at differing scales. This is obviously relevant for the commercial production of recombinant proteins but also acts as a useful tool to investigate and effects of differing oxygen and mass transfer at larger scales (Hatfield & Roth, 2007)

Secondly, alterations in the host strains have been investigated for their impact on recombinant protein production. Although ease of the down-stream process is often not the concern of the fermentation scientist, there are a number of researchers interested in improving the expression of soluable recombinant proteins or conversely the ease of separation and activity of proteins forming inclusion bodies. Some workers have specifically looked into the up-regulation of chaperone proteins increasing the solubility and secretion of proteins and thus allowing less complex purification methods to be employed (Petersson, et al., 2004 and de Marco, 2007). Allied somewhat to this there are many groups investigating the folding of recombinant proteins expressed in bacteria, and even the possibility of glycosylation of protein products in bacterial hosts, something that has traditionally been seen as only easily achievable in mammalian cell culture (Baneyx, & Mujacic, 2004).

Finally the up-regulation of metabolic pathways to overproduce certain proteins or molecules is well researched. Where one protein of interest exists there may be scope for its overproduction or the overproduction of its precursors/cofactors that would improve yields and activity. The use of metabolic engineering for the overproduction of naturally occurring compounds such as riboflavin and folate (Burgess, et al., 2004) or L- valine (Kang, et al., 2007) is well documented.

In context, the following work sought to design a semi-high throughput approach to determine the effect of plasmid and host cell on protein production in a number of *E. coli* strains. This needed to be robust enough to allow changes of medium, growth regime, plasmid construct and strain whilst generating comparable results. The initial model system was designed to express GFP under the lac promoter system control. As shown above, GFP has long been used to construct fusion proteins with less easily detected protein products and has been used successfully as an on line measure of their presence (Chae, et al., 2000). In the model system described in Chapter 4, GFP was used as the recombinant protein of interest.

This system was altered in Chapter 5, expressing the solA gene product, *N*-methyl tryptophan oxidase (MTOX). The interest generally in flavoproteins and the reactions they catalyse is increasing and will be discussed later in this chapter (section 1.6) but with their wide range of reaction mechanisms and their dependence on the cofactor FAD for activity they made an interesting target for study.

Finally the work sought to use flux balance analysis (FBA) to investigate the relationship between carbon uptake rate and biomass accumulation, FAD flux and MTOX production. With this information a window of operation of active MTOX production was generated with the solution space defining the likely levels of all three variables assuming a constant carbon uptake rate. This data was compared with the experimental values in an attempt to see if modelling could be used to complement experimental work when looking at recombinant protein production.

1.3. *E.coli* as a production host and other alternatives

As mentioned in section 1.2 the growing number of therapeutic recombinant protein products utilise a variety of host systems. Which one is used often depends on a number of decisions including ease of culture, ease of purification of the final product, and what post-translational modifications are necessary for an active product.

A list of the most commonly commercialised recombinant proteins shows a bias to a few expression sytems, namely, *E. coli, Saccharomyces cerevisiae,* murine cells, CHO cells or BHK21 cells, (Walsh, 2002) however there are many more available (Merten, et al., 2001). Yeast species are being more widely used such as *Kluyveromyces lactis* and *Hansenula polymorpha* (Gad, 2007) over and above the common use of *Pichia pastoris* and *S. cerevisiae*, (Steinborn, et al., 2006) and other Gram positive and psychrophilic bacteria are becoming more widely used (Yang, et al., 2006 and Cusano, et al., 2006). There are also some cell free expression systems but these remain less well utilised,

mainly due to their cost compared to more commonly used expression systems (Endo, & Sawsaki, 2004).

E. coli remains one of the most important expression systems in use currently. Through Blattner's work (Blattner, et al., 1997) the full sequence of the K-12 strain MG1655 is known and with a long history of use in biotechnology the range of information and molecular biology tools available to the researcher is vast.

Drawbacks commonly associated with the *E.coli* expression system have been the formation of insoluble inclusion bodies, (Ferrer-Miralles, et al., 2009) of recombinant protein expressed at high concentrations and the inability of prokaryotic cells to perform some post-translational modifications such as glycosylations, (Ferrer-Miralles, et al., 2009). For these reasons mammalian cell culture has been used for many complex protein products; however these cultures are much slower growing requiring more specialist conditions and supplements than bacteria so are not ideal. Recently however work has been done to engineer *E.coli* giving it the ability to correctly fold and glycosylate a number of proteins, causing one researcher to state, "…guarantee that *E. coli* will remain an important host for the production of both commodity and high value added proteins.", (Baneyx, & Mujacic, 2004).

1.4. Green Fluorescent Protein as a model system

A number of the previously mentioned studies have investigated the production of recombinant proteins where no practical assay for the direct measurement of protein concentration or enzyme activity exists. In these cases it is common for a reporter gene to be spliced into the vector used to act as a signal when protein expression, of the reporter and also the recombinant target has occurred.

Green Fluorescent Protein has been used extensively in this role (Chae, et al., 2000) but is used in work described here as a product itself. The ease of measurement and availability of commercial plasmid constructs made this an ideal protein to base the model system on for the initial section of study.

GFP was isolated from the jellyfish *Aequorea aequorea* in 1962 (Shimomura et. al., 1962) where it was found as a companion protein to aequorin, the well-known chemiluminescent protein of the same species. It was noticed that living *A. aequorea* tissue had an emission spectrum peaking at 508nm and looking green but pure aequorin peaked in the blue range, at 470nm (Tsien, 1998). This then led Shimomura's group to discover GFP and suggest radiation-less energy transfer as the mechanism for exciting the protein.

Its structure has been determined to consist of an 11 stranded β-barrel containing the chromophore made up of a single α helix as shown in Figure 1.1.

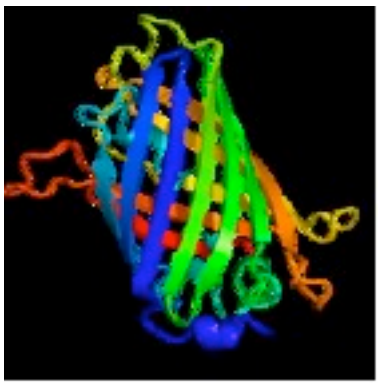


Figure 1.1) Protein structure of GFP (Source: Protein data bank)

Its use as a tool in molecular biology was not realised until 1992 when Prasher reported the cloning and sequence of GFP (Prasher, et al., 1992). Since 1994 GFP has been used as a reporter protein (Chalfie et.al., 1994) flagging its own presence and therefore also proteins under the same control, by emitting green light (λ_{em} = 508 nm) upon excitation with near ultraviolet light (around 395 nm) or blue light (around 470 nm) (Ito et.al, 1999). Since then many mutations have been developed looking to improve the emission or to focus it to a single wavelength (Heim, et al., 1995) or to change the colour of the emitted light itself. Since the first mutations, blue, cyan and yellow fluorescent proteins have been developed as Figure 1. 2 displays.

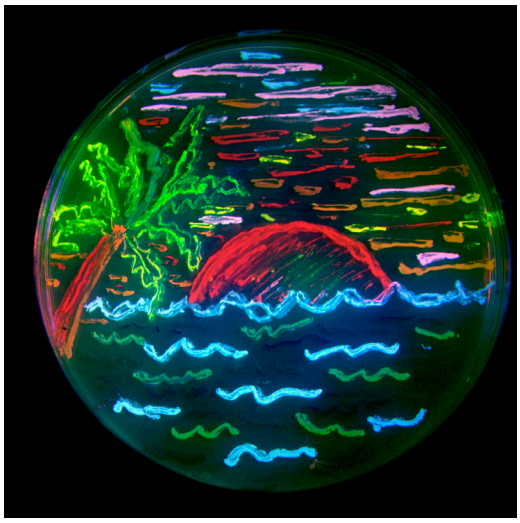


Figure 1.2) San Diego beach scene created using various GFPs and the red-fluorescent coral protein dsRED (Source: Tsien laboratory, University of California)

It was planned to use GFP not as a reporter as in many studies, signalling if a protein of interest was present or absent, or to determine a protein's cellular location, but as a model protein product, easily detected, that could be studied with differing genomic backgrounds. GFP has been used in this way before and can be used as an online measure of recombinant protein production in batch fermentations (Chae, et al., 2000 and Chalfie, et al., 1994).

A system was set up for recombinant protein expression, which will be described in Chapter 4. This also established a baseline for MG1655 growth parameters in the wild type and the plasmid carrying strain and allowed methods that would be used throughout the project to be developed.

The plasmid construct was such that the GFP gene could be excised and the solA gene, (coding for the MTOX protein) inserted to allow for the continuation of the work into the

flavoprotein MTOX. The rational behind this planned change of recombinant protein is detailed below.

1.5. The Keio collection: construction, history, and significance to the work

An *E.coli* K-12 functional genomics project was initiated in Japan in the late 80s to provide further information into the organisation and control of the K-12 genome and to look at potentially developing a set of gene deletions of non essential genes and to attempt to determine the function of a number of unknown genes. The Kohara clones, a clone library describing the physical map of the entire *E.coli* genome, had been developed (Kohara, 1987) and after the publication of the K-12 sequence in 1997, (Blattner, et al., 1997) and the W3110 sequence in 2003 the Keio collection was constructed in 2005 (Baba, et al. 2006).

The Keio collection contains a set of over 2000 precisely defined single gene deletions, of non-essential genes, in the *E.coli* strain BW25113 (a close relative to MG1655 and W3110). This was designed to provide a basic resource for systematic functional genomics but also as a data source for systems biology approaches and as such, allowed a number of gene targets to be chosen for investigation.

The plan was to decide on a number of deletions that should affect metabolism, initially central metabolism but moving to riboflavin synthesis for the later work, and source the BW25113 mutant strains from Reading University (the UK depository for the Japanese

collection). After the development of the plasmid construct in MG1655 it would be a simple task to transform the BW25113 "wild type" and mutant strains to be able to investigate changes in their growth or protein production parameters.

During the course of the initial work it became clear that the likelihood of single gene deletions fundamentally affecting the production of a recombinant protein in a bacterial cell was small. It has been shown that *E. coli* is relatively tolerant to the necessary changes in metabolic flux required to overexpress a certain desired product, e.g. pyruvate (Causey et.al. 2004), and it has a remarkable level of plasticity when altering the fluxes through what has been referred to as the high flux backbone (Almaas et.al., 2004). This finding suggest that only large scale, multiple deletions will have fundamental consequences on cellular carbon flux, whereas the effect of single gene deletions in central metabolism will be very small. It also has been shown that *E. coli* exhibits a high level of genetic epistasis and phenotypic plasticity which may well mask any effect of any single gene alterations in central metabolism (Remold and Lenski, 2004). Phenotypic plasticity is defined as an organisms ability to activate different phenotypes through the selective expression of specific genes in response to changes in the environment, (Agarwal, 2001). Genetic epistasis is an interaction between genes of an organism, in such that an effect on one locus can be altered or masked by effects on another, (Cordell, 2002). This was thought to be a potential issue in that single gene deletion mutants may behave in very similar ways to the wild type organism simply because of the input of these two systems. As discussed in Section 1.7, it was thought that a number of deletion mutants in the riboflavin pathway may have a significant effect on carbon flux to make a

difference in the levels of riboflavin, FMN and FAD. This rationale was then the focus of mutant choice from the collection (see section 5.7 for details).

1.6. MTOX: a cofactor dependent flavo protein.

With this, more directed mutation approach in mind rather than searching for single, non-lethal, gene deletions in central metabolism, and after discussions with GSK, the focus altered to another protein target, namely N-methyl tryptophan oxidase (MTOX).

Active flavoenzymes have a flavin group covalently bound to them that alters the characteristics of their active site and conferring enzymatic activity (Hassan-Abdallah, 2005). The suggestion was therefore, single gene deletions in the riboflavin pathway might have an effect on, not the expression of MTOX but the level of activity if FAD is in short supply or in excess. The fact that the riboflavin pathway is much less complex than central metabolism, although obviously in need of precursors from these pathways, made this goal seem attainable and also gave a much more defined system of reactions for application of *in silico* techniques such as flux balance analysis, covered in chapter 6.

MTOX is a 42 kDa flavoprotein coded for by the gene solA (Swiss-Prot accession number P40874) with, as yet, unknown physiological function in *E. coli*. It catalyses the oxidative demethylation of N-methyl amino acids using oxygen as an electron acceptor (see Fig. 1.3). It is this reaction that allowed the use of the assay technique detailed in section 2.4.5 with the demethylation of N-methyl-L-tryptophan producing, among other compounds, hydrogen peroxide which can be detected in a colourimetric assay.

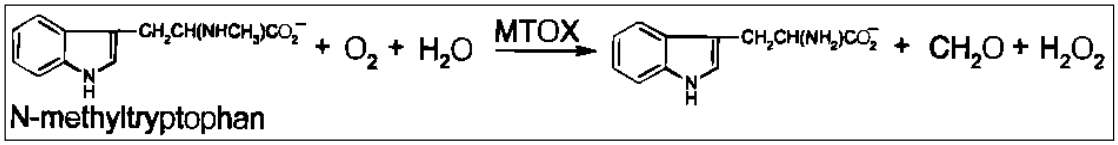


Figure 1.3) Reaction scheme catalysed by MTOX with the production of hydrogen peroxide

One reason for using MTOX was that the group of flavinated proteins is large and possesses a number of possible commercial uses. There is obviously no commercial gain in increasing the level of Green Fluorescent Protein in a bacterial fermentation; however it provides an ideal model system to investigate metabolic pressures imposed by recombinant protein production. Flavinated enzymes however, have a number of suggested uses in the pharmaceutical and agrochemical industries.

Bacterial sarcosine oxidases have been used for diagnostic measurement of creatinine in serum, in combination with creatininase and creatinase (Koyama, 1996) and flavoenzymes can be used in the biosynthesis of a relatively new group of antibiotics based on Nikkomycin (Venci, et al., 2002). These antimicrobials are potent inhibitors of chitin synthase (Venci, et al., 2002) and therefore find their uses as therapeutic antifungal agents (for the treatment of human conditions), but also easily degradable insecticides for the agriculture industry.

Flavoenzymes have been linked with soil detoxification where aromatic compounds have been found to be broken down via hydroxylation catalysed by flavoenzymes (Massy, 2000). These include many of the polycyclic aromatic hydrocarbon compounds often

found in oil spills or decommissioned industrial sites, which are known to have carcinogenic, mutagenic and teratogenic properties. Linked to this are the biocatalytic uses of flavoprotein monooxygenases in the desulferization of fossil fuels, (Gray et al., 2003). The conventional chemical hydrodesulphurization process used often for diesel oil is not adequate for the complete removal of these compounds, (Berkel et al., 2006) however it has been shown that overexpression of a flavin reductase in *E.coli* and *P.putida* can greatly enhance the overall rate of desulphurization in these processes, (Galan et al., 2000).

Flavoenzymes have also been seen to catalyse many of the same reactions known to be characteristic of cytochrome P450. These involve a diverse range of substrates and are generally the first step in the electron transport chain. Examples are seen in erythromycin biosyntheses and also in the first part of drug metabolism by the body where the action of cytochrome P450 has been shown to change the relative potency of the drug substance and could make a crucial difference in drug delivery and dosing. A striking catalytic feature of the P450 enzymes and other flavin-dependent monooxygenases is their ability to catalyse regioselective hydroxylations and entioselective sulfoxides, (Berkel, 2006). Regio – and/or enantioselective insertion of oxygen is not straightforward to be accomplished when using chemical techniques which makes the biocatalytic uses of flavo proteins even more attractive.

Enantiopure styrene oxides are important building blocks for the pharmaceutical industry, (Panke et al., 1998) where different enantiomers of drug compounds can have wildly different effects in the body. An obvious example is that of Thalidomide where the R –

enantiomer is effective against the symptoms of morning sickness whereas the S – enantiomer has powerful teratogenic properties, (Stephens et al., 2000). Although in the thalidomide case the production of a single enantiomer would have little effect clinically as the two forms readily interconvert *in vivo*, (Stephens et al., 2000) for other compounds this may be an important part of their production and commercialisation.

From all the flavoprotein monooxygenases the group known as Baeyer – Villiger monooxygenases have been studied extensively for their wide range of usable substrates and, as described above, their enantio and regioselectivity, (Berkel et al., 2006). Recently a new flavin dependent biocatalyst was discovered known as phenylacetone monooxygenase which has been shown to be both thermo stable and tolerant towards solvents, (Fraaije et al., 2005). Biocatalytic conversions using previously discovered BVMOs are often hampered by their instability so this is an important discovery which may further expand the uses of these enzymes.

Finally, flavoproteins have been linked with bioluminescence in many organisms, phototropism in plants, nucleic acid repair in prokaryotic and eukaryotic cells and also in programmed cell death (apoptosis) (Petushkov, & Lee, 1997; Christie et al., 1998; Massey, et al., 2000; Naudi, et al., 2007). It is easy then, to see why there is much interest in these cofactor requiring enzymes in the above fields and other areas, such as biotransformation, resulting in a strong reason for studying this area.

As shown, various flavinated proteins may have valuable uses and this study looks to understand alterations in the pathways producing their common cofactor, FAD. Although

MTOX will be the focus of the work, the knowledge gained on cofactor metabolism should be relevant for other members of the group.

1.7. Riboflavin pathway and genes of interest

To be able to determine which deletion mutants in the Keio collection (Baba, et al, 2006) had the potential to impact MTOX activity via FAD flux, the riboflavin pathway was examined (Figure 1.4).

Referring to the pathway there are 5 obvious genes that may have a direct impact on riboflavin production and linked to that, both FMN and FAD production.

The gene ribC codes for the enzyme riboflavin synthase transferring an alkyl group from 6,7-Dimethyl-8-(1-D-ribityl)lumazine and producing riboflavin as one of the resultant compounds. It was obvious that a deletion mutant without this enzyme would prove lethal as the flux to riboflavin would be curtailed. Thus this mutant was also not included in the Keio collection.

Therefore the next gene in the pathway to riboflavin was investigated; ribF codes in the majority of prokaryotes for a bifunctional flavin kinase (Martinez – Julvez, et al., 2003) occasionally occurring as a monofunctional FAD synthetase but this function is more common in eukaryotic cells than prokaryotes (Solovieva, 2003). In *E.coli* K-12 W3110

this enzyme catalyses the conversion of riboflavin to FMN utilising ATP and subsequently the conversion of FMN to FAD, again using phosphate from ATP, is catalysed. Although it would be interesting to see the effects of up-regulation of this gene (see Chapter 7) it again was clear that a deletion in ribF would cause a cessation in FAD and FMN production. These two cofactors play such key roles in areas such as mitochondrial electron transport, fatty-acid oxidation, and metabolism of vitamin B6, vitamin B12 and folates, (Kearney, 1952) that a deletion mutant would be lethal.

The two genes downstream of riboflavin cobB and cobT both coding for 5, 6, -dimethylbenzimidazole phosphoribosyltransferase and catalyzing the reaction of beta-nicotinate D-ribonucleotide + 5,6-dimethylbenzimidazole to nicotinate + alpha-ribazole 5'-phosphate were considered as possible targets.

Another possible target for mutation was thought to be appA, coding a phosphoanhydride phosphorylase and catalyzing the non-reversible conversion of FMN to riboflavin with a phosphate being donated to ADP to form ATP. Very little information is available in the literature about the rate of back conversion from FMN to riboflavin facilitated by appA and it is not included in Palssons reaction network that is used as the basis for the modeling work described in chapter 6. Although cobT and cobB are included no mention is made of appA.

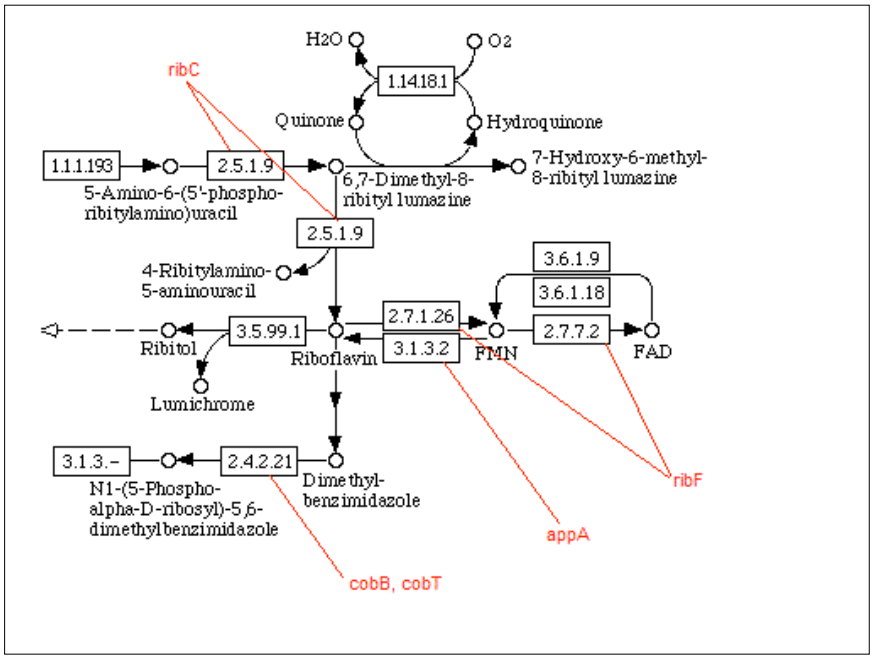


Figure 1.4) Detail of the Riboflavin pathway showing genes of interest for FAD production in E. coli K12 (Source: Kegg Database)

As non-lethal mutants, appA, cobB and cobT were contained in the Keio collection and available for use. It was thought that potentially removing appA may reduce the back conversion of FAD to FMN and therefore result in an accumulation of FAD. cobT and cobB are known to work in tandem (Chen, et al, 1995) therefore whilst removing both would prove lethal to the host, stopping production of vitamin B12, a deletion of either one would probably reduce, but not stop, the flux downstream of riboflavin. Again the hypothesis was that a 'constriction' in the pathways downstream of riboflavin may cause an accumulation of riboflavin and FMN/FAD levels may rise as a consequence of this.

Further rationale behind the deletion choices is included in chapter 5 (section 5.11) where a discussion of the results gained from each mutant strain is also found.

1.8. Constraint-based metabolic modelling

The construction of *in silico* models has allowed the interpretation of a flood of genomic, transcriptomic, and other data generated in high-throughput experiments. The use of these models allows an insight in to large amounts of complex data and can be used to place that information in the wider context of cellular physiology (Reed & Palsson, 2003).

Models have been constructed using kinetic information, cybernetic methods, stochastic models or those based on biochemical systems theory but none have proved as robust or grown to such size as many constraint based systems (Reed & Palsson, 2003). Relevant to the planned work is the history of using constraint based metabolic modelling to predict changes in metabolic flux of gene knockouts or deletion mutant often compared to the parent strain (Segre, et al, 2002).

The flux balance analysis (FBA) method has been developed in the light of the explosion of "-omics" data over the last few decades. It can be used to assess the capabilities and systemic properties of a metabolic genotype (Edwards and Palsson, 1999), and at its heart relies on achieving a steady state mass balance for a defined number of metabolic reactions.

FBA methods only require information about metabolic reaction stoichiometry, metabolic requirements for growth and potentially strain specific parameters such as seen in this study with the production of GFP and MTOX. This information determines a domain of stoichometrically allowable flux distributions that can define a strain's, "metabolic genotype" (Varma and Palsson, 1994), this genotype can be further constrained where additional kinetic or regulatory information is available.

Also relevant to the planned study, *E. coli* has been the focus of many early and developing constraint based models which has then broadened to organisms such as *Helicobacter pylori, Streptomyces coelicolor* and the industrially important, *Saccharomyces cerevisiae*. (Edwards, et al, 2002; Borodina, et al, 2005 and Förster, et al, 2003)

As a comparison, kinetic based models can give a very detailed understanding of metabolic processes but are often limited by the amount of specific information available. Constraint based modelling however, does not attempt to define one single solution for a set of governing equations, rather calculating a collection of allowable solutions that ultimately define a solution space. This can then be further reduced with the imposition of greater constraints such as mass balance, stochiometric information, regulation of the enzymatic reactions concerned and details of enzymatic capacity (with appropriate V_{max} data) (Bonarius et al, 1997).

The solution space can be further defined and reduced in size using a number of different techniques. These are better described in Figure 1.5 but include extreme pathway analysis, where the edges of the solution space are used to define extremes in metabolic behaviour and phenotypic phase plane showing what conditions the network operates under different limitations.

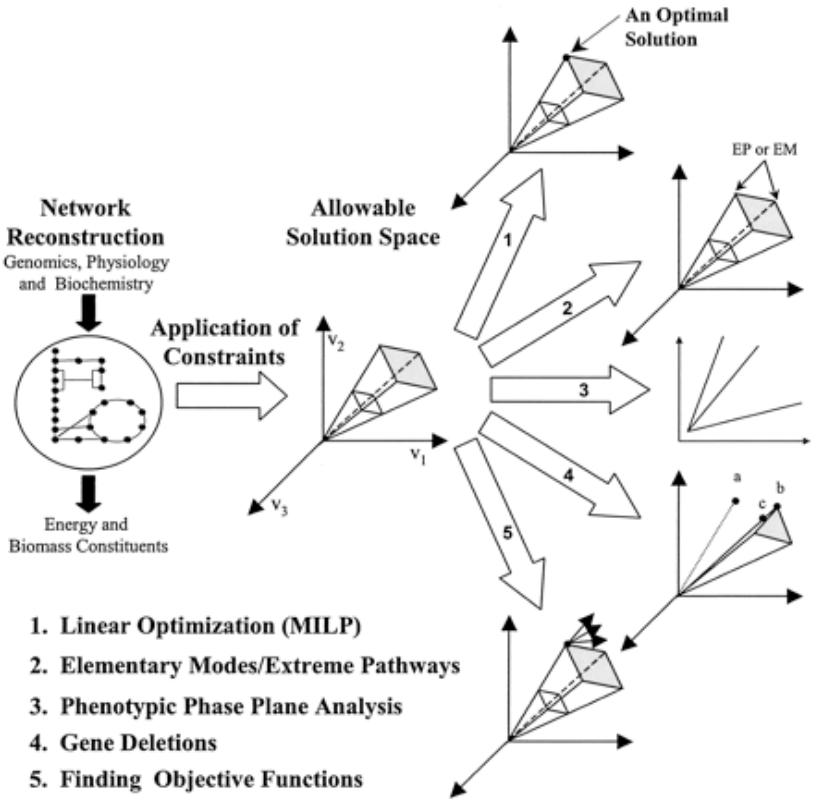


Figure 1.5) Constraint based modelling: Application of constraints to reduce the previously defined solution space in which a cells network must operate, (Source: Reed and Palsson, 2003)

The effect of gene deletions can be computed which is further explored in Chapter 6, and finally linear optimisation can be used to find optimal conditions (allowing the definition of optimal yield of products or optimal biomass accumulation as has been attempted

again in Chapter 6) this is often referred to as flux balance analysis (Varma and Palsson, 1994). In this chanpter both gene deletions and linear optimization were used to shrink the potential solution space with constraints of FAD and MTOX production and maximum specific growth rate.

1.9. Aims of the work and layout of thesis

The following work is organised into 4 main results chapters 3 – 6 which describe the development of a number of methods used in the project, experimental results from the GFP construct, the MTOX producing strain and the modelling work respectively. These chapters are supplemented by chapter 2 which describes the methods used in the study and finally chapter 7 which gives a general discussion of the work as a whole and attempts to place the study in current context.

The aim of the work as a whole was to investigate flavin cofactor supply in a system expressing a recombinant flavo-protein and to understand how different strains and/or mutations may affect the supply of FAD and subsequently the activity of the recombinant protein product. Following this aim a number of experimental objectives were considered:

- To develop a model system in *E. coli* for the controlled expression of GFP with the potential to apply it for other proteins

- To establish a protocol and develop methods for comparable, growth and recombinant protein production in a variety of *E. coli* host strains, with a variety of plasmid constructs and the potential for a semi-high throughput method
- To engineer the same plasmid construct used for GFP production for the expression of *N*-methyl – l – tryptophan oxidase and develop a reliable assay for both active and inactive forms of the expressed enzyme
- To measure FAD, riboflavin and FMN levels at certain points in fermentations and relate this information to MTOX activity levels
- To model the test system using Palsson's *E .coli* reaction network, altered for the strains used and for MTOX production, and compare values generated with experimentally data.

2. Materials and methods

2.1. Introduction

The following chapter outlines methods used throughout the study. It initially concentrates on the approach used to characterise growth of *E. coli* strains and the media chosen. It goes on to explain the analytical methods used for the measurement of cellular products and metabolites. It covers statistical analysis of data sets, which are used throughout the work and explains the background of modelling software used to generate a metabolic model of *E.coli* for use throughout chapter 6.

2.2. Microorganisms used and cultivation methods

2.2.1. Selection of host strains

After initial experiments with MG1655, 3 additional strains were used for the continuation of the work all four strains and their respective mutations from the *E. coli* wild type (Source: CGSC website (http://cgsc2.biology.yale.edu/index.php)) are shown in Table 2.1.

Table 2.1) E.coli strains with genotypes used in the study (Source: CGSC website, http://cgsc2.biology.yale.edu/index.php)

Strain	Genotype
MG1655:	F^- LAM^- rph-1
W3110:	F^- LAM^- rph-1 In(rrnD-rrnE)
BW25113:	F^- LAM^- rph-1 DE(araD-araB)567 lacZ4787 (del) (::rrnB-3) DE(rhaD-rhaB)568 hsdR514
TG1:	F' traD36 laclq Δ(lacZ) M15 $proA^+B^+$/supE Δ(hsdM-mcrB) 5 (r_k^- m_k^+ mcrB-) Δ(lac-proAB)

E. coli K-12 MG1655 was chosen as the initial host strain having been obtained from the American Type Culture Collection (ATCC no. 47076). This was used as it formed the basis of the first *E.coli* full genome sequence completed, (Blattner et.al., 1997) and was also the basis of the metabolic model being developed in house.

TG1 was provided by GSK as it was the optimised strain for the production of the recombinant proteins, (see section 5.1) and as such was the only non K-12 strain to be included in the study.

W3110 was used as it is very closely related to MG1655 but also as it was the starting point for the Keio collection of mutants, (Mori, 2000; Baba, et al., 2004 & 2006). This was mutated into BW25113 before undergoing further alterations to produce the numerous Mori single gene deletion mutants available.

The table above clearly shows the relationship between the first 3 strains, which are all used in the mutation work (covered in Chapter 5, section 5.7). It also underlines the difference between the K-12 strains and GSK's production strain TG1.

2.2.2. Cell bank preparation

The original stocks were streaked out on nutrient broth (Oxoid #2, Oxoid/Fischer Scientific UK) with agar added at 1.2 g/L giving a solid nutrient medium. These were grown overnight at 37^{o}C. A single colony was then transferred to a fresh plate and incubated again overnight at 37^{o}C. Using 3.5mL of 20% (v/v) glycerol and a sterile glass rod, cells were removed from the surface of the agar plate and the resulting suspension transferred, using a sterile pipette, into small aliquots stored at -80^{o}C. This constituted the master cell bank (MCB). For further experiments a working cell bank (WCB) was produced by taking a number of colonies from the MCB using a sterile loop and streaking them onto plates, incubating overnight at 37^{o}C, and again preparing glycerol stocks to store frozen.

After the construction of plasmid carrying strains (puvGFP, pMEX8, as described in section 2.3) the same protocol was used for the preparation of bacterial stocks and cells frozen at -80^{o}C for storage.

2.2.3. Media preparation

For all fermentation experiments, *E. coli* was grown initially from the WCB on plates of nutrient broth agar (as described in section 2.2.2) or nutrient agar (Oxoid).

After the initial overnight growth a single colony was transferred into M9 or M6 minimal media (see Table 2.2 and 2.3), or Luria Bertani broth (LB) with the addition of 4% (w/v) glucose for the liquid cultures. M9 was made up of four components: - M9 salts, prepared as a 10x stock solution (pH 6.95), and calcium, magnesium salts and glucose, as a 100x solution. These were autoclaved separately (121^{o}C for 15 min (Denly Autoclave)) and then could be stored on the bench for future use.

Table 2.2) M9 medium composition

Component	Stock Concentration (g/L)	Working Concentration (g/L)
M9 Salts		
$Na_2PO_4.7H_2O$	152.00	15.20
KH_2PO_4	30.00	3.00
NaCl	5.00	0.50
NH_4Cl	10.00	1.00
Glucose	400.00	4.00
$CaCl_2$	1.47	0.02
$MgSO_4.7H_2O$	24.64	0.25

M6 medium was made up of five different components. Similarly to M9, a salt solution was made up to a 2x stock (any higher concentration proved difficult to adjust the pH effectively) pH 6.95. Again the magnesium and calcium salts were added separately for the same reasons as above. Glucose was added from a 100x

stock. Finally a 100x of trace elements was prepared. This concentration was used to enable accurate quantities of constituents in low concentrations.

Table 2.3) M6 medium composition

Component	Stock Concentration (g/L)	Working Concentration (g/L)
M6 Salts		
$(NH_4)_2SO_4$	10.4	5.2
NaH_2PO_4	7.72	3.86
KCL	8.05	4.025
$MgSO_4$	104	1.04
Citric Acid	8.32	4.16
$CaCl_2$	25	0.25
Glucose	400	4
TRACE ELEMENTS 100x Stock solution		
Citric acid	104	1.04
$CaCl_2$	5.22	0.05
$ZnSO_4$	2.06	0.02
$MnSO_4$	2.72	0.03
$CuSO_4$	0.81	$8.0x10^{-3}$
$CoSO_4$	0.42	$4.0x10^{-3}$
$FeCl_3$	10.06	0.10
H_3BO_3	0.03	$3.0x10^{-4}$
$NaMoO_4$	0.02	$2.0x10^{-4}$

LB broth is made up of 10 g/L tryptone, 5 g/L yeast extract and 10 g/L NaCl, however this was available as a premixed powder from Oxoid.

Antibiotics, where necessary were made up to working concentrations, sterile filtered, aliquoted into single use volumes and stored at -20ºC until needed. Ampicillin stocks were made up to 50mg/mL and used at 500µg/mL for solid media or 300µg/mL for liquid culture. Kanamycin was made up to a working concentration of 37.5mg/mL and used at 30µg/mL for all media. Plasmid induction was enabled

through the addition of IPTG which was made up as a stock concentration of 200mg/mL and used to a final concentration of 1mM (238 mg/L) for all experiments.

2.2.4. Shake flask fermentation experiments

Initial shake flask fermentations were carried out in 2L baffled vessels however the size of these was unnecessary for the planned work and made large numbers of replicates inconvenient.

The seed train was essentially similar for the shake flask experiments and the stirred tank reactors, however the scales were obviously altered for the larger cultures. A single colony was picked from an overnight plate cultured from the working cell bank and was used to inoculate 2.5mL of medium in a 25mL Universal tube, this was then grown overnight at 37^{o}C in an orbital shaker (200rpm, 5cm throw, New Brunswick Scientific, Co., Inc., USA). At the start of the experiment this inocculum was added to 22.5mL fresh medium in a pre-sterilised 250mL baffled shake flask and returned to the shaker at 37^{o}C. Samples were taken every hour for a period of 8 hours, and then left overnight for a final 24h reading. To lessen the effect of diminishing culture volume sample volumes were either 100µL (for samples analysed for optical density and fluorescence only) or 1mL (where the sample needed to be lysed for further analysis) at alternate hours.

2.2.5. 5L stirred tank bioreactor experiments

Larger volume fermentations were run using a pair of 7L reactors (New Brunswick, New Brunswick Scientific Co., Inc., USA) with an operating volume of 5L. The bioreactors were of glass construction with stainless steel head plates containing the ports used for sterile additions. The vessels were filled with the appropriate amount of RO (reverse osmosis) water prior to heat sterilisation. The DOT, pH and temperature were monitored by Ingold electrodes (Ingold, Switzerland), linked to real time data acquisition system (RT-DAS) (Acquisitions Systems, Reading UK). Off gas concentrations of oxygen, nitrogen, carbon and argon were monitored by mass spectroscopy, (Fisons Scientific Instruments, Prima 600) and parameters such as oxygen uptake rate (OUR), carbon dioxide emission rate (CER) an respiratory quotient (RQ) were calculated. The vessels that would be used for the addition of acid, base, media components and inocula were sterilised alongside the main vessel. Post-sterilisation, media components were added via sterile transfer and finally the inoculum. The fermentations were then run at 37^{o}C with an airflow rate of 1.5vvm and an agitation rate of 500rpm. Polypropylene glycol (PPG) was added as antifoam at 4mL/L if required.

The fermentation was monitored remotely and automatic pH control, to 7, was achieved with the addition of 2M NaOH and 2M H_2SO_4. Off gas from the foam trap was filtered through a 2µm filter before entry into the mass spectrometer.

Sampling of the fermentations was carried out every hour with a 4mL sample taken from the aseptic sample port and line.

2.2.6. Sample preparation for HPLC analysis

Samples from shake flask or stirred tank reactor destined for HPLC analysis of FAD, FMN and Riboflavin levels (see section 2.5) were quenched using perchloric acid to "freeze" all cellular metabolic activity at the point of sampling (Vinkler, et al., 1978). For shake flask fermentations 0.2mL of 1.6M $HClO_4$ was added to sterile Eppendorf tubes and 1mL of sample added to the tubes upon sampling. For stirred tank reactors 0.8mL of 1.6M $HClO_4$ was added to sterile universal tubes and 4mL of sample collected directly into the prepared tubes. For each case, 1 minute after quenching the equivalent volume of 10M KOH + 0.25M Tris was added to neutralise the solution. After neutralisation samples were stored at 4ºC, protected from light for no longer than 24 hours before lysis and further analysis.

2.2.7. Cell lysis

Bacterial lysis was carried out via sonication (Soniprep150, Sanyo, Japan) in lysis buffer (10mM Na_2PO_4, 1mM EDTA, 10mM $MgCl_2$) (Feliu et al., 1997).

Cells were pelleted (10min, 16500 x g, 4°C), washed twice with 1mL sonication buffer and resuspended in a further 1mL buffer. Sonication was carried out, on ice,

for 10 s at 8 kHz. These conditions were determined in experiments carried out and described in chapter 3 section 3.2.

Cell debris was removed where necessary by a further centrifugation step (10min, 16500 x g, 4°C) to produce a clarified cell lycate.

2.3. Plasmid construction

The initial plasmid construct was uvGFP (Green Fluorescent Protein optimised for a high UV signal) (BD Biosciences) incorporated into a pBR322 type plasmid. The construct had an ampicillin resistance gene to allow selection and control of induction was sought with the inclusion of the lac repressor. A map of the original plasmid is shown in Fig. 2.1.

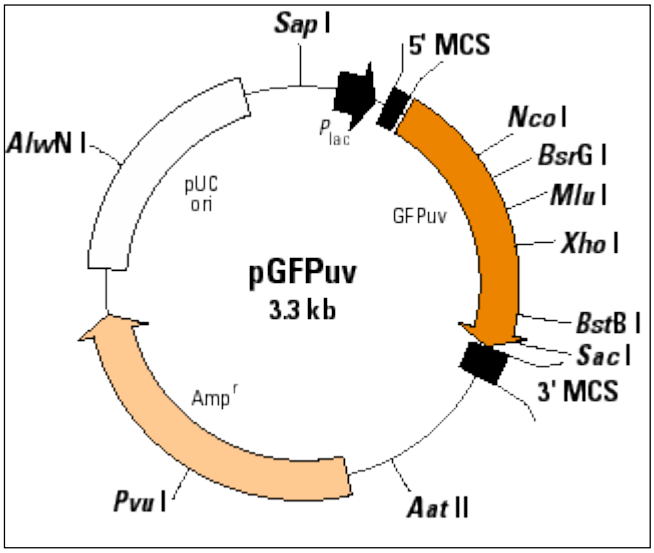


Figure 2.1) Plasmid map of original GFP construct (Clontech)

The multiple cloning sites (MCS) allowed the easy removal of the GFP gene if required

Once the focus of the work moved to the MTOX protein (see Chapter 5), solA (gene coding for N-Methyltryptophan oxidase) was sourced by Prof. John Ward and amplified using PCR and the primers below, based on the published data (Gene Bank accession number EG12669).

N-terminus:

5’ AAGCTTATAAGGAGATATACATATGAAATACGATCTCATCAT 3’

Hind III Ribosome binding site

C-terminus:

5’ GAATTCTTATTGGAAGCGGGAAAGC 3’

EcoR1

The PCR product was then incorporated into pCR®2.1-TOPO® cloning vector (Figure 2.2) (Invitrogen) and transformants (for method see page 41) plated onto selective AMP, Xgal, IPTG plates. These were grown overnight and white colonies picked for further work.

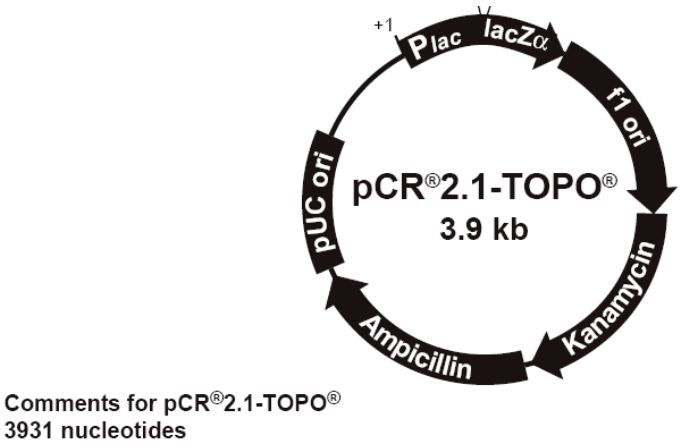


Figure 2.2) Plasmid map for the pCR®2.1-TOPO cloning vector

Plasmid DNA purification was then carried out using the QIAprep Spin Miniprep Kit (Qiagen, UK) according to the manufacturer's instructions. Plasmid DNA was eluted into TE buffer (10 mM Tris-HCL (pH 7.5), 1 mM Na_2EDTA), and stored at -20°C. Restriction digests of the resulting DNA were carried out with EcoR1 and HindIII.

Digestion of plasmid DNA with restriction endonucleases was performed using enzymes from New England Biolabs. Typically 10µL plasmid DNA was digested in a total volume of 20µL, comprising 1µL restriction enzyme, 2µL 10X restriction buffer and 7µL TE buffer. The solution was then incubated at 37°C for 2-3 hours.

The rationale behind the digests was that a 1200bp fragment would be seen on a gel from a HindIII digest if the PCR product had incorporated into the cloning vector with the correct orientation. A 60bp fragment would indicate incorrect alignment. With the EcoR1 digest a 1200bp fragment would be seen if solA was present, nothing if not.

As a final check plasmid DNA was purified and quantified using the Picogreen assay (Labarca, 1980) and then sent for sequencing to verify the construct was correctly aligned. (Data not shown)

Once these were confirmed minipreps were again done for the TOPO clone and the GFP carrying strain. Restriction digests were again completed with a mix of EcoR1 and HindIII to release the inserts from each plasmid. Competent cells of *E. coli M*G1655 and W3110 were prepared following a protocol provided by Prof. John Ward. Briefly this consisted of a 5mL overnight culture of MG1655 and W3110 in nutrient broth transferred to 20mL of fresh broth containing 20mM $MgCl_2$. This was grown for one hour and then spun down at 7000 x g , 10min, 4^{o}C and kept on ice.

The pellet was then resuspended in 2mL of ice cold, sterile 75mM $CaCl_2$, 15% (v/v) glycerol, aliquoted into 200µL amounts and stored at -80^{o}C until needed.

From the digest of the TOPO clone a gel was run (see figure 2.3), the band for the solA gene excised from the gel and an extraction protocol (Qiagen) completed to

release the DNA. The expression vector was then heat inactivated for 10min at 70°C before being mixed with the MTOX fragment, T4 buffer, (50 mM Tris-HCl, 10mM $MgCl_2$, 1mM ATP and 1mM Dithiothreitol) and 1µL T4 ligase in a 20µL rection and stored at 16°C overnight for the ligation to occur.

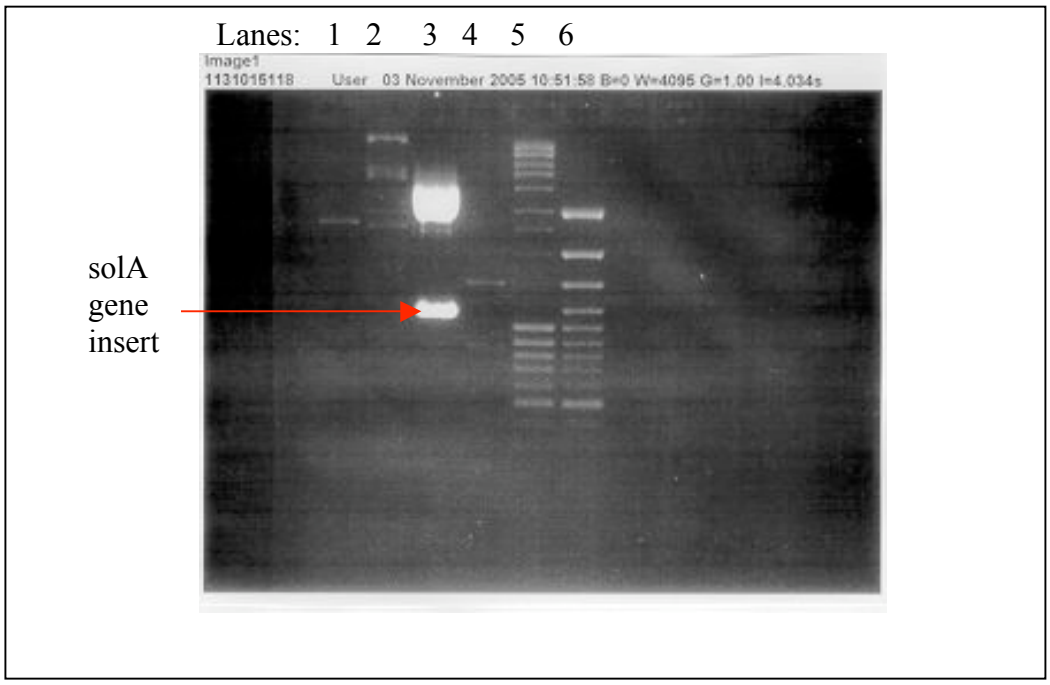


Figure 2.3) Digest of MTOX TOPO clone showing band for excision

Key:

Lane 1: GFP plasmid

Lane 2: λ PST marker (bands visible =11490, 5777, 4749, 2888bp)

Lane 3: TOPO digest with solA gene insert

Lane 4: Low molecular weight ladder (BioRad)

Lane 5: Mass ruler (BioRad) (bands visible = 10000, 9000, 8000, 7000, 6000, 5000, 3000, 1000, 500, 400, 300, 200, 100bp)

Lane 6: Gene ruler (BioRad) (bands visible = 5000, 3000, 2000, 1500, 1000, 500, 400, 300, 200, 100bp

The final stage of the process was to mix the ligated DNA with an aliquot of transformed cells and again follow a protocol suggested by Dr Ward. After this stage cells were plated out and 8 samples taken for a final miniprep. This step was again digested with EcoR1 and HindIII before being run on a gel to confirm the correct insert had been ligated into our host. (results not shown)

All plasmids used including the pQR constructs generated during this work are summarised in Table 2.4.

Table 2.4) Plasmid used in this study and its characteristics

Plasmid (size in kb)	Characteristics	Source
pGFPuv (3.3 kb)	LacZ Ap^r	Invitrogen, UK
pMEX8 (5.7 kb)	LacZ Km^r	Dr Chris Preston GSK, UK
pQR497 (5.0 kb)	solA in TOPO 2.1	Ward (unpublished)
pQR498 (3.6 kb)	solA in pGFPuv Ap^r (GFPuv excised)	This study

Abbreviations: Ap^r, ampicillin; Km^r, kanamycin; kb, kilobase

2.4. Analytical methods

2.4.1. Optical Density

As a measurement of cellular growth the optical density of the broth at 600nm was measured using a spectrophotometer (Cecil Aquarius 7000, Cecil Instruments UK).

After taking a sample from either shake flask or fermenter, the sample was immediately placed on ice and the absorbance read at 600nm against a blank of sterile media. The data obtained, in conjunction with dry cell weight (see below) could then be used to generate growth profiles of the cultures and calculate important growth parameters such as the maximum specific growth rate (μ_{max}) and conversion yield on glucose ($Y_{x/glu}$). Maximum biomass concentration (X_{max}) was determined by referring to an OD/DCW correlation (Section 2.4.2)

2.4.2. Biomass correlation

Part of the early work sought to develop a reliable correlation between optical density (OD, 600nm) and dry cell weight (DCW, g/L) to enable the calculations described above to be performed. After OD measurements had been taken, 1mL duplicates were made and added to pre-weighed, pre-dried (90^{o}C for 24h) eppendorf tubes. These were placed in a microfuge (Biofuge 13, Heraeus, Germany) at 17900 x g for 10min. Supernatant was transferred into clean eppendorf tubes and stored at -20^{o}C for future analysis. The cell pellets were then resuspended in RO water and centrifuged once again. The supernatant was discarded and the walls of the tube carefully dried of any residue. These tubes were then weighed for a measurement of wet cell weight (WCW). They were then returned to the oven, removed and allowed to cool in a desiccator, and weighed periodically until a stable weight was reached, this being the dry cell weight (DCW g/L). A linear correlation between optical density at 600nm and dry cell weight was obtained as shown in Fig. 2.4.

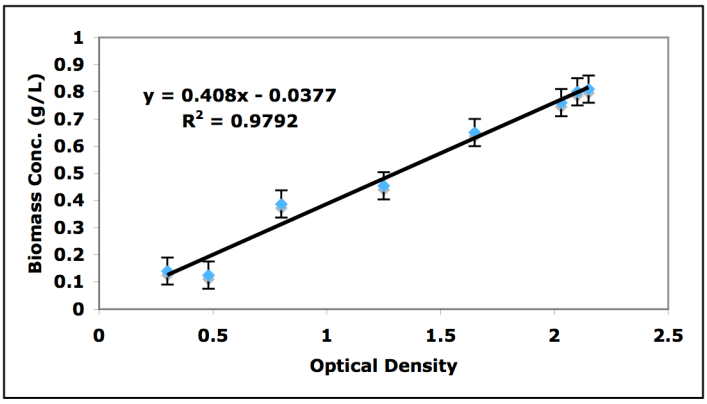


Figure 2.4)Calibration curve showing Biomass concentration plotted against OD 600nm

This gave a reliable correlation between optical density (standard error bars plotted, n = 3 independent runs) and dry cell weight for *E.coli* cultures with the R^2 value being 0.9792. The gradient, as shown on the graph was 0.408 (i.e. 1 unit OD600 is equal to 0.408 g/L in culture) This is useful as in practice it is much less time consuming to measure the OD of a sample and then calculate DCW, compared to measuring the cell weight directly.

2.4.3. Protein analysis

Two methods were used to determine either the level of total protein present in any sample or specifically the level of MTOX protein present. The level of total cellular

protein was determined with the Bradford assay (Bradford, 1976; Reed and Northcote, 1981). The principle of the assay is based on the absorbance shift from 465 – 595 nm that occurs when Coomassie blue binds to proteins in acidic solution.

Protein levels in the sample were determined by comparison to a standard curve generated with serial dilutions of bovine γ globulin with the OD measured at 595nm. Cells were first lysed as described in section 2.2.7 by sonication. After sonication the cell lysates were centrifuged (17900 x g, 10min, 4^{o}C) and the supernatant used for further analysis.

Stacked SDS-PAGE (12% with 4% stacking, precast gel, BioRad) was used to visualise the protein make up of the samples after staining with Coomassie Brilliant Blue R-250 and washing in 5% (v/v) acetic acid. After sonication the supernatant contained soluble protein. Resuspending the pellet gave the insoluble protein fraction, whereas the sample after the initial sonication could be run as total protein. Protein gels were then stained with Coomassie before being photographed.

For specific measurement of the MTOX protein the BCA assay was used (Thermo Scientific) following the supplier method to generate total cellular protein data. A SDS polyacrylamide gel was then run, stained with Coomassie brilliant blue and analysed using the Bio-Rad Gel Doc system. This compares the intensity and distance run of each band seen in the gel against a known concentration of control protein (Bovine serum albumin, BSA). From this comparison it is possible to

quantify the size and relative amount of protein present on the gel and therefore in the lysed cell sample. Results are shown in section 5.7.

2.4.4. GFP fluorescence assay

GFP was chosen as the model system for recombinant protein production partly due to its ease of measurement. 150µL of undiluted cell lysate was added to a 96 well plate and read with an excitation wavelength of 365nm ± 5nm and an emission wavelength of 510nm ± 5nm using a Tecan plate reader (Tecan, Switzerland) with Xfluor4safireII software. The signal was compared to a blank of wild type *E.coli* cell lysate.

2.4.5. Peroxide assay

The MTOX enzyme or N-methyl tryptophan oxidase is known to be able to convert N-methyl-L-tryptophan (sometimes known as L-Abrine) into L-tryptophan and peroxide. Therefore the assay for the detection and measurement of active MTOX protein was based on the production of this by-product shown in the following reaction scheme (Fig. 2.5)

$$CH_3NHCH(R)CO_2^- + O_2 + H_20 \Rightarrow NH_2CH(R)CO_2^- + CH_2O + H_20_2$$

Figure 2.5) Reaction scheme for the oxidative demethylation of N-methyl amino acids as catalysed by MTOX.

Samples of lysed cells were diluted to 1:100 and 1:1000, 50µL of this dilution being added to relevant wells in 96 well plates.

The substrate, N-methyl-L-tryptophan was prepared by dissolving to a concentration of 10mM in a 50mM sodium phosphate buffer (pH 8.0) and placing in a sonicating water bath (37^{o}C, 10min) to aid solubility. Once prepared, 0.5µl of substrate solution was added to each well and left at room temperature for 30min. After incubation with N-methyl-L-tryptophan the protocol for the Enzolyte™ ADHP Hydrogen Peroxide assay Kit (Anaspec, USA) was followed. This involved preparing a reaction mix of 50µL ADHP (10-Acetyl-3,7-dihydroxyphenoxazine) and 100µL Horseradish peroxidase, made up to 5mL with 0.01M PBS (pH 7.4) and then adding 50µL to each well containing peroxide. The reaction proceeds as follows (see Fig. 2.6)

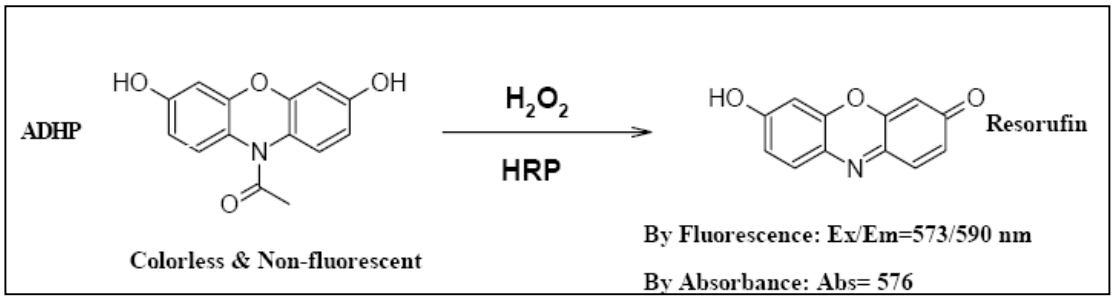


Figure 2.6) Reaction scheme describing the colorimetric reaction which is the basis of the MTOX enzyme assay

Readings were then taken after a further 30 min with the plate reader (Tecan, Switzerland) at 576nm for absorbance and emission at 590nm, excitation at 540nm for the fluorescence.

The readings compared with a dilution series of peroxide gave results in terms of the amount of peroxide evolved and are therefore difficult to compare across differing strains/conditions. Quantification was undertaken however, based on Wagner's method for the measurement and reporting of MTOX and MSOX activities, (Wagner, et al., 1999). In this method, units of activity are quoted; one unit of activity being defined as the formulation of 1μM peroxide/min at 25ºC following standard assay procedures. In chapter 5 this measure is mainly used in conjunction with a measure of total amount of MTOX protein present. These two data sets can then be combined to produce a specific value of units of enzyme activity/mole MTOX protein present which is comparative across strains. It was realised however that the above assay was only measuring active enzyme, since MTOX requires the co-factor FAD to bind to confer activity and therefore exists in an apo- or inactive, or a holo-/active form. In answer to this problem the assay was modified using information about the ability of FAD to restore activity to apo enzyme produced (Hassan-Abdallah, et al, 2004).

With the information that flooding the sample with FAD would convert ca. 90% of any apo enzyme to the holo form, the assay was altered to include a further 30min incubation with 1mM FAD into relevant wells of the 96 well plates after the addition

of substrate. An equivalent volume of RO water was included in each control well. After this incubation the peroxide assay as described above was carried out allowing the treated cells to give a reading of total enzyme produced and the control cells a reading of active enzyme originally produced. The results of these assays and comparative data before and after addition of FAD are presented in section 5.4.

2.5. HPLC assays

High performance liquid chromatography was used for two different separations throughout the work. Initially a column and a UV detector were used for the separation of organic acids and the determination of glucose and acetate levels at various points throughout the fermentation. Further into the work a second rig was used with fluorimetric detector to measure levels of FAD, FMN and riboflavin in clarified culture lysates.

2.5.1. HPLC analysis for Glucose and Acetate

A Dionex HPLC rig was used with an Aminex HPX – 87P ion exclusion column (300 x 7.8mm, 9µm pore; specific for organic acid and alcohol separations (BioRad, UK)) with a mobile phase of 5mM H_2SO_4 running at 0.8mL/min. Fermentation samples were centrifuged (17,900 x g, 10min) and the supernatant sterile filtered through a disposable 2µm syringe filter. Samples were also prepared with serial

dilutions of a known concentration of glucose and acetate. The standard curves generated are shown in figure 2.7.

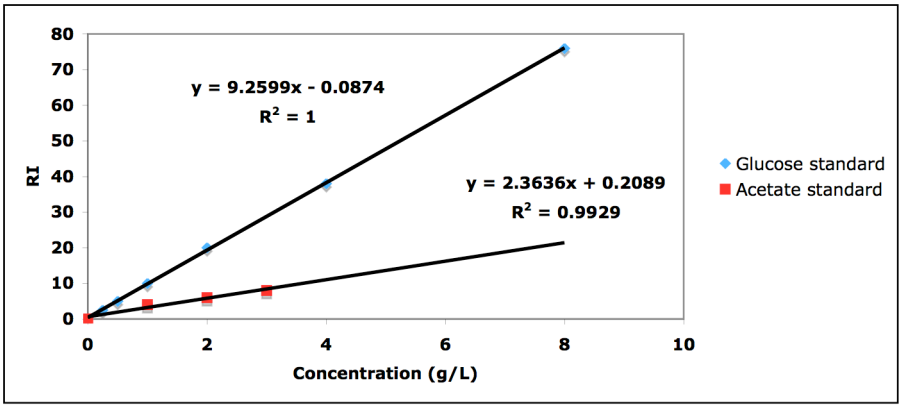


Figure 2.7) Standard curve of Glucose and Acetate using HPLC

Detection was with refractive index (RI) (as shown above) or UV 210nm with the acetate peak having a retention time of about 9 minutes and the glucose, 14 min.

2.5.2. HPLC analysis for FAD, FMN and Riboflavin

As much of the work in chapter 5 aimed to increase levels of cellular FAD it was important to have a robust method for the separation and measurement of FAD, FMN and riboflavin. Samples were prepared after perchloric acid quenching (described in section 2.2.6) by lysing the cells, centrifuging to remove cell debris and then sterile filtration to avoid clogging the HPLC column. The method employed was an adaptation of Barile's paper of 1997, (Barile, et al. 1997) but also

drew on other work by the same author and others, (Barile, et al. 1993, King, et al 1962).

Separation was achieved using a Chronusil-S ODS-2, 5µm x 25mm x 0.46mm HPLC column connected to a Dionex P580 pump and AS350 autosampler which was kept chilled and protected from light at all times. Detection was possible using a Jasco fluorimetric detector (using excitation and emmision wavelengths at 450nm and 520 nm respectively). All hardware was controlled using Chromelion software (Dionex Corp, Sunnyvale, CA) except the detector wavelengths and gain, which had to be set manually.

A gradient was used for the mobile phase consisting of two solutions, A and B. Both were a 20mM potassium phosphate base with either 25% (v/v) (A) or 50% (v/v) (B) methanol added. All standards and samples (40µL) were injected and eluted at room temperature for 6 min 40 sec with solvent A followed by 13 min 20 sec for solvent B at a flow rate of 0.75 mL/min.

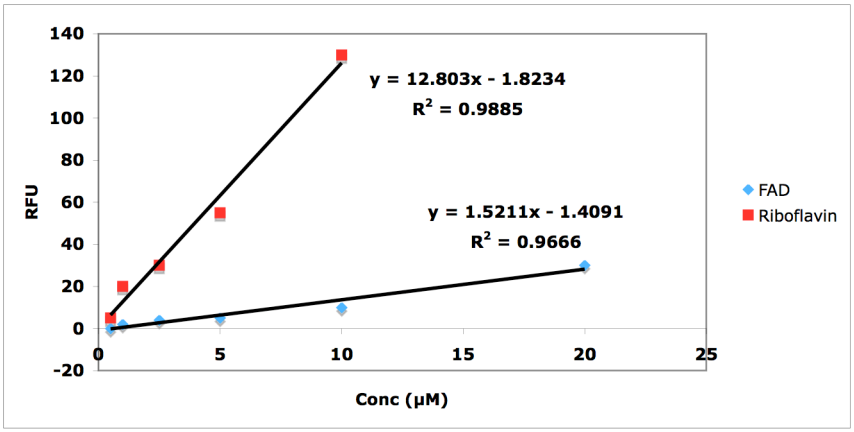


Figure 2.8) Standard curves generated by HPLC of Riboflavin and FAD

Standards curves were generated for riboflavin and FAD and are shown in Figure 2.8. Solutions of FMN standards however did not generate one single peak, with peaks forming at retention times known to correspond to FAD and riboflavin. It was therefore assumed FMN was too unstable in solution, even storing all sample solutions refrigerated and in the dark , to not partially break down into the two other compounds.

2.6. Statistics and calculations

A number of statistical manipulations were necessary to analyse and compare data sets generated in the course of the study, these are explained below.

2.6.1. Coefficient of variance

Coefficient of variance can be described as the degree to which a set of data points varies. The larger the final value the more variation there is in the data set compared. Although not a measure of statistical significance, % CV can be used as a useful tool for the initial comparison of a number of data sets. It is calculated using the equation below.

$$c_v = \frac{\sigma}{\mu}$$

Equation 2.1) Formula for the calculation of the dimensionless number, coefficient of variance

As shown C_v is the ratio of the standard deviation (σ) and the population mean (μ). Multiplying the result by 100 gives percentage C_v.

2.6.2. Student's T-test

The Student's T-test is commonly used to test the null hypothesis that the means of two normally distributed populations are equal generally assuming the variance of both populations is also equal. As shown in equation 2.2 this need not always be the case as the heterscedastic T-test, (often referred to as Welsh's T-test) can be used where variances are unequal.

Common to all T – tests is the fact that the result achieved from the calculation needs to be compared to a table of critical values which will vary depending on the confidence level (generally taken as 95%) and the degrees of freedom. If the calculated value from the T – test is lower than the critical value the null hypothesis is upheld and no significance is seen between the two sample means compared.

$$T = \frac{\overline{X} - \overline{Y}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Equation 2.2) Formula used to calculate the Student's T-test statistic for two sample populations assuming equal variance (Homoscedastic T-test)

X, Y = population means

S_p = pooled standard deviation

n = population size for each data set

For this test, the calculation of degrees of freedom is simple being $n_1 + n_2 - 2$.

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}}$$

Equation 2.3) Formula used to calculate the Student's T-test statistic for two sample populations assuming unequal variance (Heteroscedastic T-test)

S^2_x = standard deviation of X

S^2_y = standard deviation of Y

The calculation of degrees of freedom for this statistic is much more complex than for the homoscedastic test and is shown below.

$$df' = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Equation 2.4) Formula used to calculate the degrees of freedom used for the heteroscedastic T-test as shown in equation 2.2.

S = standard deviation

n = population size

2.7. Modelling

Much of the work done in Chapter 6 is based on a metabolic reconstruction of the MG1655 K-12 strain of *E.coli.* The reaction network used was developed by Bernard Palsson (Edwards, and Palsson, 2000) and is a constraint based model of *E.coli* metabolism. This model was developed using published information such as the annotated gene map and other physiological and biochemical information, (Joyce, et al, 2006 and Reed, et al, 2006). It therefore consisted of a reaction network for the maintainence of the *E.coli* cell and the requirements of carbon sequestering during growth. Crucially for the work planned it also included the riboflavin pathway described in Chapter 1.

The model was constrained by the known ATP requirement for cell maintenance of 7.6mM/g(DCW)/h and an initial glucose uptake rate of 2mM/g(DCW)/h. Also a stociometric equation was generated based on the amino acid sequence of the solA

gene product, (Koyama, and Ohmori, 1996). This, in addition to the riboflavin pathway would allow the comparison of MTOX production with FAD levels and also demonstrate how demands through these two pathways would effect the maximum specific growth rate.

Mutations could be included and as such the model was altered slightly to accommodate for the mutant strains background of BW25113 (see table 2.1).

Once the initial set up was achieved questions were asked of the model in terms of Biomass production, MTOX production and also FAD production. Plots were then generated in an attempt to determine the relationships between these three carbon sinks.

As detailed in Chapter 6 initially a maximum specific growth rate was calculated by setting all carbon flux into biomass accumulation. This allowed an estimation of the maximum theoretical growth rate of 0.5 h^{-1} as shown in Figure 6.1. Then as growth rate was incrementally decreased and all available flux redirected to MTOX protein production an estimation of maximum MTOX molar flux was determined alongside a plot of MTOX production against growth rate.

This approach was then also used to compare flux to the MTOX protein and FAD with two differing glucose uptake rates and additionally no growth, (shown in Figure 6.2). Finally a comparison of the three variables was performed which gave

information on the production of the MTOX protein, the cofactor needed for its activation and also the potential growth rate achievable at certain levels of protein production. This then gave a window of operation which could be compared with experimental values generated in the laboratory investigations. A discussion of where the predictions of the model deviated from the data generated in the laboratory is included in Chapter 6.

3. Method Development

3.1. Introduction

This chapter describes experiments designed to develop methods enabling further work to be carried out on the host strains and the proteins that were being produced.

It initially details work done to investigate the optimal conditions for bacterial lysis, necessary for the further analysis of intra-cellular proteins, co-factors and enzymes. The optimum conditions determined in these experiments were used for the lysis of all bacterial samples throughout the work.

It goes on to detail and explain the development of an HPLC method for the separation and measurement of FAD, from its related compounds, FMN and riboflavin. As the level of active MTOX enzyme was directly related to the level of FAD present accurate measurement of the concentrations of this co-factor was vital to the study.

Methods and results described in this chapter directly relate to the work detailed in the following chapters.

3.2. Bacterial lysis using sonication

Bacterial lysis was carried out via sonication (Soniprep150, Sanyo, Japan) in lysis buffer (10mM Na_2PO_4, 1mM EDTA, 10mM $MgCl_2$) (see materials and method section 2.2.7)

The initial experiment was undertaken to determine the optimum conditions for the lysis to take place. Eight conditions were looked at with varying sonication frequencies and durations. After lysis an estimation of total cellular protein concentration released was undertaken (see section 2.4.3) and also a measure of enzyme activity, calculated using the method for MTOX enzyme activity described in section 2.4.5. The data is presented below.

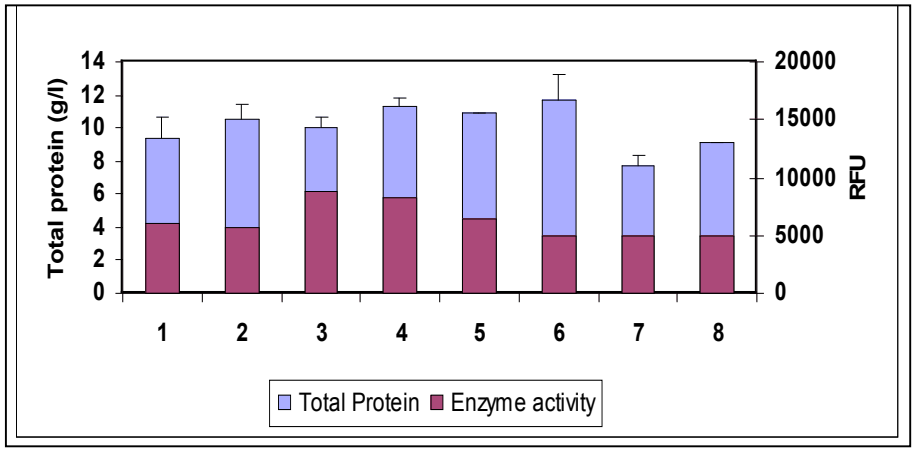


Figure 3.1) Comparison of 8 different lysis conditions with the resulting total protein yield and enzyme activity (Conditions were: 1 – 6kHz : 5 sec, 2 – 6kHz : 10sec, 3 – 8kHz : 5 sec, 4 – 8kHz : 10 sec, 5 – 10 kHz : 5 sec, 6 – 10kHz : 10 sec, 7 – 4kHz : 5 sec, 8 – 4kHz : 10 sec) All consisting of 5 cycles (see materials and method section 2.2.7 for details)

It was shown that condition 6 (10 kHz for 10 s) produced the largest total protein concentration, suggesting the most efficient cell lysis. However for this condition MTOX

activity was low potentially due to enzyme denaturation caused by extreme lysis conditions. Instead it was decided to use condition 4 (8 kHz for 10 s) for all further cell breakage because it gave a good compromise of efficient cell breakage coupled with minimal damage to the enzyme product.

With levels of cellular breakage and total protein yields being adequate for the work planned this methodology was thought sufficient. Chemical lysis agents were investigated such as the Bugbuster® reagent (Novagen) and others. Chemical agents are known to be efficient and reliable for cell lysis but as no sample clean up of the cell suspension was planned with the exception of centrifugation it was thought that compounds used in the lytic process may remain and could have an effect on the assays carried out post-lysis.

3.3. Flavin Stability

Initial work was needed to adapt the assay described in Barile's paper (1997) and to generate data from standard solutions of FMN, FAD and Riboflavin and also ensure complex mixtures could be separated using the same technique.

As described in the Materials and Methods (section 2.4.5) a method was developed based on work done by Barile et al, (1997) but also drew on earlier work from the same author and others (Barile et al., 1993, King et al., 1962). It relied on HPLC separation of FAD,

FMN and riboflavin and fluorimetric detection of these flavin metabolites from clarified lysates generated in shake flask fermentations of the *E. coli* strains used in this study. It was known that riboflavin and its derivatives fluoresced strongly and that a simple measurement at 450 nm could be undertaken, however this would have given no indication of how the flavin component of the cell is partitioned at any one time into FAD, FMN and riboflavin. As this information was crucial to the activation of MTOX, as the enzyme requires FAD as a co-factor conferring activity, HPLC had to be employed.

Initially standard solutions were run which highlighted the importance of storage conditions for both the samples in preparation, and once awaiting analysis. It was understood that FAD would be utilised by the cells, especially expressing a protein requiring the compound to confer activity, which therefore required the use of perchloric acid quenching in the eventual experimental fermentation runs (see section 2.2.6). However the degree to which FAD would degrade naturally throughout normal laboratory preparation was highlighted after the first standard run.

As shown in Figure 3.2 a sample of 20μm FAD when subjected to normal laboratory preparation but 24 hours delay (caused by equipment failure) showed significant degradation of FAD (amounting to 93% loss) to FMN (90%) and riboflavin (3%), the three main peaks shown matching retention times quoted in literature. Barile et al. (1997) stated retention times for all three compounds as “about 4, 5 and 8 minutes for FAD, FMN and riboflavin respectively” with a flow rate of 1mL/min. Correcting for a lower flow rate of 0.75mL/min these times equate to approximately 5.3, 6.7, and 10.7 minutes

giving a good correlation to figure 3.2. Although unplanned this gave a useful insight into the importance of speed of analysis and storage conditions when dealing with further samples. It is true that some degradation is inevitable however after this finding all samples were treated equally and analysed as soon after sampling took place as possible.

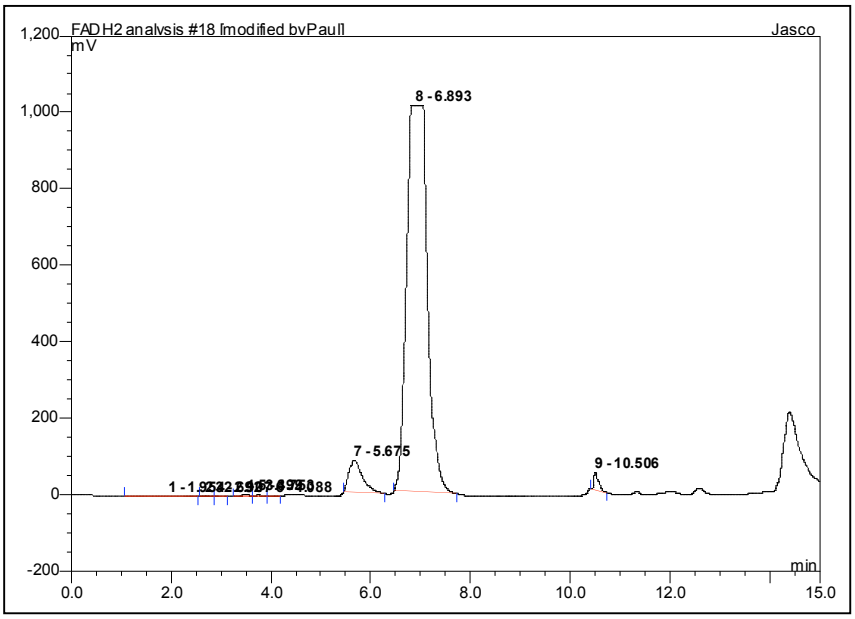


Figure 3.2) Chromatogram of 20µm FAD standard solution after HPLC separation and fluorescent detection (Peaks relate to FAD[5.68 min], FMN [6.89 min], and Riboflavin [10.51 min])

Shake flasks were then run in the same way as described previously with samples being taken at various time points throughout growth. As described in section 2.2.6 samples were immediately quenched with a 1.6M solution of perchloric acid, neutralised and then stored at 4ºC in the dark until analysis could be performed.

3.4. FAD spiking experiments

As an important control, spiking experiments were carried out to determine the percentage recovery of FAD achievable using the modified HPLC method. As detailed in section 2.4.5, 10μM FAD was added to each plasmid free strain cell suspension prior to lysis. 10μM FAD was also added to fresh media and subjected to the same lytic procedure. Finally the same amount of FAD was spiked into fresh media and then sterile filtered before analysis via HPLC. The results are shown below.

Table 3.1) Percentage recovery of 10μM FAD added to samples of E.coli cell suspension prior to lysis.

Strain/Matrix	FAD recovered [μM]	% recovery
MG1655	6.0	60.9
TG1	6.5	66.2
appA	6.3	64.2
cobB	6.6	67.5
cobT	6.0	60.5
Media (lysed)	8.0	81.0
Media (unlysed)	9.9	98.4

It was shown that regardless of the *E.coli* strain used approximately 60-68% of the FAD added (6.0μM – 6.6μM) could be recovered after sonication and analysis via HPLC.

It was concerning that approximately 35% of the FAD seemed to be lost but the consistency over the 6 strains suggests that measurements of cellular FAD previously made throughout and during the course of the study would be accurate for all strains used with a potential underestimate of ca. 35% for each. It was suggested that part of the loss could be explained by the degradation of FAD to FMN and riboflavin (discussed further

in section 5.8). To investigate this, fresh media with FAD added was subjected to sonication prior to analysis. A better result was obtained than with the cellular extracts but still 20% less than the FAD spiked into fresh media and analysed without lysis was recovered (see table 3.1). This suggested that part of the loss was directly related to the degradation of FAD during the sample preparation steps. On re-analysis of the chromatograms increased peaks relating to FMN and riboflavin were seen to account for almost all of this 20% loss. This suggests that degradation was caused by sonication but only accounts for half the loss seen in the cell extracts. It is known that there are a number of intracellular proteins that will utilise free FAD, proteins such as fumarate reductase and succinate dehydrogenase both found in *E.coli* have flavoprotein subunits which may be utilising the additional FAD (Brandsch, & Bichler, 1989). There may also be a small amount lost to non-specific binding. Since all samples were from plasmid free strains MTOX will not be present in large amounts to account for any significant loss. A certain amount of work has been done studying FAD stability in differing temperatures in blood plasma, these broadly agree with the trends suggested here (Akimoto, et al. 2006).

They showed a much higher rate of FAD hydrolysis to FMN at raised temperatures, (37˚C) than seen at low temperature, (4˚C). With FAD concentrations dropping to almost zero after 30 minutes at elevated temperature. This underlined the importance of refrigerating the samples whether during storage or during HPLC analysis where the auto sampler was also cooled to between 2 and 8˚C.

Akimoto also showed that predominantly all FAD is hydrolysed to FMN and the subsequent conversion to riboflavin is relatively slow (even at higher temperatures). This agrees with the findings presented in Figure 3.2 where the peak of FMN is significantly larger than the one attributable to riboflavin.

3.5. Summary

The experiments described here were necessary for the analysis of cultures expressing either GFP or MTOX proteins. These investigations were vital to develop and optimise two techniques that would be used throughout the course of the work.

It was shown that with the sonication method of cell lysis there was a trade off between maximum cell breakage and intracellular protein integrity. This resulted in a compromise to be used where high degree of lysis was achieved without affecting the integrity of the desired protein product severely.

Secondly the transfer and optimisation of an HPLC method for separation of FAD, FMN and riboflavin was shown. This was crucial to the further understanding and measurement of intracellular levels of the three compounds, directly relevant to riboflavin metabolism and the activity of MTOX. This work also gave an insight into FAD stability and altered the way samples were collected, stored and the timeframe in which they needed to be analysed.

Finally a spiking experiment was performed to determine what proportion of FAD could be recovered prior to HPLC analysis. A loss was seen throughout the sample manipulation but this was consistent across the range of *E.coli* strains tested. This loss was attributed partly to the degradation of FAD caused by the exposure to light and elevated temperatures during the sonication process, and partly to FAD binding to intracellular flavoproteins inherent in the *E.coli* cell. This level of loss has been accounted for throughout the following chapters, particularly were FAD levels estimated *in silico* are compared to experimentally derived values.

4. GFP recombinant strain as a model system

4.1. Introduction

The rational behind the work in this chapter was to provide data about the initial strain of *E.coli* used in the investigation and to develop methods that would be used throughout the course of the work. Experiments determining the growth characteristics of *E.coli* MG1655 (Blattner et. al, 1997) wild type strain and the level of native protein present were carried out to establish a baseline for bacterial growth before the introduction of a foreign protein expressing plasmid.

As mentioned earlier Green Fluorescent Protein (GFP) was chosen to act as a model protein product due to its ease of detection and plasmid availability. The plasmid construct p*uv*GFP was therefore introduced into MG1655 and differences in growth characteristics, total cellular protein production, carbon utilisation and GFP production were studied.

Shake flask experiments made up the bulk of the fermentations due to the ability of using these for a higher number of duplicate runs than would be possible with larger scale stirred tank reactors This fitted with the aims of developing a higher throughput method for screening and selection of mutant strains, differing media, and differing plasmid constructs that were foreseen at this early stage of the work. However data such as dissolved oxygen tension (DOT), off gas analysis or even automatic pH control is not available when using shake flasks. Therefore a number of 5L scale stirred tank cultivations were undertaken to allow comparison of the two systems.

4.2. Growth Characteristics and Plasmid effects on cellular growth in SFs

Initial experiments were carried out using the wild type *E.coli* strain MG1655 on M9 minimal medium (see Table 2.2 for constituents). As shown in Table 4.1, growth on this medium was slow and achieved low final cellular concentrations. Growth at this level was thought not ideal for the planned work. It has been discussed earlier that mutations made in the *E.coli* genome may only have a small effect on cellular growth therefore a medium that provided faster, more reliable growth to a higher final OD was sought.

The obvious choice for an alternative would have been complex media. This was initially rejected on the grounds that a detailed composition of such media was not available. Accurate amounts of carbon input were required for the planned modelling work (see Chapters 1 & 6) and therefore ruled out the use of undefined, complex media. Therefore M6 minimal medium was investigated as an alternative. It was shown that cultures on M9 grew to a significantly lower final OD than those on M6 (T-test result: 4.97 > 2.132: 95% df=4). The rate of growth was less convenient to analyse having non-equivalent variance. To overcome this problem ANOVA was employed to determine whether M9 values came from the same population as the M6 data. The test returned a critical F value of 5.99 (>5.208, F crit 95%, 6 degrees of freedom), and was therefore scored as significant.

Table 4.1) Maximum specific growth rates and confidence intervals for 4 differing growth investigations of E.coli MG1655 grown in shake flask cultures.

Strain and Media	μ_{max} (h^{-1})*	Mean OD (600nm)	Mean CV (95%) +/-
E.coli MG1655 WT, M9	0.42 ± 0.11	1.168	0.096
E.coli MG1655 WT, M6	0.62 ± 0.06	2.660	0.062
E.coli MG1655 puvGFP, M6 uninduced	0.56 ± 0.08	2.313	0.079
E.coli MG1655 puvGFP, M6 induced	0.55 ± 0.06	2.247	0.060

* = example calculation included in appendix 3

At this stage a plasmid construct p*uv*GFP (see materials and methods section 2.3) was introduced into the host strain. Growth of this plasmid carrying strain, with and without induction (1mM IPTG at an OD ~ 1.0), and the plasmid free strain, was compared by determining biomass concentration and growth rate.

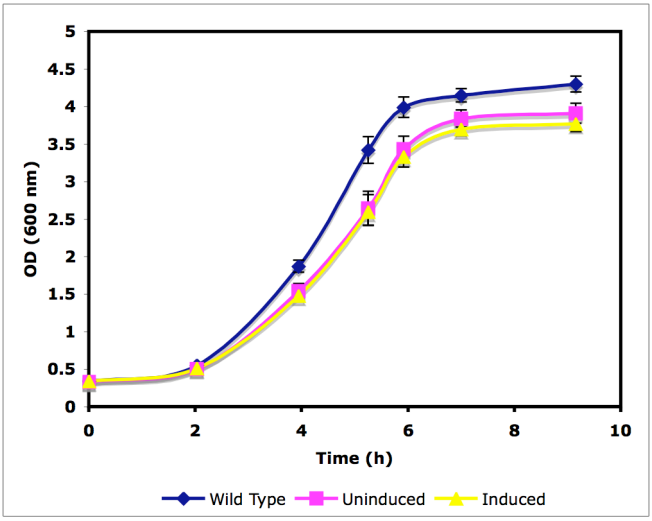


Figure 4.1) Growth profiles for three independent runs (n=3) of MG1655 WT and plasmid carrying strain on M6 media (induction with 1mM IPTG at T = 3h) grown in SFs.

It was noted that the variability of each shake flask run was generally reduced as is shown with a comparison of percentage CV in Table 4.1 above.

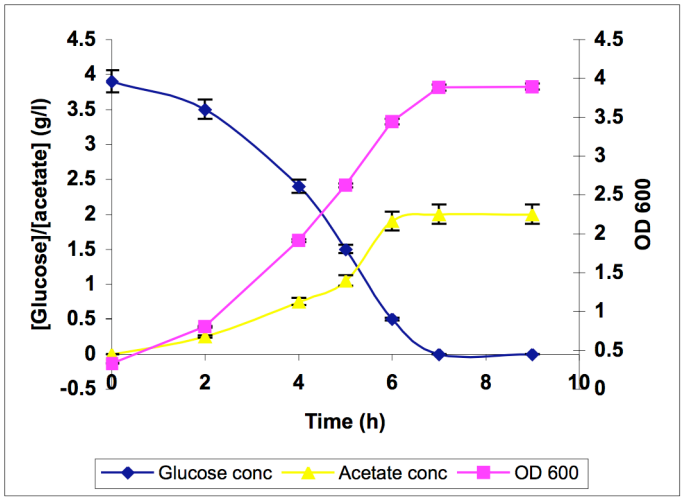
From the above data sets a baseline was produced for the growth of the non-plasmid carrying strain and the plasmid containing strain grown in M6 minimal medium. It was noticed that bacterial growth was slightly reduced in the plasmid carrying strain compared to the wild type strain, but no significant difference was seen when GFP protein production was induced with 1mM IPTG at a culture OD of approximately 1.

The initial shake flask fermentations provided some basic data on the growth characteristics of *E. coli* MG1655 wild type and plasmid containing strain. The growth curves in Fig. 4.1 show no significant difference in shape between the different strains tested. Generally, as would be expected, a lag phase is seen followed by a period of exponential growth which lasts until approximately 8 – 9 hours into the experiment. At the end of the fermentation a stationary phase is observed where the rate of cell division equals that of cell death and no increase in the broth turbidity is seen.

4.3. Analysis of carbon consumption and acetate formation of SF cultures

In an attempt to further constrain cellular growth in order to understand and observe potentially small changes originating from single gene deletions to be evaluated (see Chapter 5) the cultures were designed to be carbon limiting. In order to measure the amount of carbon source available at various points throughout the growth experiments, and to monitor the accumulation of acetate, a known carbon sink often produced when a cell is undergoing metabolic stress (Kirkpatrick et al., 2001); glucose and acetate levels were measured as described in Chapter 2, section. 2.5

Results in Figure 4.2 show that glucose levels dropped during the course of the shake flask fermentations due to carbon consumption by the organism. After 7 hours this carbon source has been exhausted which corresponds well with the beginning of stationary phase. Acetate levels rose over time reaching a plateau at 6 hours, again roughly corresponding with the end of exponential growth.



* Note all error bars are variation due to triplicate reading of a single sample

Figure 4.2) Growth profile of three independent runs (n=3) of MG1655 WT showing OD, glucose consumption and acetate production throughout a 10 hour shake flask fermentation in M6 medium.

Using sets of data for each set of shake flask samples and the calibration of OD to dry cell weight, yields of biomass on glucose and also acetate on biomass were calculated. (see Table 4.2).

Table 4.2) Summary of parameters characterising growth of MG1655 WT and MG 1655 puvGFP grown in shake flasks.

Strain and Media	μ_{max} (h^{-1})	X_{max} (g/l)	$Y_{x/glu}$ (g/g)*	$Y_{ac/x}$ (g/g)*
E.coli MG1655 WT, M6	0.62 ± 0.06	1.72	0.43	1.17
E.coli MG1655 puvGFP, M6 uninduced	0.56 ± 0.08	1.56	0.39	1.28
E.coli MG1655 puvGFP, M6 induced	0.55 ± 0.06	1.50	0.38	1.30

* = example calculations included in appendix 3

Values stated in table 4.2, seemed to suggest a small increase in acetate production between the plasmid free strain and that with the plasmid, also another increase between the uninduced and induced cultures however none of these were shown to be statistically significant, (values for the T-statistic were calculated between WT and puvGFP induced = 0.53 < 2.132 [P=0.95: d.f.= 4], values between the uninduced and induced strains = 0.09 < 2.132) It has been suggested that acetate production increases with rising metabolic stress (Hahm, et al., 1994) as acetate metabolism can be used as an additional source of energy and important metabolic precursors (Chang, et al., 1999) this result seemed to support that hypothesis.

4.4. Total Protein analysis of SF cultures

In order to quantify the amount of GFP produced, total protein analysis was carried out in the different strains used. This was performed on six *E.coli* culture samples of each type to ascertain the total amount of intracellular protein present in the wild type, plasmid uninduced and plasmid induced strains growing on M6 media.

The data is presented in the table below.

Table 4.3) Total protein in samples of E.coli MG1655 WT and MG1655 puvGFP uninduced and induced with IPTG (OD ~ 1.0) T=0 refers to the sample taken at the start of the fermentation, T=3 is taken immediately after induction with 1mM IPTG T=4 being one hour after this. The final timepoint T = 7 equates to the stationary phase of the culture

	Total protein conc [mg/ml]			
Strain	**0h**	**3h**	**4h**	**7h**
WT	0.83 ± 0.06	1.03 ± 0.41	1.12 ± 0.18	1.91 ± 0.09
Uninduced	0.64 ± 0.09	1.23 ± 0.15	1.62 ± 0.22	2.30 ± 0.24
Induced	0.61 ± 0.12	1.01 ± 0.10	1.75 ± 0.25	2.41 ± 0.13

Due to the nature of the low volume shake flask cultures it was not possible to analyse every time point of the fermentations. Since 1mL was the minimum sample size that was found to be adequately lysed via sonication (see Materials and Methods section 2.2.8), four time points were chosen and analysed.

When stationary phase was reached a small increase was seen in total intracellular protein between the plasmid free and the plasmid carrying strain, (Table 4.3). Performing a Student T-test on these two data sets gave a result of 8.80, above the tabulated value of 2.132 (df=4) and is therefore a significant difference at the 95% confidence level.

A less clear difference was seen between the uninduced and induced plasmid carrying strains. It was assumed that, with the production of green fluorescent protein the amount of total protein measured from the induced sample would be considerably higher than for the uninduced. This was not the case. On performing

the Student T-test on these sets of data a result of 4.23 is returned (n=3) and as such showed this to be a non-significant difference, (again P=0.95).

4.5. Verification of GFP production in SF cultures by SDS-PAGE

As verification that GFP was indeed being produced after induction SDS-PAGE was carried out and run with total protein and soluble protein for each of the conditions.

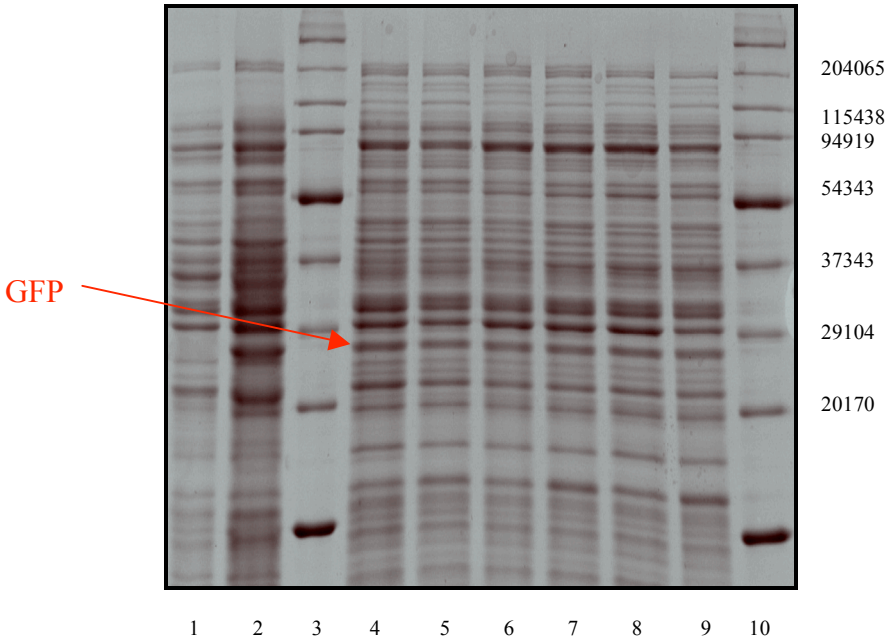


Figure 4.3) SDS – PAGE of MG1655 plasmid free (lane 1), and MG1655 puvGFP showing the 27kD protein. GFP Samples were taken from 3 hours post induction; the lanes were loaded as follows:

Lane 1: MG1655 plasmid free clarified cell lysate
Lane 2: MG1655 puvGFP total cell lysate sample
Lanes 3&10: Molecular weight ladder (Daltons)
Lanes 4 – 6: MG1655 puvGFP clarified cell lysate samples (uninduced)
Lanes 7 – 9: MG1655 puvGFP clarified cell lysate samples (induced)

At this stage, gels were used to verify the presence of the protein rather than to quantify the amount present so no further analysis was completed. A band was seen

at approximately 27kD daltons in all the sample lanes (numbers 2, 4 – 9), which was not seen in the plasmid free control (lane 1) suggesting the presence of green fluorescent protein, which has a molecular weight of 27 kD. There is also a band seen (although less clear than at 27kD) at approximately 54kD. This is thought to be caused by the propensity of UV optimised GFP to form dimers, a phenomenon not seen in other variants according to the manufacturer (Source: BD catalogue). However, rather unexpectedly a clear band can be visualised at 27kD for samples induced with IPTG and also those not induced. This implies a lack of control over expression of the construct used, and has been investigated further (see Section 4.10).

4.6. Determination of GFP expression in SF cultures by Fluorescence assay

The assay (described in Chapter 2, section 2.4.4) was designed to be a fast and effective way of comparing different *E.coli* strains expressing GFP. Compared to the Bradford assay for total protein or SDS-PAGE, this fluorescence assay could be performed simply and quickly for a range of cell lysates in a 96 well plate. Data was collected for non-lysed samples and, although the trends were similar to that obtained from lysed cells, the signal was much reduced. Using non-lysed cells would increase the throughput of the assay but the potential loss of precision was thought to be too great especially as very small differences in protein expression were expected with some of the deletion mutants to be used. Therefore only data from lysed and

clarified samples are shown in figure 4.4. All samples were obtained from 3 hours post induction with IPTG.

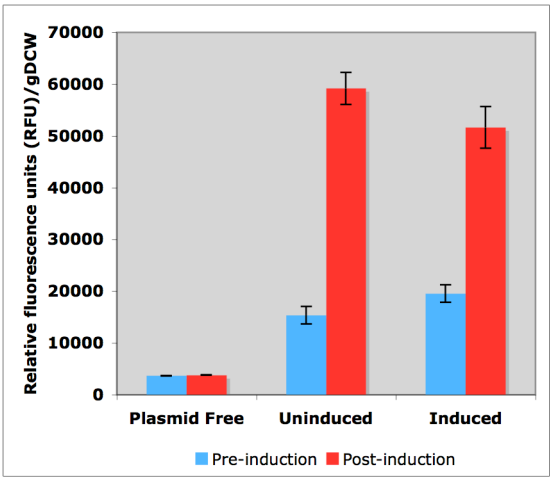


Figure 4.4) Fluorescent signals for three conditions of MG1655 puvGFP(un-/induced) and MG1655 (plasmid free) grown in SFs on M6 medium. Samples were taken immediately before and 3 h after time of induction.

Expanding upon experiments done to assess total protein production and presence of GFP, Figure 4.4 shows the difference seen in fluorescence signal at two points in the growth of *E.coli* strains. RFU was used as stated units as no reference standard for GFP was available and comparative data, all that was deemed valuable.

A low signal is seen from the plasmid free strain indicating that wild type MG1655 has a small amount of inherent fluorescence at the wavelengths used but this was less than 10 times the response seen in the other two samples, and was therefore not thought significant. No change is seen after the addition of IPTG and the one tailed T-test (see Materials and Methods, section 2.6) returns a value of 0.96. This is well below the threshold value of 2.132 (P=0.95: n=4) thus the differences between the

pre – and post – induction plasmid free strain samples are not significant and this acts as a valid control.

As the variance of the uninduced and induced data sets of the plasmid carrying strains was not equal, the heteroscedastic t-test was employed, (see Chapter 2, section 2.6). The obvious differences seen were confirmed with a value of 21.63 for the uninduced and 12.81 for the induced samples. Both of these values were above the critical value of 2.353 (d.f.=3: 95%) and therefore showed significant differences in the means of each sample group.

There were some interesting findings when the same time points for uninduced and induced samples of plasmid containing strains, were analysed. As expected (since no induction had taken place) the differences between uninduced and induced pre-induction were not significant ($T=2.51 < 5.841$ [P=0.95: n=3]). Again due to differing variances between the plasmid free and the plasmid carrying the heteroscedastic test was used and returned a result of 31.26, again above the critical value of 2.353 (d.f.=3 : 95%) showing a significant difference between the two. What was more unexpected is that no significant difference was found between the uninduced and induced samples, 3 hours post induction, ($T=12.36 > 5.841$ [P=0.95: n=3]). GFP seemed to be expressed regardless of induction with IPTG in these shake flask cultures. There was an assumed significant difference between the plasmid free strain and both the p*uv*GFP strains even before induction, and no significant

difference between the induced and uninduced strains 3 hours after induction had taken place.

As stated earlier experiments were mainly undertaken in shake flask culture however a limited number of stirred tank reactors were utilised to give a comparison between small and medium scale culture. The data from these experiments follows.

4.7. Characterisation of growth parameters in 7L stirred tank reactors

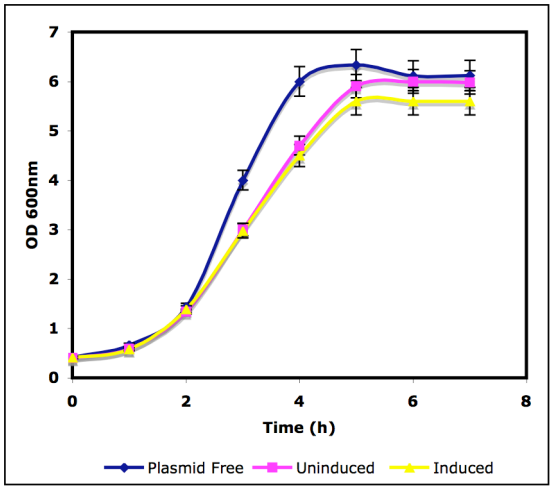


Figure 4.5) Comparative growth curves for plasmid free MG1655 and MG1655 puvGFP uninduced and induced grown in M6 medium in 7L stirred tank reactors, induction with 1mM IPTG occurs at T=2 h.

Compared to shake flask cultures a higher final OD, faster growth rate and a shorter lag phase was seen in bioreactor cultures. On the other hand, similar to results obtained in SFs there was a difference between the plasmid free and the plasmid carrying strains. Unlike earlier experiments a significant difference in growth rates is seen between the plasmid induced and uninduced strains as well as the plasmid free

control. However due to the time involved running these medium scale fermentations the sample population for each was too low (n=2) to quote statistical values.

As with earlier shake flask cultures the consumption of carbon source and the production of acetate were determined. The data is presented in Table 4.4.

Table 4.4) Summary of parameters characterising growth of MG1655 WT and puvGFP on M6 medium in 7L stirred tank reactors (STR)

Strain and Medium	μ_{max} (h^{-1})	X_{max} (g/l)	$Y_{x/glu}$ (g/g)	$Y_{ac/x}$ (g/g)
E.coli MG1655 WT, M6 (STR)	0.71 ± 0.090	2.53	0.63	0.88
E.coli MG1655 puvGFP, M6 uninduced (STR)	0.63 ± 0.065	2.40	0.60	1.00
E.coli MG1655 puvGFP, M6 induced (STR)	0.59 ± 0.035	2.24	0.56	0.95

The better control over culture temperature, pH and levels of dissolved oxygen (DOT) in STRs commonly produces faster growing batches which achieve higher final ODs and therefore higher yields of biomass per unit carbon (Michel Jr. et al., 1990). Furthermore, the yields of acetate were reduced compared to the shake flask cultures (see Table 4.2), potentially a sign of lower metabolic stress caused by growth conditions being closer to optimal (e.g. Kirkpatrick et al., 2001). The trend that was seen in shake flasks where acetate production was higher in the plasmid carrying strain relative to the plasmid free strain is borne out although no significant difference is seen between the two states of plasmid carrying cultures. In addition, a greater difference in $Y_{x/glu}$ between wt and plasmid containing strain was obtained.

In contrast to SF cultures there seemed to be clearer difference between un- and induced state that may suggest an additional metabolic burden and better control of GFP expression.

4.8. Total protein analysis of bioreactor samples

As with experiments run under shake flask conditions total protein analysis was carried out to determine the amount of total intracellular protein present in each sample. Table 4.5 data shows total protein data obtained after 3 hours of cultivation.

Table 4.5) Total cellular protein for plasmid free and two plasmid containing cultures in 7L stirred tank fermentations 1 hour post induction

Strain	Total Protein (mg/ml)
Plasmid Free	2.24 ±0.11
Uninduced	2.46 ±0.12
Induced	2.54 ±0.13

The trend is very similar to that seen earlier with the small-scale cultures. Analysis using the T-test is possible and comparing the protein yield of the plasmid free with that of the uninduced a value of 2.27 is obtained, this being above the threshold value of 1.943 (P=0.95) a significant difference was seen. Conversely, comparing the uninduced and induced data sets a value of 0.235 is calculated proving that there is no significant difference in the two means.

4.9. Determination of GFP expression in bioreactor cultures using fluorescence assay

In a similar way to shake flask, samples were analysed in triplicate for fluorescence and the data generated is shown below.

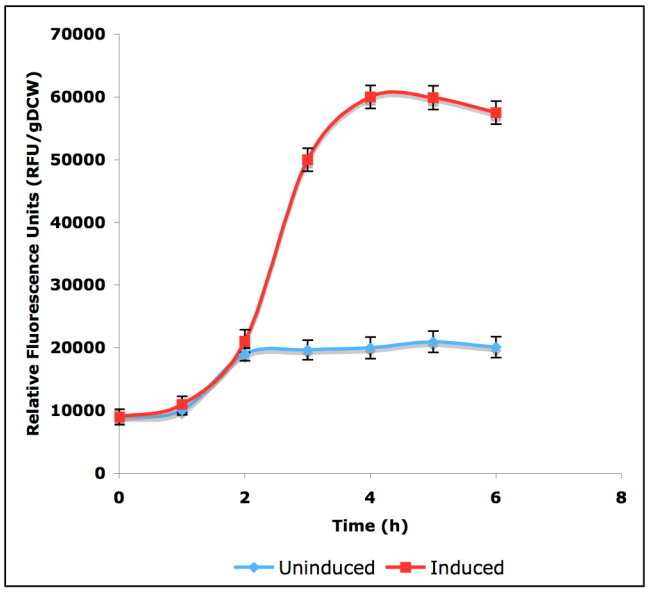


Figure 4.6) Comparative plot of normalised fluorescence for MG1655 puvGFP 7L stirred tank cultures in M6 medium (one induced at t=2h with IPTG, the other uninduced)

Fig. 4.9.1 shows a marked difference in fluorescence profiles over the course of the experiment between the induced and uninduced cultures. The possible reasons for these differences are discussed in section 4.11, however it was thought that the difference in growth rate seen may have been causing this effect. In fact, the most marked difference between the growth of MG1655 p*uv*GFP in shake flasks and stirred tank reactors was that of maximum specific growth rate and, to some extent

linked to this, final cell density. It was therefore decided that comparative runs of MG1655 p*uv*GFP would be undertaken using Luria Bertani broth (LB), in an attempt to increase specific growth rate to a level approximating that seen in stirred tank reactors.

4.10. Shake flask experiments in LB medium

These experiments were conducted in exactly the same way as those previously replacing M6 medium with LB. A comparison was then made between the fluorescence profile of the shake flask culture and that of the bioreactor cultures grown in M6 medium (see Fig. 4.7). In addition the specific growth rates and other parameters of the cultures grown under different conditions were compared (see Table 4.6),

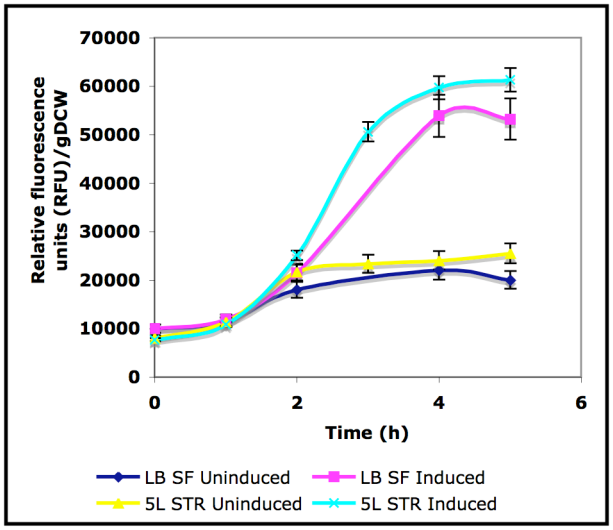


Figure 4.7) Comparison of normalised GFP fluorescence in cultures of MG1655 puvGFP grown in SFs and STRs.

The profile of fluorescence (implying the production of GFP) was seen to follow similar profiles for the shake flask experiments using LB medium compared with the data collected from the stirred tank experiments using a defined medium. All cultures show a slow increase in fluorescence from time 0 (where the seed culture was added) until 1 hour. At time point 1 IPTG was added at a concentration of 1mM and the two groups were seen to diverge. The uninduced cultures showed a slowing of increase of fluorescence, which reached a plateau 2 hours into the fermentation. The induced cultures however displayed an exponential increase in fluorescence between 1 and 3-4 hours, reaching plateau after 4 hours at approximately 3 times the level seen in the uninduced cultures.

Maximum specific growth rates, final OD values and yield coefficients were also determined, the results being tabulated in Table 4.6 below.

Table 4.6) Summary of growth parameters of MG1655 (WT& puvGFP) grown in two different media and culture conditions. N.d = not determined

Strain and Media	μ_{max} (h^{-1})	X_{max} (g/l)	$Y_{x/glu}$ (g/g)	$Y_{ac/x}$ (g/g)
E.coli MG1655 WT LB shake flask	0.70 ± 0.087	2.92	0.73	N.d.
E.coli MG1655 puvGFP LB shake flask uninduced	0.64 ± 0.095	2.80	0.67	N.d.
E.coli MG1655 puvGFP LB shake flask induced	0.60 ± 0.022	2.78	0.67	N.d.
E.coli MG1655 WT, M6 (STR)	0.71 ± 0.090	2.53	0.63	0.88
E.coli MG1655 puvGFP, M6 uninduced (STR)	0.63 ± 0.065	2.40	0.60	1.00
E.coli MG1655 puvGFP, M6 induced (STR)	0.59 ± 0.035	2.24	0.56	0.95
E.coli MG1655 WT, M6 shake flask	0.62 ± 0.064	1.72	0.43	1.17
E.coli MG1655 puvGFP, M6 shake flask uninduced	0.56 ± 0.082	1.56	0.39	1.28
E.coli MG1655 puvGFP, M6 shake flask induced	0.55 ± 0.063	1.50	0.38	1.30

The comparison of all nine sets of data showed significant differences in growth rates between M6 shake flask cultures and M6 stirred tank reactors (T=6.70 > 1.943, 95%, df. 6) and also between the M6 and LB shake flasks cultures (T=5.96 > 1.943, 95%, df. 6). No difference was seen between the three conditions in any of the 3 cultures. This showed that regardless of growth conditions the incorporation of the plasmid had no significant effect on growth parameters. However the trend was seen in each condition whereby maximum specific growth rates are seen to be reduced with the incorporation and subsequent induction of the recombinant protein. Significant differences in maximum biomass concentration were seen which followed the same trend seen for growth rates. It is a pattern of the data that wild type values were consistently higher than uninduced values, which were again higher then induced values however these differences were not statistically significant. Previously is has been stated that it was thought that this difference in growth rate might be causing the lack of expression control that can be seen in Figures 4.5, 4.6 and 4.7. A hypothesis explaining how a reduced growth rate could affect recombinant protein expression control is suggested and discussed in the following section.

4.11. Summary and Conclusions

The aims of this section of work were to set up a model bacterial system where growth characteristics could be investigated for the MG1655 strain of *E.coli* containing a GFP producing plasmid in a number of culture conditions. It also established a baseline for growth of *E. coli* MG1655 before and after transformation with the plasmid p*uv*GFP that could be used for comparison throughout the project. To answer these aims a minimal medium was chosen that displayed moderate and reproducible growth in shake flask culture, which was designed to be limited in carbon source. HPLC analysis indicated that the cultures were carbon limited after ca. 7 hours of cultivation. This limitation was thought to be necessary to analyse mutants with deletion in genes of central metabolism. *E.coli* is known to demonstrate a remarkable level of plasticity when carbon flux is altered, by mutation for instance, (Almaas et al., 2004), and therefore a certain level of carbon stress was desired to make the physiological effects of any central metabolic deletion more easily observable.

It is well recognised (e.g Seech & Trevors, 1991) that the incorporation of a plasmid constitutes a metabolic load to the cell, with the most often seen phenotypic effect being reduced maximum specific growth rate (e.g. Betenbaugh, 1987, Alldrick & Smith, 1983, Clerch et al., 1996). The data presented confirms this observation for two different types of media and two culture techniques, however there is a universal similarity in all cultures in that the apparent load (assumed by reduction in μ_{max} and

$Y_{x/glu}$) is greater between the plasmid free and the plasmid containing strain before induction than that apparent between uninduced and induced cultures of the plasmid carrying strains.

This may not be an unsurprising result in the light of the plasmid protein expression vector system utilised. Referring to the plasmid construction (materials and methods, section 2.3) it can be seen that p*uv*GFP is based on a pBR322 type plasmid containing an ampicillin resistance gene to allow selection for the successfully transformed cells and culture plasmid maintenance. The enzyme β - lactamase that confers resistance to penicillin-based antibiotics is constitutively expressed or antibiotic resistance would not occur. It can therefore be seen that the plasmid transformed strain is already expressing a foreign protein before any construct is induced. Therefore the slowing of growth rate between the plasmid free and plasmid uninduced strains can be attributed to the extra load of plasmid replication itself and also the production of β - lactamase. It is possible then that a smaller difference should be seen when a single extra protein is produced after induction but this will depend strongly on the expression level. Similar results were found by Roskov et. al. (2004) whilst attempting to describe changes in growth rate and overflow metabolism in a plasmid expressing strain of *E.coli* DH1.

This extra protein production is seen in analysis of total intracellular protein. It was seen that at the beginning of stationary phase, plasmid carrying strains produced approximately 30% more protein than the wild type strain. However this was not

continued when recombinant protein production was successfully induced. As shown in Figure 4.8.1 the trend in increased total protein is also seen between the plasmid free and plasmid carrying strains in stirred tank reactors. When GFP production is induced, and confirmed by the increase in fluorescent signal no significant difference is observed in total protein production. It is thought that in bacterial cultures producing a foreign protein, some native protein production can be shut down to allow the cell to divert resources into the formation of the new product (Terpe, 2006). It is suggested that this is the reason for trends shown in the data here.

The real interest however lies in the comparison of GFP expression based on fluorescence signals in different culture conditions. In the slow growing MG1655 p*uv*GFP in M6 medium in shake flasks; no significant difference was displayed between the uninduced and induced cultures. However, with the same medium growth rates increased by the use of stirred tank reactors and a marked difference was observed in GFP signal before and after induction with IPTG. The question was then raised whether this phenomenon was dependant on the growth rate or whether a feature of the reactor culture caused the results observed. A third set of experiments were then performed in shake flasks by using the complex Luria Bertani broth. Table 4.10.1 shows the similarities in growth rates between the bioreactor culture using minimal media and the small-scale shake flask culture using LB. In addition, Figure 4.10.1 shows a similarity of fluorescence profiles over the course of both fermentations. Taking together it may be concluded that the GFP induction is

affected by a decrease in growth rate regardless of other external variables. One hypothesis to explain this is based on the plasmid construct itself.

Control of recombinant gene expression is based on the lacI gene. In the absence of lactose, or its analogue IPTG, Lac I is produced and binds to the operator in the lactose operon. This prevents DNA polymerase moving along from the promoter region and therefore, if a gene of interest is inserted downstream, can prevent transcription of the recombinant gene. However it is proposed that as plasmid replication is not linked to cellular division, slow growth may see a relative increase in copy number from the 20 copies per cell expected. As the number of plasmid copies increases the amount of lac I in the cell may be insufficient to block all of the lac operator sites and therefore this could allow transcription of the GFP construct downstream. At higher growth rates the plasmid copy number stays generally constant and therefore only the addition of IPTG or lactose will allow the gene product to be expressed. It was thought that this could explain the fact that control of GFP expression was only observed at the higher growth rates seen in the stirred tank reactors and complex media shake flask cultures (Ward, personal communication).

In conclusion the results demonstrate the development of a robust model system to study recombinant protein production in *E.coli*. GFP was chosen as a convenient protein product due to its ease of detection and the plasmid system based on pBR322 is a proven vector platform. The data showed that M6 medium could deliver adequate and predictable growth leading to carbon limitation after about 7 hours.

The issue of plasmid expression control was important to understand and the impact on further work assessed. As has been stated, no significant difference was seen in the level of GFP expression between the induced and uninduced cultures using minimal media and shake flasks. It was suggested that this may be as a result of slow growth rates seen in such cultures, a hypothesis that was borne out by results in faster growing cultures either in stirred tank reactors or using complex media in shake flask conditions. Although these results need to be noted the impact on this work and subsequent investigations were thought to be minimal. Relative levels of GFP production (measured by fluorescence) were seen to be equivalent in all three conditions, the differences being only found in uninduced cultures that were growing more rapidly. For future investigation this lack of expression control should be noted as a characteristic of growth in minimal media shake flasks but protein expression has been shown to be not significantly affected by the lack of a tight control on the promoter system. Therefore the plasmid and growth conditions were thought to provide a solid foundation for further investigation into heterologous protein production with differing host strains, which are described in the following chapter.

5. Production of the flavoprotein MTOX in *E.coli*

5.1. Introduction

The following chapter describes the evolution of the model system developed in the previous chapter, in which the experimental set up was defined. It was shown that the strain of *Escherichia coli* MG1655 could produce a foreign protein through the incorporation of a plasmid vector and that, although it was thought the expression of the antibiotic resistance gene, the recombinant protein and the maintenance of the plasmid, caused a significant metabolic load, good growth and protein production was achieved. The phenomenon of uncontrolled expression in slow growing shake flask cultures was seen and an explanation put forward.

Based on these results the following chapter looks to build on this platform of *E.coli* expressing recombinant protein utilising the parallel approach of low volume shake flask cultures but with a better understanding of their limitations and continuing to use a small number of stirred tank reactors for comparison.

After reviewing results obtained in earlier chapters it was decided that as a progression of the work another protein target was required. This protein, namely N-methyl tryptophan oxidase (MTOX), was chosen as it was thought that the production and optimisation of the protein product, a member of an important group of enzymes, the flavoproteins, could have more potential benefit and may be of more use than GFP. Although GFP has many uses in molecular biology, and has proven itself to be a convenient protein to visualise and measure due to its fluorescent nature there is no commercial gain from designing a strain of *E.coli* that will overproduce the protein.

More important than the potential commercial interest in MTOX, it was noted that as MTOX was part of the group of flavoproteins found in many organisms there would be a specific pathway that could be targeted when mutant strains were chosen for the continuation of the study.

The group of flavoproteins rely on the binding of a cofactor, namely FAD or FMN. As described in Chapter 1 (section 1.5), the initial thought that single gene deletions in central metabolism could be shown to have significant effects on recombinant protein production was questionable, therefore a more directed approach was sought. The fact that flavoenzymes require a flavin cofactor to be bound to confer activity allowed a directed approach to investigating gene deletions as an obvious pathway to explore would be that producing riboflavin and its derivatives.

Further explanation of the rationale behind the move away from GFP and to MTOX is included in Chapter 1 including its potential uses in industry, structure, and function. However a brief overview will be included here for clarity.

Flavoenzymes such as N-methyl tryptophan oxidase have a flavin group (commonly flavin adenine dinucleotide or FAD) covalently bound to them that alters the characteristics of their active site, conferring enzymatic activity (Hassan-Abdalah, 2005). This chapter therefore describes experiments designed to measure the amount of MTOX protein produced by a number of *E.coli* strains and also to determine the proportion of expressed protein that is active, i.e. has a

FAD moiety bound. It also presents data from characterisation of strains containing mutations in the riboflavin pathway. Again covered in the initial chapter, mutations in the pathway directly responsible for the production and recycling of riboflavin, FMN and FAD were expected to have some effect on the amount of FAD available to the cell to bind to MTOX and produce a viable, active enzymatic product.

Various flavinated proteins exist and may prove to be industrially relevant. What was hoped was that information gained on the optimisation of production of one flavinated enzyme may be relevant to other proteins in the group as they all have the common necessity for FAD as their cofactor which confers activity.

5.2. Initial growth characteristics of the new strains and constructs

GSK donated a strain of *E.coli* (TG1) containing a MTOX expressing plasmid (pMEX8), which had been used previously for investigations into flavoprotein expression. However the planned modelling work and the mutant strategy, which are both explained in detail elsewhere, required the use of a K-12 strain such as MG1655. To this end a second plasmid construct was developed in house and was named pQR498, (for details see Chapter 2, section 2.3). Also although data was collected for strains in both M6 media and LB, the latter was used for much of the analysis due to the ability to control expression under conditions supporting faster growth. As the new plasmid was developed from puvGFP plasmid initial experiments sought to compare the growth rates of the 3 different

strain/construct combinations in use. The comparative growth rates are shown in Table 5.1 below.

Table 5.1) Comparison of maximum specific growth rates of 3 strain/plasmid combinations grown in SFs in LB medium (data includes n = 6 independent replicates)

Strain and Media	μ_{max} (h^{-1})
E.coli MG1655 pQR498, LB uninduced	0.64 ± 0.032
E.coli MG1655 pQR498, LB induced	0.62 ± 0.024
E.coli MG1655 puvGFP, LB uninduced	0.64 ± 0.095
E.coli MG1655 puvGFP, LB induced	0.62 ± 0.022
E.coli TG1 pMEX8, LB uninduced	0.67 ± 0.040
E.coli TG1 pMEX8, LB induced	0.61 ± 0.018
E.coli MG1655 pMEX8, LB uninduced	0.66 ± 0.056
E.coli TG1 pQR498, LB uninduced	0.67 ± 0.089

No significant difference was seen between the MG1655 pQR498 and MG1655 puvGFP constructs (induced) retuning a value of 0.0 with a one tailed test (Pcrit = 2.132, 95%, d.f.=4). This was expected as only the gene expressed was changed between the two constructs.

No significant difference was also seen in the induced state between the two different strains containing the same plasmid (pQR498), returning a value of 0.54 again less than the critical value of 2.132.

Finally no significant difference was seen when induced between either strain containing the other plasmid. GSK supplied a TG1 strain transformed with pMEX8; pQR498 being initially developed in MG1655. Thus it was shown a TG1 strain transformed with pQR498 would behave similarly to a MG1655

strain containing pMEX8. The T test returned a result of 0.1 (<2.132). Comparable values of growth rates are seen in both strains showing no bias to either plasmid construct.

5.3. Peroxide assay: Initial findings

As explained in Chapter 2, an assay was developed for the measurement of MTOX activity using the demethylation of N-methyl L-tryptophan to producing hydrogen peroxide as a by-product allowing colourimetric detection of peroxide to act as a signal for the presence of the enzyme.

With this reaction as the basis of the assay, specific units of enzyme production are difficult to calculate as there was no direct measure of enzyme concentration, only activity. Further work was undertaken in later fermentation runs that measured directly the amount of MTOX produced which could be related to activity to calculate specific activity. In the earlier work a simpler method was employed to gain a quick insight into the production of the enzyme. The method for quantification is explained in section 2.4 of materials and methods. For the following work some data is presented as concentration of peroxide evolved and some in Units of activity.

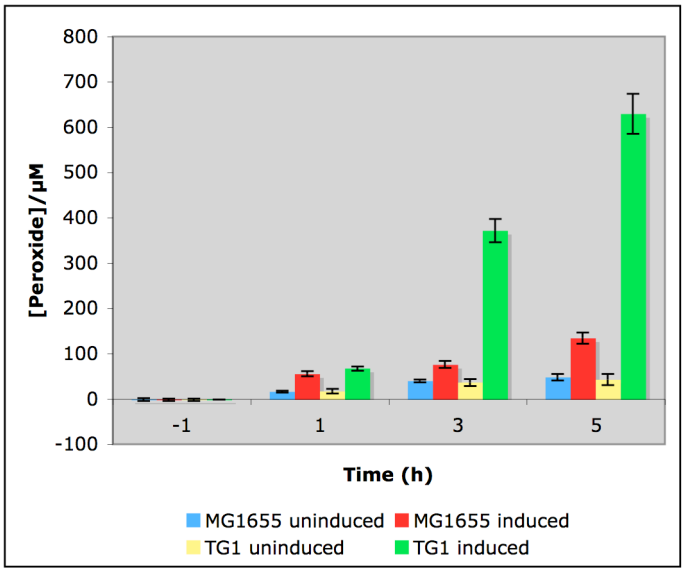


Figure 5.1) Comparison of peroxide production for MG1655 pQR498: LB and TG1 pQR498: LB in shake flask cultures (Time is pre/post induction with IPTG at time = 0, n = 6 independent replicates)

Initial work (in Figure 5.1) showed that, as expected, the amount of peroxide produced increased over the time of the fermentation. The first experiment also showed lack of control of the promoter system, as peroxide levels increased significantly from baseline throughout growth without the presence of IPTG as an inducer.

The first comparison was made between MG1655 and the TG1 strain (see Figure 5.1. The graph shows time point -1 or one hour prior to induction, this was commonly the point the cultures were seeded but not exclusively. Following this, data collected 1, 3 and 5 hours post induction are shown. A marked difference between the two strains was seen. MG1655 was seen to exhibit low levels of MTOX activity, which increased steadily up to a maximum at 5 h post induction. TG1 however showed similar levels of activity 1 h after induction but then activity increased dramatically to a final level over 4 times that of MG1655. On

examining these results further investigations were thought necessary to form a more complete picture of MTOX production and activity in the cell.

5.4. Peroxide assay development

There were a number of questions that arose from the early MTOX assay work, the most important involving the activity of the enzyme itself.

As described in the introduction, MTOX, like all flavoproteins requires a change in conformation of its active site caused by the covalent binding of FAD, to confer activity. The assay therefore, by definition, was only measuring the active or holo enzyme. Work by Hassan – Abdallah (2005) showed that although enzyme activity is severely reduced in the absence of FAD, activity could be recovered in over 90% of protein molecules with the addition of FAD into solution.

With this as a basis it was decided that the assay should be repeated, with FAD added to one set of samples before analysis as described in Chapter 2 section 2.5. The hypothesis being any apo forms of the enzyme produced by the host would be converted to the active holo form and then become detectable by the assay technique. Data could then be produced comparing the total amount of enzyme produced originally and the proportion of that enzyme that was active at that time.

Other controls/assay developments were investigated, firstly including a short study investigating the minimum amount of FAD that could be added to give full conversion, secondly a cell breakage experiment to ensure the lysis conditions developed in Chapter 3 were indeed the most efficient at cell disruption and thirdly collection of data from later than 5 h post induction (generally 6 – 7 hours culture time) to confirm enzyme production did not continue to rise after stationary phase was reached. These are covered in detail in Chapter 3, "Method development".

Initially the conditions from the above experiment were revisited. Taking the sample from 3h post induction on Figure 5.1 and analysing with and without FAD addition the following was seen.

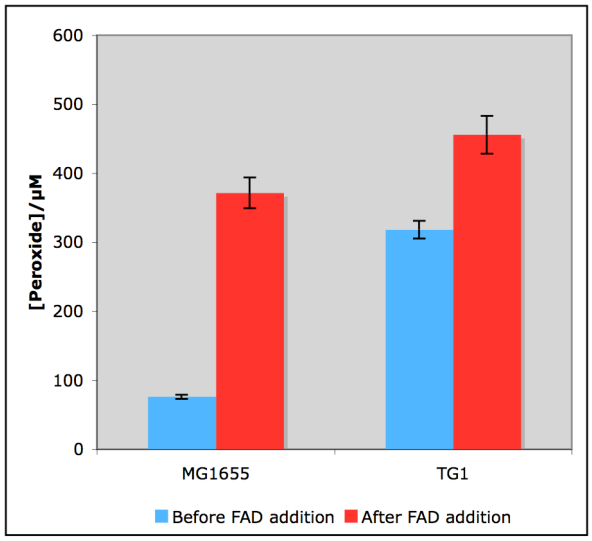


Figure 5.2) Comparison of MTOX activity (measured by peroxide production) in two strains of E.coli before and after addition of FAD (100µM)

It was noted that although the levels of MTOX enzyme activity was still greater in the TG1 strain the total amount of protein produced, irrespective of its activity,

was comparable between the two strains. The difference is significant between MG1655 with 100μM FAD added and TG1 with the same addition at the 95% level returning a result of 20.73 (> 2.132: df=4). It was therefore suggested that the TG1 strain has a greater pool of cellular FAD or a greater capability to supply FAD on demand than the MG1655 strain; this will be discussed further at the end of this chapter (section 5.11)

The hypothesis was then put forward that a host strain producing more FAD or having the capability to supply more FAD would not necessarily produce more MTOX protein but a greater proportion of that protein would be the holo or active form.

Before the mutant strategy was pursued, an easier option was considered. A cheap and convenient solution to the problem of supplying FAD may be to grow the strains on a media enriched with a FAD precursor such as riboflavin.

5.5. Growth on riboflavin enriched media

Riboflavin was added to LB media at a concentration of 1mM and the two comparative fermentations described in section 5.3 and 5.4 repeated in this medium. Riboflavin was used in favour of FAD in that it is relatively stable in solution and considerably less expensive than FAD. A industrial process with a riboflavin enriched media was thought to be economically viable where an FAD enriched media would not be.

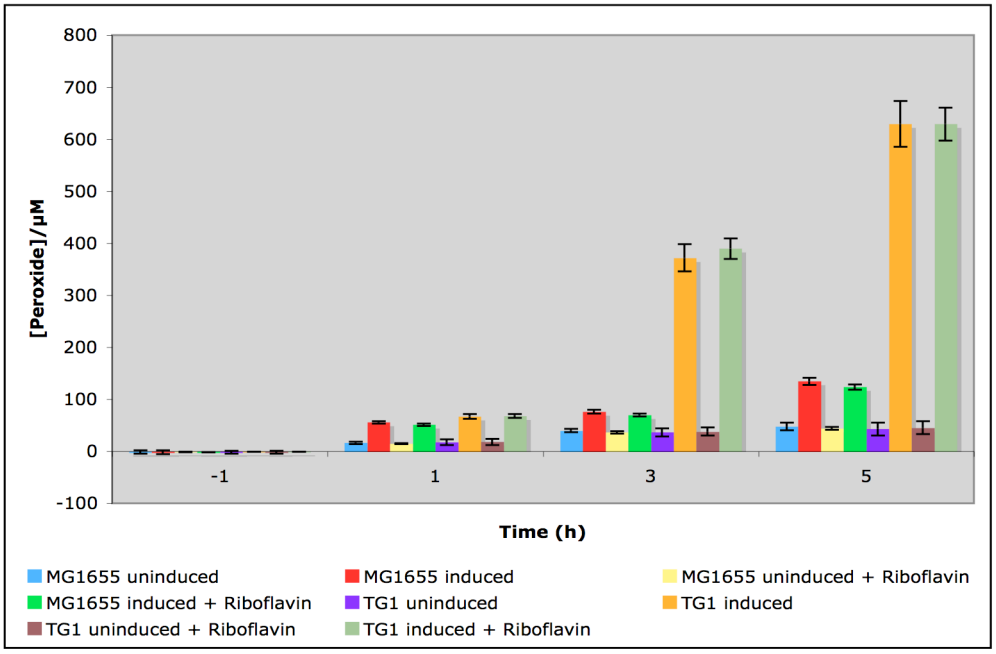


Figure 5.3) Comparative enzyme activity (measured by peroxide production) in two strains of E.coli using riboflavin enriched (1mM Riboflavin) media (n = 6 replicates)

No significant differences were seen between the media enriched with riboflavin and that without, whether induced or not induced. What this shows about cellular adsorption of riboflavin is discussed later (section 5.11) but what is clear is that the production of a FAD dependent enzyme cannot simply be improved by growth on a rich riboflavin medium. This was an important control experiment as it gives a basis for completing the more complex deletion mutant utilising strategy that will be detailed in the continuation of this chapter. It has been shown that the simple act of enriching growth medium does not lead to an increase in active MTOX production therefore other avenues needed to be explored.

5.6. Development and expansion of an experimental matrix

In response to the expanding nature of the plasmid, strain, media combinations a simple matrix was designed to better track the growing number of experiments carried out. An example of this matrix is included in Appendix 1.

The aim of much of the work was then altered to investigate the level of active enzyme production in a small number of deletion mutants from the Mori collection (Mori et al., 2000). As discussed in the introduction (section 1.5) three K-12 *E.coli* strains were used, MG1655 as discussed in Chapter 3, W3110 and BW25113 (used as host in the Mori collection). Mutations from the *E.coli* wild type are noted in section 2.2 of the materials and methods chapter.

The comparison between the first three strains was not only important for the protein production work but also the modelling work, which is described in the following chapter. Therefore, before any extended experimental plan was made, a short bridging study was undertaken to ensure similar growth was seen in W3110 and BW25113 before transformation with the plasmid construct pQR498, and then that similar levels of MTOX protein was produced both before and after induction.

The data from the study is summarised in Table 5.2 and shows no significant differences between any of the strains in any of the conditions investigated. On the basis of these results it was decided that earlier data generated with the

MG1655 strain would be comparable with the related strains, W3110 and BW25113 that would be used increasingly in the continuation of the work.

Table 5.2) Comparative growth rates and MTOX protein production for three strains of E.coli used in the study

Strain and Media	**μ_{max} (h^{-1})**	**X_{max} (g/l)**	**Units of MTOX activity (pre FAD)**	**Units of MTOX activity (post FAD)**
E.coli MG1655 uninduced LB shake flask	0.64 ± 0.1	2.8	0.002	0.003
E.coli MG1655 induced LB shake flask	0.60 ± 0.02	2.78	0.71	2.4
E.coli W3110 uninduced LB shake flask	0.66 ± 0.1	2.97	0.004	0.006
E.coli W3110 induced LB shake flask	0.64 ± 0.2	2.78	0.77	2.3
E.coli BW25113 uninduced LB shake flask	0.64 ± 0.3	2.66	0.003	0.004
E.coli BW25113 induced LB shake flask	0.64 ± 0.4	2.59	0.86	2.7

As described in the materials and methods, units of activity in this context refers to a measurement of MTOX activity defined by the rate of peroxide production. Wagner describes this rate as "One unit of activity being…the formulation of 1μM peroxide/min at 25°C following standard assay procedures". Prior to analysing the protein makeup of the cell lysates via SDS-PAGE this measurement was used to give an accurate measure of the level of expression and activity of the MTOX enzyme.

5.7. Mutant strategy

The strategy for selection of single gene deletion mutants is described elsewhere, (Chapter 2.2.3) and the use of mutants lacking the genes, appA, cobB and cobT explained.

The emphasis now altered from evaluating differences in growth rates caused by the incorporation of the plasmid as it had been shown that, TG1 aside, all host strains grew at approximately similar rates and produced comparable amounts of MTOX protein with statistically equal amounts of the active and inactive forms expressed.

The initial work calculated the enzyme activity over triplicate shake flask fermentations for 7 strains with samples taken at 1, 3 and 5 hours post induction with a final value stating enzyme activity after adding 100μM FAD to the final sample (5 hours post induction). This then gave not only a maximum enzyme activity at the end of fermentation, but also the proportion of enzyme bound with FAD at that time.

The data from these initial experiments is presented in Table 5.3.

Table 5.3) Units of MTOX activity for seven E.coli strains carrying the pQR498 plasmid (with the exception of the wild type control, BW25113)(all figures are dimensionless)

Time points / Strain	1h	±SE	3h	±	5h	±	Final (+ FAD)	±
MG1655 pQR498	0.38	0.020	0.67	0.040	0.71	0.057	2.44	0.034
BW25113 pQR498	0.44	0.034	0.79	0.056	0.86	0.099	2.68	0.100
appA	0.20	0.010	0.47	0.023	0.49	0.050	1.84	0.076
cobT	0.75	0.044	1.57	0.089	1.70	0.077	2.78	0.099
cobB	0.54	0.001	1.14	0.032	1.20	0.085	2.92	0.050
TG1 pQR498	1.12	0.089	2.07	0.076	2.10	0.053	3.00	0.089
BW25113 wt	0.001	N/A	0.002	N/A	0.002	N/A	0.003	N/A

Observing the data, the trend shown in Figures 5.1 and 5.2 is underlined with the TG1 strain producing a higher enzyme activity at the end of fermentation before the addition of FAD when compared to the MG1655 strain.

All other strains (coming from the Mori collection) are based on BW25113 and part of the data generated in the bridging study mentioned in section 5.6 is included. From this no significant difference can be seen between the MG1655 and BW25113 strains carrying the plasmid, in terms of initial enzyme activity, the figure of 0.65 being calculated from the T-test (<2.132, 95%: d.f.=4). Also the capacity for enzyme production (when flooded with FAD) seems alike as the T-test returns a value of 1.12, again a non-significant result.

Further analysis of the data was deemed unnecessary, as the aim was to generate a specific activity by combining the data above with information on the total amount of MTOX protein produced in each strain.

In light of this and following the peroxide assay findings, gels were run (as described in the materials and methods) with the clarified lysates from time points 1, 3 and 5. A protein standard of BSA (66.4 kDa) was used and total MTOX protein concentration of the lysates could be calculated using the standard curve generated. This data (quoted in mg/mL) is shown in table 5.4.

Table 5.4) MTOX protein concentrations in clarified lysates of 3 separate E.coli strain fermentations (mg/mL)

Time points / Strain	1	±	3	±	5	±
MG1655 pQR498	2.00	0.120	3.47	0.170	4.12	0.090
BW25113 pQR498	2.30	0.189	3.99	0.108	4.54	0.177
appA	1.24	0.111	2.57	0.011	3.22	0.990
cobT	1.87	0.078	3.55	0.123	4.67	0.200
cobB	1.67	0.065	3.01	0.189	4.89	0.199
TG1 pQR498	2.00	0.098	3.67	0.077	4.98	0.077
BW25113 wt	N/A	N/A	N/A	N/A	N/A	N/A

With these two sets of data it was therefore possible to calculate a specific activity for each of the protein expressing strains in terms of Units of activity (table 5.3) per milligram of MTOX protein present (Table 5.4). This calculation produced data that was truly comparable across all strains as it is displayed in specific units of activity, units/mg.

Table 5.5) Specific units/mg of MTOX activity over all E.coli strains

Time points / Strain	1	±	3	±	5	±	Final (+ FAD)	±
MG1655 pQR498	3.78	0.303	3.84	0.298	3.44	0.287	11.86	0.305
BW25113 pQR498	3.83	0.432	3.98	0.301	3.78	0.207	11.81	0.439
appA	3.23	0.331	3.63	0.180	3.01	0.324	11.43	0.591
cobT	8.02	0.577	8.85	0.588	7.29	0.454	11.93	0.622
cobB	6.47	0.252	7.58	0.521	4.90	0.401	11.94	0.713
TG1 pQR498	11.20	1.046	11.25	0.477	8.42	0.250	12.05	0.430
BW25113 wt	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

The data shows some interesting trends with the six strains falling into two distinct groups in relation to specific MTOX activity.

Firstly it was noted that the capacity for active MTOX production was not significantly different across the range of mutants. Although it has been shown (in Tables 5.3 and 5.4) that there were significant differences in the amount of holo enzyme between a number of the mutants, after the addition of FAD the differences became noticeably smaller underlining what was seen early in the investigation and shown in Figure 5.2.

Apart from the appA mutant all strains showed equivalent levels of total MTOX protein production at the end point of the fermentations. Differences were observed where MG1655, BW25113 and BW25113 appA (hereon referred to as group 1) displayed levels of specific activity ca. 30% below the maximum activity, calculated from the samples with FAD added prior to assay. In comparison cobT, cobB and TG1 (group 2) displayed activity between 41 and

70% of the maximum level throughout the relevant cultures. Another feature of the fermentations seemed to be that for group 1, specific activity seemed constant over the time course of the fermentation. Group 2 however displayed very high levels of specific activity early on in culture growth, which seemed to drop late on in the fermentation. Potentially this data suggests Group 2 has a higher capacity for supplying FAD and conferring activity on the MTOX protein produced but that this supply is becoming limiting towards the end of fermentation where MTOX levels are highest. Group 1 strains showed a generally more stable specific level of activity suggesting although activity is lower; any feedback mechanisms used to replenish cellular FAD are adequate to meet the demand put on them. Little is known about the specific mechanisms of regulation of the riboflavin pathway in prokaryotes but it has been suggested that a type of feedback mechanism may be in place utilising an ancient piece of cellular machinery known as the riboswitch, (Vitreschek, et.al, 2004). Riboswitches are thought to regulate a number of systems; riboflavin and thiamine metabolism as examples, by undergoing conformational changes when bound to effector molecules. It is thought that this change often triggers the transcription of an mRNA coding for either a biosynthetic or transport protein that increases the level of the relevant metabolite (Fischer et al., 2005).

In the data presented in Figure 5.5 below there is a noticeable rise in the levels of intracellular FAD after the induction of MTOX in all but one of the host strains. This may well be evidence in the experimental data collected for some kind of FAD feedback mechanism although whether this is linked to a riboswitch is unknown.

One important observation is that one of the chosen mutant strains (cobB) seemed to be able to compete with the TG1 strain in terms of MTOX activity throughout the course of the culture. Possible reasons for this are discussed later (section 7.3) in the light of further investigational work.

5.8. FAD analysis of all strain samples

Fermentations were carried out with all six strain/mutant combinations previously used and samples taken at various time points throughout growth. Figure 5.4 below summarises the results from these fermentations for each strain without plasmid and with plasmid both uninduced and induced. The data presented is from the mid exponential phase of all cultures, therefore 3-4 hours into the fermentations.

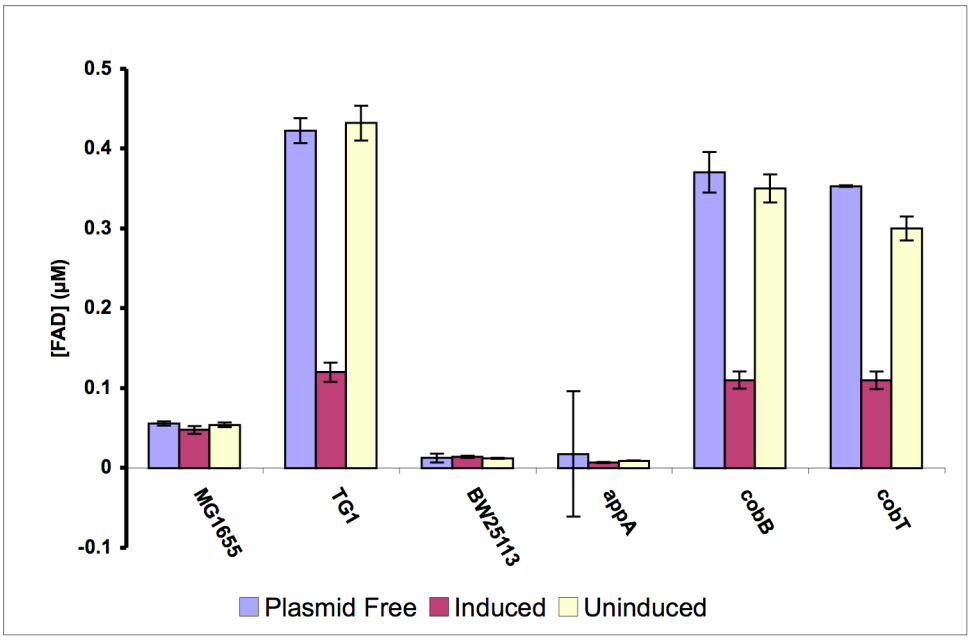


Figure 5.4) Comparative plot of FAD levels in mid exponential phase samples(3-4 h of fermentation) of 6 E.coli strains (n = 6 replicates)

The data gives a clearer picture of the two groups of cultures as described in section 5.7 of this chapter. Examining the first set of data for the plasmid free strains it was clear MG1655, BW25113 and the appA mutant all contained significantly lower amounts of unbound intracellular FAD than the TG1 strain and the cobB and cobT mutants. No significant differences were seen throughout the two groups themselves.

There were also no significant differences in any value comparing plasmid free and uninduced, which was expected. After the induction of the plasmid and expression of the MTOX protein, FAD levels drop significantly in the TG1, cobB and cobT mutants again as may have been expected in light of the findings in section 5.7. These three strains showed a significantly higher level of enzyme activity without a significant increase in the actual level of MTOX expression. This suggested that a greater proportion of MTOX in these strains is bound to FAD and therefore active. The finding that free FAD levels are elevated in cobT, cobB and TG1 strains of *E.coli* is therefore an expected one. Although there was a general trend of a slight decrease in FAD concentration in the MG1655, BW25113 and appA, these losses were too small to be significant.

5.9. HPLC analysis over fermentation time courses

Once the HPLC method was established shake flasks fermentations were run with more frequent sampling time points to establish a profile of cellular FAD levels over the course of a 6 – 8 hour fermentation.

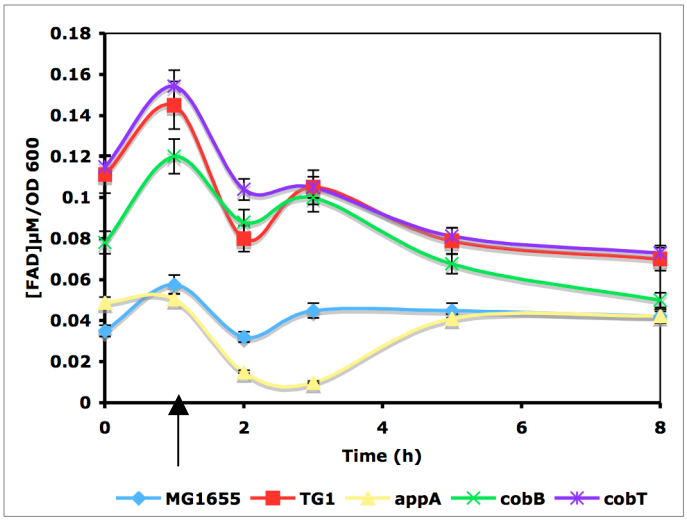


Figure 5.5) Free intracellular FAD levels (normalised to Biomass conc.) in 5 strains of E.coli over the first 6 hours of MTOX fermentation (arrow indicates induction at T = 1 h with 1mM IPTG)

A pattern was seen in intracellular FAD levels seemingly in agreement with earlier data. All strains increased the level of free FAD available in the cell from the time the seed culture was introduced into fresh media (T = 0) until induction via IPTG added at T = 1. During the first hour of MTOX production FAD levels were depleted in all strains, however all except the BW25114 appA mutant increased FAD levels during the following hour suggesting a strong feed back system to be present.

After this initial increase and the continuing demand for FAD the two groups as described in section 5.8 were seen to differ. Although the profiles of FAD levels were similar in all strains TG1, cobB and cobT were all seen to contain higher initial levels of the cofactor, which increased with cellular growth until MTOX induction. This caused a drop in free FAD as it bound to the protein which partially recovered 2 hours after induction. However levels then decreased gradually for the rest of the fermentation.

The MG1655 and appA strains also showed an increase in FAD levels before induction and a reduction after, however the recovery was significantly slower in appA and seemed to plateau at approximately 0.5 μM FAD/OD 600nm. MG1655 reached this plateau 1 h after induction, the level of FAD being constant throughout the rest of the fermentation.

5.10. MTOX production in 5L stirred tank bioreactors and comparison with SF fermentations

As was the case in chapter 4 a small number of 5L stirred tank fermentations were undertaken to determine any differences from the previous data set caused by the differing culture conditions. For the comparison only 3 strains were investigated and all were induced at 1 hour after the addition of the seed culture into the reactor vessel. The three strains thought to be most interesting after reviewing data from the shake flask fermentations were the original MG1655 strain containing the pQR498 plasmid as much of the baseline data was generated with this construct. Secondly the TG1 strain was used as a comparison

as it was the only non-K12 strain in use throughout the work. Finally the cobT mutant strain of BW25113 carrying the pQR498 plasmid was used as it proved to be the most effective construct for the production of MTOX in the shake flask experiments.

Table 5.6) Comparison of 4 parameters related to MTOX production in 3 strains of E.coli under stirred tank and shake flask conditions

Strain and Media	μ_{max} (h^{-1})	Mean Units of Activity	Mean Specific activity	Mean FAD level (µM)
E.coli MG1655 pQR498 (STR)	0.64	0.77 ± 0.10	3.78	0.06 ± 0.01
E.coli TG1 pQR498 (STR)	0.67	2.41 ± 0.25	10.55	0.15 ± 0.03
E.coli BW25113 cobT (STR)	0.65	1.74 ± 0.20	8.43	0.09 ± 0.04
E.coli MG1655 pQR498 (SF)	0.56	0.58 ± 0.05	3.69	0.04 ± 0.01
E.coli TG1 pQR498 (SF)	0.58	1.76 ± 0.44	10.29	0.10 ± 0.02
E.coli BW25113 cobT (SF)	0.55	1.34 ± 0.29	8.05	0.07 ± 0.01

From the data obtained a pattern is clear although statistical analysis was impossible due to the small sample population size, all stirred tank data was generated from n = 2 independent runs and all shake flask n = 6. The trends seen match those also shown in section 4.10 where a significant difference in growth rates was seen between shake flask cultures of *E.coli* and stirred tank fermentations. All three strains exhibit faster growth under stirred tank conditions when compared to small-scale shake flasks.

Relevant to MTOX production three other parameters are compared. Firstly the units of activity of MTOX enzyme, averaged over the course of the fermentation (1, 3, 5 and 7 hours after induction). These show an increased enzyme production in stirred tanks as expected due to the faster growth and higher final biomass. Specific units of activity are therefore quoted with the figures normalised by the amount of total MTOX protein produced. These figures show a minimal increase

in specific MTOX production over the two fermentation regimes suggesting specific MTOX production may be more strain specific than growth rate/regime specific.

The cellular FAD levels seem to be consistently higher in the stirred tank fermentations, although as stated above the sample size prevented any statistical analysis of the data, however this may be expected with the cells being in a less stressed state than seen in shake flask cultures (possibly due to poor pH control, poor oxygenation or other drawbacks of the culture technique (Anderlei, et al., 2001, Garcia-Arrazola et al., 2005, Büchs, 2001). It is this increased level of intracellular FAD that may account for the increased levels of MTOX activity, however it seems that the amount of FAD that can be bound as a cofactor is determined not by the level of FAD available but is strain specific.

5.11. Summary and Conclusions

A natural progression from model system to a group of products with greater potential for commercial interest and also the opportunity to investigate factors affecting many cofactor- requiring enzymes was the driving force behind the switch from GFP to the MTOX protein. It was possible to determine the impact of the host strain on the supply of cofactors and whether different levels of cofactors, or their precursors, in the cell could affect active enzyme production.

As a comparison, initial experiments were similar to those run for the GFP construct. Maximum specific growth rates were calculated from measurements

of optical density over time for a number of strain/plasmid combinations. No significant differences in growth rates were seen in MG1655 or TG1 expressing either the pMEX8 or pQR498 plasmid constructs proving no plasmid bias in the testing. Also equivalent values were obtained for the pQR498 and puvGFP constructs as expected considering the only change was the protein of interest.

Without a conveniently fluorescing protein product, an assay was developed to measure the activity of the expressed MTOX enzyme, reliant on its ability to demethylate methyl tryptophan producing peroxide that could be measured, this is covered in more detail in chapter 2.

Initially a low level of constitutive expression was seen but upon induction raised levels of enzyme were measured in both strains starting from a low level early in fermentation to relatively high levels 5 – 7 hours post induction. The TG1 strain showed a level of enzyme production initially 4 times that of MG1655 but it was quickly realised that the assay, as performed, would only be measuring active enzyme and may be insensitive to large amounts of expressed protein present but without the FAD moiety bound conferring activity.

In answer to this problem the assay was improved to measure the amount of active enzyme as a proportion of total enzyme present, which gave some interesting findings. The large difference in protein levels seen in figure 5.1 between MG1655 and TG1 was much reduced when total enzyme was assayed. Although TG1 was still producing significantly more MTOX protein than MG1655 it was shown that the latter had produced similar amounts of the

enzyme but the majority of this protein being in its inactive form. This suggested either TG1 had a larger capacity for storing FAD or that it may react faster to a demand for FAD than MG1655. Work has been done on what is known as the Rib operon (Mack et al, 1998) and also the role of flavokinases for the conversion of riboflavin to FAD/FMN (Bacher, 2001) but little has been written on the differences seen in FAD supply in different strains of the same organism.

On this basis, and after exploring the possibility that a simple enrichment of the culture medium with riboflavin proved ineffective, the Mori collection (Mori, et al., 2001) was searched for deletion mutants deficient in genes thought important to the production of riboflavin or its conversion to FAD.

Three deletion mutants were chosen for further study based on three important genes in the Riboflavin pathway, appA, cobB and cobT. The choice is discussed in Chapter 1.

With the increasing number of strains/mutations it was important to have a robust method of comparison between results. The chosen method is fully described in Chapter 2 section 2.4.5 and allowed direct comparison of enzyme production in each strain. The results showed two distinct groups; MG1655, appA and BW25113 all showed low levels of specific activity as shown in Table 5.5 and the cobT, cobB and TG1 strains that displayed significantly higher levels of enzyme activity than this first group. This then suggested a direct link between MTOX activity and the supply of FAD in the cell. Although directly enriching the growth media had no impact on enzyme activity (see section 5.5) it seems the

intracellular level of FAD can have a direct impact on protein activation in this case.

MG1655 and BW25113 were both unaltered host strains compared to appA (deficient in the enzyme catalysing the back conversion of FMN to riboflavin), and cobB and cobT (both involved with the degradation of riboflavin into ultimately vitamin B12 (Fischer, et al., 2002)). Therefore an HPLC method was developed to directly measure the amount of FAD in the cell at specific times throughout the fermentations and also discriminate it from riboflavin and FMN.

Although the method was shown to have some drawbacks, most notably the low level of spiked FAD recovery (see method development chapter) it underlined the differences discussed above clearly. The low producers of active MTOX (MG1655, BW25113 and the appA mutant) all had very low levels of FAD available at mid exponential phase (see figure 5.4) compared to the much greater values determined for the TG1 strain and cobB and cobT mutants. The amount of FAD remained roughly constant for those strains producing mainly apo enzyme regardless of the level of expression seen (Group 1). Those strains that were producing mainly holo enzyme (Group 2) however displayed a dramatic drop in the level of free FAD when MTOX was expressed. Useful as this snapshot was, multiple samples were then taken throughout the time course of the fermentations the aim being to see how FAD levels fluctuate over an MTOX producing culture.

As shown in Figure 5.5 the holo enzyme producing group saw a sharp increase in FAD levels between the time seed cultures were added until protein expression was induced at t = 1. Following that there was a sharp decrease until presumably a feedback mechanism such as the riboswitch suggested by some (Vitreschek, et.al, 2004) registers a lack of FAD (Mack. et al. 1998) and levels started to recover. From that point on however MTOX is still being produced and is still utilising the supply of FAD, therefore after 3 hours there was a steady decline in FAD levels until a plateau at about 0.08 μM for TG1 and cobT or 0.06 μM for cobB was reached. Whether this is a baseline level of FAD specific to each strain is unknown however it does seem to support the slight differences seen in specific activity seen at the end of fermentation in the cobB and cobT strains (see Table 5.5). In comparison the MG1655 strain follows the same pattern of increase, use, feedback but then quickly reaches a plateau significantly lower than any of the holo enzyme producing strains. The appA mutant is the only exception to this profile displaying a much elongated feedback recovery. It is feasible the deletion in the FMN – riboflavin back reaction may have affected the ability of the cell to react to a FAD demand. This result was not investigated further but may be of interest in future studies.

Finally a small number of fermentations were carried out in 5L stirred tank reactors for comparison with the above data from shake flask cultures. Although insufficient runs were completed to be analysed statistically they provide a useful insight. As shown in chapter 4, maximum specific growth rates were expected to be higher in the stirred tank experiments than in shake flasks. Potentially linked to this faster growth and a higher final biomass level the mean enzyme activity

over the culture time was increased compared to the shake flask runs. Following this trend mean levels of FAD were also increased. However, the mean specific enzyme activity was essentially unchanged over the life of the culture. Regardless of FAD levels or amount of enzyme produced, specific levels of active enzyme remained fairly constant for each strain.

In conclusion it has been shown that levels of active MTOX enzyme can be increased by the use of certain host strains or specific deletion mutants in the riboflavin pathway. TG1, a strain that has been used in the past for recombinant protein production studies (Hoffmann and Rinas, 2004) showed high levels of FAD and subsequent high levels of active enzyme produced. For the mutant strains the activity of cobT and cobB mutants was interesting as they both displayed similar levels of free FAD and consequentially similar levels of active MTOX enzyme as TG1.

The function of cobT is to synthesise the lower ligand of vitamin B12 (Chen, 1995) encoding the enzyme 5,6, dimethylbenzimidazole (DMB). cobB codes for a phosphoribosyltransferase also involved with vitamin B12 production. However Chen (1995) stated cobB seems to require a high concentration of DMB for activity. He went on to say that cobT has been seen to possess some ability to transfer ribose phosphate, as well as cobB.

In light of this work the small differences in the cob mutants could be explained. In the case of the cobT mutant DMB is not produced and although cobB will produce phosphoribosyltransferase this will function with a lowered activity.

Thus carbon flux should be massively reduced downstream of riboflavin allowing the cell to partition more into the production of FAD, and therefore resulting in the higher specific activities seen in Table 5.5. In the case of the cobB mutant, it would seem cobT may take the role of both enzymes and although this may result in a partially reduced carbon flux through to B12 this reduction should be significantly less than in the previous case, specific activities are therefore increased with respect to the non-mutant strains but to a lesser degree.

As has been said, the appA mutant has no direct effect downstream of riboflavin and as such will not affect flux towards B12. Whether the removal of this enzyme reduces the ability of the cell to respond to lower FAD levels is a hypothesis put forward but not proven currently.

The above work therefore formed a significant data set describing growth, MTOX production and cellular FAD levels in a number of *E.coli* strains and with differing culture conditions. This was then used as an experimental basis for comparison with model data generated using the Palsson *E. coli* reaction network (Edwards & Palsson 2000, Wilback et al., 2004, Feist et al., 2007) and also investigating how increase in growth rate or intracellular FAD concentrations could alter the expression and activity of the MTOX recombinant protein product. The use of the model network and its comparison with laboratory-generated data is discussed in the following chapter.

6. Constraint based modelling for MTOX strains

6.1. Introduction

The metabolic model of *E.coli* used in the following work (Edwards and Palsson, 2000) was first developed by Palsson's group in 1990 and consisted of 695 genes, coding enzymes responsible for 720 individual reactions involving 436 metabolites. This in turn is based on published information about the annotated genome sequence of the K -12 strain of *E.coli* MG1655 (Blattner, 1997) and other physiological and biochemical information.

Metabolic modelling is becoming more prevalent and with the development of more sophisticated models, is becoming an important tool in metabolic engineering research. The benefits are clear, in that once a model is created and its designers are confident it can behave in a way comparable to the organism under experimental conditions, it can be used to perform many *in silico* investigations in a matter of hours, work which in the traditional laboratory may take days or even months.

It is partly for this verification, that the model is included in this work. The model can generate carbon fluxes (e.g. to biomass, to a specified protein, or to another metabolite), which can be directly related to information gained in the laboratory environment. A good level of agreement between the experimentally and computationally derived data would allow a greater level of confidence in predictions made by the model when investigating other reactions.

As previously stated, the model is based upon 720 reactions found in MG1655, therefore initial work was needed to adapt the reaction network to allow for gene deletions introduced in BW25113 (the strain that is the basis of the Keio collection) not present in MG1655. This was achieved by the removal of pathways in the model relating to gene deletions in the BW25113 mutants employed.

Following this it was possible to include the stociometric equation for the production of MTOX (see Appendix 2) and search for solution spaces using linear programming techniques that would maximise the production of the protein or maximise biomass production. As the model included the reaction networks responsible for the riboflavin pathway (detailed in Chapter 1), it was also possible to calculate the flux into FAD and how that might change when the other 2 output variables were altered. It was also possible to alter the carbon uptake rate and thereby constraining the model. This input variable could also be calculated from experimental work relying on the HPLC analysis of organic acids to determine the rate of carbon utilisation *in vivo*.

Finally the data generated was compared to the experimental data and any discrepancies discussed.

6.2. Effect of growth rate on MTOX and FAD fluxes

Firstly the relationship between cellular growth and MTOX protein production was explored. A range of specific growth rates up to a maximum suggested by the model constrained by the glucose uptake rate were used and the corresponding flux of carbon into MTOX was calculated.

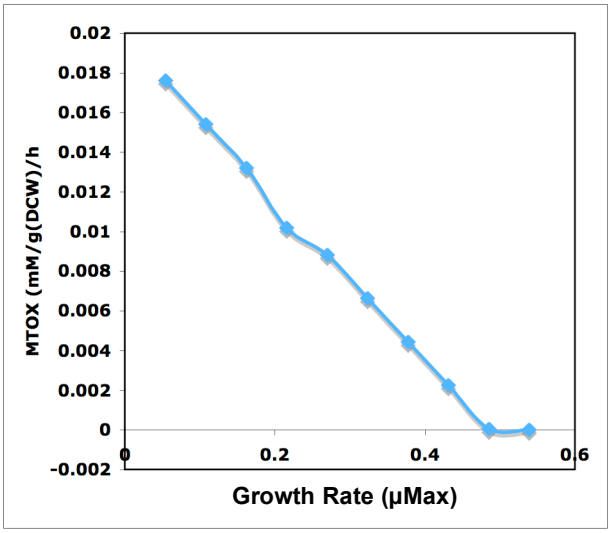


Figure 6.1) Plot of specific growth rate vs specific MTOX production rate predicted by the BW25113 model reaction network add glucose uptake rate used to constrain model

The model predicts that as the rate of growth increases the flux of carbon into the production of MTOX decreases approximately linearly to a point where maximum specific growth rate is approximately 0.5 h^{-1}, where no more MTOX is produced. This was based on the assumption of a glucose uptake rate of 6.6 mM/g(DCW)/h which was calculated from experimental data described in Chapter 5.

A second scenario was examined with no growth but estimating the possible maximum flux to FAD given a certain flux to MTOX. This simulation was thought particularly relevant, as 1 molecule of FAD is needed to confer activity to 1 MTOX protein.

As has been discussed earlier the cofactor FAD confers activity on the MTOX enzyme. For the work undertaken there would be no relevance for the optimisation of MTOX production without also considering the implications for cellular FAD levels. This simulation underlined the fact that from a finite supply of carbon there will always need to be a compromise between any number of carbon sinks present.

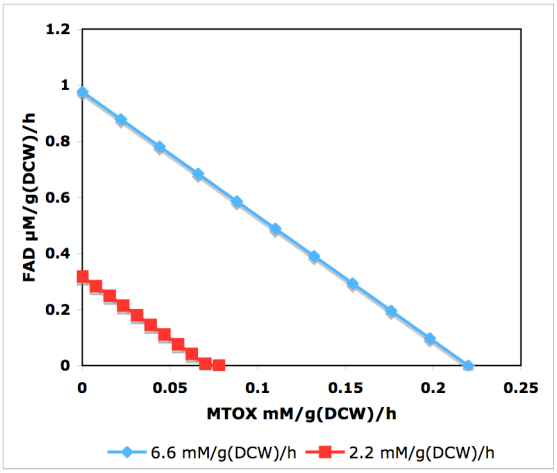


Figure 6.2.) Plot showing the predicted flux balance between MTOX and FAD for 2 different carbon uptake rates.

The initial glucose uptake rate of 2.2 mM/g(DCW)/h was the default setting on the model system. After analysing HPLC data it was shown that the strains were utilising carbon at an average rate 3 times higher, (see Figure 3.2). Fig. 6.2. shows as

expected an increase in available carbon with the higher uptake rate and also the partitioning of that resource between two carbon sinks. With the higher uptake rate a maximum FAD flux of close to 1µM/g(DCW)/h is predicted if MTOX is not produced and conversely the model predicts that 0.022mM/g(DCW)/h MTOX is achievable if FAD is not produced. As with all linear optimisation methods this produces a solution space which contains all possible outcomes in this partitioning of resource.

It was then attempted to generate a solution space that included all three variables of specific growth rate, flux to FAD and flux to MTOX in one simulation (see Fig. 6.3).

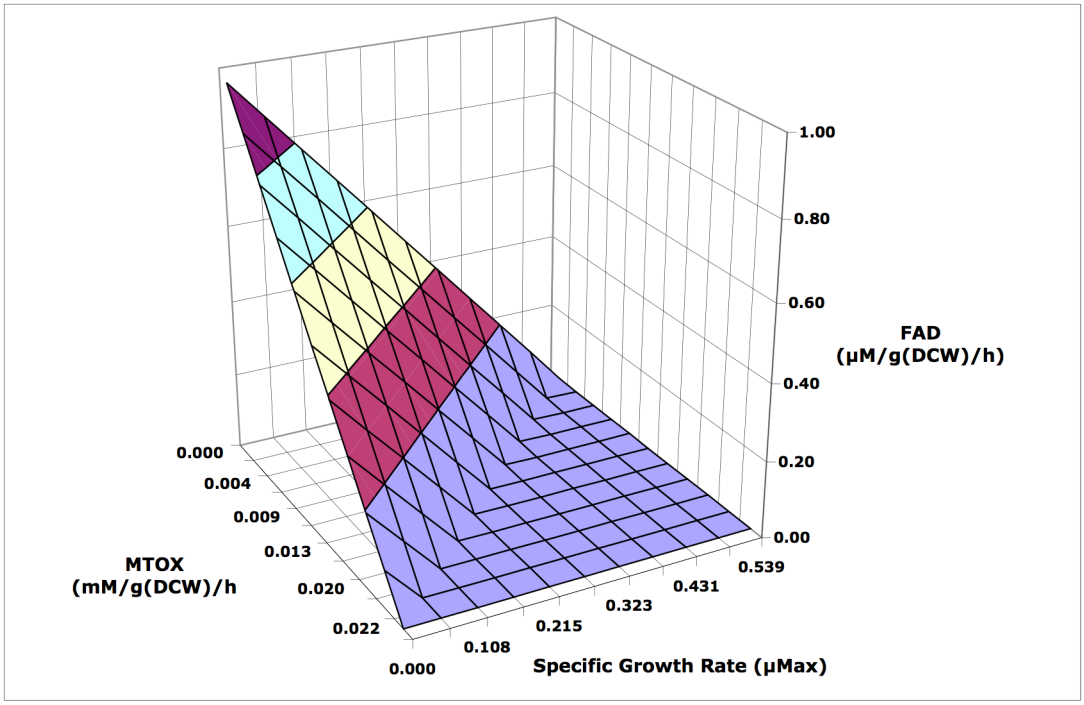


Figure 6.3) Surface plot of the solution space predicted for BW25113 displaying specific growth rate, FAD and MTOX flux.

A surface solution space was generated for MTOX production in the cell. It shows how, in the model at least, carbon flux can be channelled into cell growth, FAD or MTOX production or a combination of the three. Reading off the graph it can be seen that a maximal rate of growth will result in an absence of flux to MTOX and FAD. Similarly a maximal flux into either of the other two, will cause a zero value in the two metabolic endpoints left. In reality this is an impossible result but describes the theoretical maximum fluxes achievable under a given set of conditions

However, more interesting is the solution space on the slope. On the flat plane of the graph only two fluxes are varying the other being zero. On the slope it can be seen that carbon can be sequestered into each of the three endpoints listed, but only within a maximal carbon flux overall, constrained by the glucose uptake rate. For instance, taking a maximum specific growth rate of 0.323 h^{-1} and a FAD flux value of 0.2 μM/g(DCW)/h, the value of approximately 0.014 mM/g(DCW)/h can be read off the z-axis relating to the flux into MTOX production. This demonstrates the sectioning of carbon resource into differing sinks. A finite amount of carbon (defined by the glucose uptake rate) is sectioned into cell maintenance and growth, and specifically the MTOX enzyme being investigated but also the FAD cofactor needed for the activity of the enzyme product.

As will be discussed in section 6.5 one of the drawbacks of this simulation is that, by definition, the model is static. As mentioned in Chapter 2, for each condition a steady state is calculated, therefore this model is not inclusive of the possible

utilisation of FAD to produce an active MTOX protein, and any feedback mechanism that ensures an adequate supply of FAD for cellular maintenance when demands are put on the supply.

It does however give an idea of the maximum value that could be expected from experimental work, which will be compared in section 6.4.

6.3. Altering carbon flux through specific pathways

Part of the initial rationale for conducting the modelling work was to be able to investigate the potential for deletion or up-regulation of genes involved in the FAD pathway (see Figure 1.4) without having to create those mutants in the lab. In theory many mutants could be investigated with this high-throughput *in silico* work and only the most promising could be cloned and investigated.

Therefore as an extension to the above modelling work FAD levels were again the focus in relation to changes in the specific growth rate. However the model was then constrained by altering the flux through the riboflavin pathway. As discussed in Chapter 1 (section 1.7) it was hypothesised that altering flux through the metabolic pathways downstream of riboflavin may alter the production of FAD in the cell. Figure 6.4 shows how FAD levels are altered when the carbon flux downstream of riboflavin was constrained in the model.

Initially the flux down stream of riboflavin was set at zero. This allowed the model to compute the solution space with all available carbon being sequestered into either FAD or biomass. There were then 3 differing scenarios where the flux downstream of riboflavin was increased by 0.25 mM/g(DCW)/h in turn. The effect this loss of carbon downstream of riboflavin had on the system is shown below.

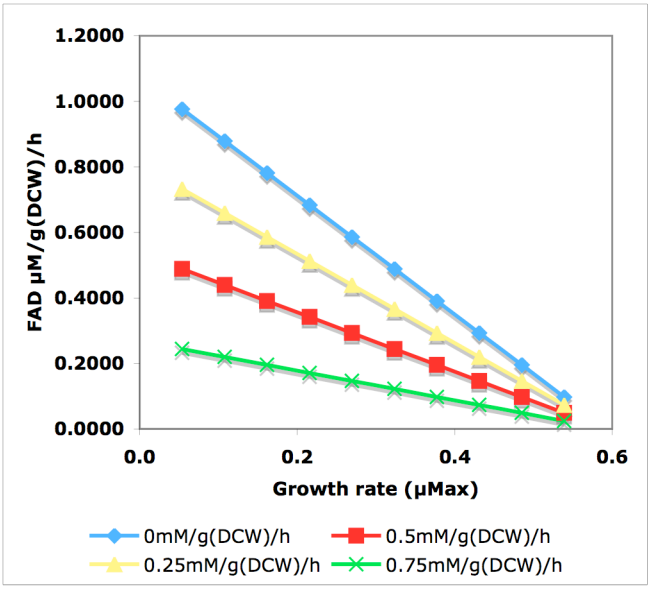


Figure 6.4) Plot of 4 different simulations showing the effect of altering carbon flux through the pathways downstream of riboflavin on FAD flux as a function of specific growth rate; with an initial carbon uptake rate of 6.6 mM/g(DCW)/h.

With no flux of carbon through the pathways downstream of riboflavin the highest levels of FAD are achieved, as seen in Figure 6.4. The graph also shows that increasing the specific growth rate reduces the amount of carbon available for FAD production and the flux decreases. It is thought that this scenario could represent a mutant strain after the deletion of both cobB and cobT genes, however this would render the resulting double mutant auxotrophic for vitamin B12, whereas reductions

in flux downstream of riboflavin caused by the deletion of either gene in isolation may cause a reduction in flux through the pathway significant enough to display an effect similar to that seen in the Figure 6.4. This was seen in the experimental work where FAD concentration and subsequently MTOX concentration increased when either of the cob mutants were investigated. It was thought that this reduction in carbon flux downstream of riboflavin caused by the cobB and cobT mutants allowed for accumulation of FAD and therefore an increased level of activity in the MTOX enzyme. The difference seen between the two cob mutants is discussed in Chapter 5 Section 11.

When the flux downstream of riboflavin was increased, a subsequent drop in the potential FAD production rate was observed. Again this is an expected result as it has been shown in Figure 6.3 that when dealing with the model data a finite amount of carbon flux is available for all sources. Above, it is shown that potentially lowering the activity of the enzymes, or using mutations to delete them altogether, could have a dramatic effect on the level of intracellular FAD and therefore may allow a greater proportion of MTOX protein molecules to be made active via binding to FAD. As previously mentioned this simulation does not allow for any recycling of metabolites in the cell or any feedback mechanisms to replenish depleting levels of riboflavin, FMN or FAD. The limitation of the model to provide this level of detail will be discussed in section 6.5.

6.4. Comparison of model predictions with experimental data

To allow comparison between actual experimental values and those generated by the metabolic model it was necessary to convert absolute values over the time course of fermentation into flux values as described in Chapter 2. These values could then be directly compared with the model predictions obtained.

Based on the experimental data shown in chapter 5 section 9 the average flux over a 7 hour fermentation period of a BW25113 culture for MTOX was $0.008 \pm 9.6 \times 10^{-3}$ mM/g (DCW)/h and $0.03 \pm 2.7 \times 10^{-3}$ μM/g (DCW)/h for FAD.

Compared to the values shown in the figures 6.2. and 6.3 for MTOX production, these are as expected. Figure 6.1, shows a value for MTOX production of 0.008 to be close to the centre of the expected range but also equates to a growth rate much lower than seen experimentally, approximately 0.29 h^{-1}. Investigating figure 6.2.2, again the MTOX value is close to the middle of the range equating to a FAD flux of 0.55 μM/g (DCW)/h. Finally using figure 6.3 the protein flux is again in the centre of the solution space but seems to equate to a lower range of values for FAD flux and growth rate than perhaps expected.

For FAD however the values calculated from the model seem much higher than those derived from the experimental data. Only at the extreme end of the solution space shown in figure 6.3 are values seen close to those calculated from the

experimental data. There may be a number of factors that could explain the differences seen between the experimental results and the model predictions. As will be discussed in the following section, no FAD recycle or feedback loop can be included in the model work as, by definition, the model seeks a steady state for FAD, MTOX and Biomass production. Also there was no constraint in the model for each molecule of MTOX to have an FAD molecule bound. However, when simulations were carried out using the model and requiring 1 molecule of FAD for each MTOX molecule produced this simply reduced the amount of FAD linearly with the amount of MTOX produced.

Also the model seemed to underestimate the maximum specific growth rate, the highest value being 0.539 h^{-1} compared with the experimentally derived value of 0.64 h^{-1}. It also seemed that the model overestimates the effect the production of MTOX or the supply of FAD would have on growth rate. Similar to the suggested reasons for differences in FAD flux it is likely that there are recycles and feedback mechanisms to lessen the impact of cofactor supply or protein production on growth rate although it is accepted that at high expression levels a certain degree of growth limitation will be seen. Whether a more complex model such as that used by Edwards and Palsson (Edwards, et al 2001) or the energy balance analysis (EBA) method, (Beard et al., 2002) would provide more accurate predictions is unknown.

The development of metabolic networks has over recent years taken two complementary directions. A number of researchers have developed much larger

models based on the principles of FBA. The most recent constraint based model of Palsson's group now contains over 200 extra genes compared with the model of the year 2000 and over 800 more genes than a decade ago (Reed and Palsson, 2003). There are also smaller models that attempt to incorporate some facets of regulation into the metabolic network such as the regulated model of *E.coli* central metabolism developed by Palsson's group, (Covert and Palsson, 2002). The EBA model mentioned above adds constraints to a metabolic network that are thermodynamically feasible, often removing the extreme pathways mentioned earlier and therefore shrinking the solution space to include fewer possible outcomes.

The nature of the approach is such that as a model or network becomes more complex, including regulatory or feedback information they become smaller and more focused on one part of metabolism, (Furusawa et al., 2008; Beard et al., 2002). These models could be useful to identify the drawbacks of the larger FBA models when focused on specific parts of the pathway, i.e. FAD metabolism and regulation but would need to be viewed in the context of a larger but less regulated metabolic network.

What is noticed is that the values generated in Figure 6.3 are generally consistent with values described in the model network. However due to the lack of regulatory constraints each demand on carbon source from one of the three factors used has a direct impact on the level of the other two. It is true that the production of a large recombinant protein will impose demands on a cellular system but in reality those

demands would be met and adapted to unless the level of recombinant protein is very high.

6.5. Drawbacks of the model system

As mentioned previously the basis of the constraint-based FBA model is to arrive at a steady state solution space for each change to the reaction network or question asked of it. Feedback loops therefore, by definition, cannot be modelled in this way, as a steady state could not be reached. Little is known about the mechanisms of regulation of the riboflavin pathway in bacteria (Vitreschak et. al, 2002) but it is thought that a feedback mechanism plays a part in keeping the pool of cofactors (i.e. FMN and FAD) at an approximately constant level. Most studies have estimated the level of free FAD to be in the region of 3μM at any point in time (Tu, and Weissman, 2002).

If a requirement is made to bind one FAD molecule to each MTOX molecule produced, the steady state level of FAD is simply reduced in proportion with the amount of MTOX produced. Using the model as shown in Figure 6.4, a reduction in flux downstream of riboflavin proportionally increases the amount of FAD available. In truth this is a very simplified view of cofactor metabolism in the bacterial cell.

As such the stochiometric reaction network is useful for relatively simple predictions of carbon fate. A certain carbon uptake rate can be sequestered into various carbon sinks. If however the reaction network is thought to be more complex, a more sophisticated model may be needed.

6.6. Summary and Conclusions

The construction and analysis of metabolic networks has drawn much attention in the past few years and this section of the work set out to apply such a model to determine theoretical values for the production of the MTOX protein, the supply of the co-factor, FAD and how these two demands on metabolism would impact on the growth rate of the organism. The aim was then to compare values generated using the metabolic model with actual experimental data, which is contained in chapter 5.

In constraint-based modeling, stoichiometric, thermodynamic, flux capacity, energy preservation and other constraints are used to limit the space of possible flux distributions attainable by the metabolic network. Flux balance analysis work, much like the study presented here, has been undertaken by many researchers (Reed, et al, 2003, Förster et al., 2003, Kaufmann et al, 2003) but as discussed above, is limited somewhat by its inability to reconstruct dynamic metabolic occurrences.

The work was based around Edwards and Palsson's *E. coli* MG1655 reaction network, which was then altered to better reflect *E. coli* BW25113 and appended to include the stoichiometric equation for the production of the MTOX protein. Following these alterations solution spaces were calculated concentrating on the three

variables of FAD and MTOX flux and maximum specific growth rate.

Initially a simplistic view of MTOX production was seen where MTOX flux decreased as the growth rate increased. This was built upon by displaying how a rise in carbon uptake rate would increase both the potential levels of MTOX and FAD proportionally.

From the data generated describing the relationships between FAD, MTOX and specific growth rate a surface solution space was then constructed displaying the interactions of the three carbon sinks under scrutiny.

In relation to the deletion mutant work covered in the previous chapter, flux was reduced (in Figure 6.4), in the pathway downstream of riboflavin to simulate a lowering of flux through this pathway containing the enzymes dimethylbenimidazole phosphoribosyltransferase (cobT) and a linked enzyme catalysing the deacetylase of acetyl-CoA synthetase (cobB). It was observed that a reduction in carbon flux downstream of riboflavin caused an increase in flux to FAD, a phenomenon also seen in the *in vivo* experimental work (see Chapter 5, section 5.8).

The comparison of model predictions, with experimental values was, in general, successful, with the predictions of MTOX flux being in line with experimental figures, as discussed in section 6.4. FAD flux did seem to be overestimated by the model after comparison with fermentation data, but this may have been caused by the model's inability to simulate a dynamic system with an active feedback

mechanism. A more sophisticated mathematical solution would have to be sought with more information about the possible feedback loops in riboflavin pathways before accurate predictions could be made. Such models are used but with greatly reduced numbers of metabolites and reactions. Often a system is broken into constituent pathways such as ethanol, glycerol and carbohydrate production in *Saccharomyces cerevisiae,* (Torres, et al., 1997) and each pathway is optimised in isolation but with increased complexity of regulatory/feedback constraints. This can be further developed into a small but multi-objective model such as that using power law formalism to simultaneously optimize several metabolic responses at one time (Vera, 2003).

This approach could potentially be used for the optimization of metabolites such as FAD or riboflavin but might be too focused on specific pathways to give relevant information about the production of complex proteins or on the growth characteristics described earlier.

In general it was shown that accurate predictions can be made by models of metabolic systems, which, in terms of carbon flux can give insight into possible metabolic engineering solutions. Models have been shown previously to accurately predict carbon flux (Edwards et al, 2001; Pramanik & Keasling, 1997) and have also been used to predict whether deletions in an organism's genome would prove fatal or not (Sergre, 2002).

It is clear from the work presented that the effect of altering flux through specific pathways can be investigated and theoretical maximal values calculated for a number of carbon endpoints. A more accurate picture of specific pathways could be generated but the focus would have to be on a much smaller well-defined system where detailed kinetic and regulatory information is known. Little has been written about the regulation of the riboflavin pathway, and in fact the information on usual cellular concentration of co-factors such as FAD is sparse. Flux balance modelling will always give an idea of the maximal potential for the production of the metabolic products under scrutiny and may help guide *in vivo* experimental work. Specific pathways could be investigated in more detail but this would rely on more kinetic information than currently available and may lose some of the overall picture the model developed here portrays.

In the following chapter all experimental work is discussed as a whole with the common themes running through the work explored and potential further work suggested.

7. Discussion and Future work

7.1. General experimental aim

The general aims for the preceding work were to develop a model system to investigate the production of potentially any recombinant protein in an *E.coli* host. A plasmid expression system was utilised to transform differing strains of host with plasmid containing the gene of interest.

Initially GFP was chosen due to the ease that it could be detected and quantified, however it was realised that little industrial gain would come from a strain capable of over expressing GFP.

Once this semi-high throughput system was developed, in collaboration with GSK the focus of the study was shifted to N-methyl tryptophan oxidase (MTOX). As discussed in Chapter 5, this was not necessarily a protein of great interest but it forms part of a group of proteins known as the flavoenzymes. This large group of FAD co-factor requiring proteins are known to be involved in a number of cellular processes such as bioluminescence, free radical reduction, DNA repair and also apoptosis (Shumyantseva, et al., 2004) and therefore made this group an interesting prospect to investigate.

Finally the third part of the work sought to use a metabolic model of *E.coli* MG1655 and investigate the theoretical fluxes of FAD, MTOX and carbon to biomass including how each output parameter reacted to changes in the others. It was planned that data generated from the model would be compared to experimentally derived values to confirm or

disprove the ability of this reaction network model to accurately predict levels of the two metabolites and the growth of the host.

7.2. Green fluorescent protein expression

The GFP chapter established a baseline for all the work that would follow and also investigated the amount of metabolic burden the chosen plasmid vector would impose on each host.

Growth profiles were generated for each strain used and differences noted between the plasmid free and plasmid carrying strains (both induced with IPTG and uninduced). From these profiles maximum specific growth rates were calculated and compared. HPLC analysis was undertaken to measure the concentrations of glucose and acetate throughout the time course of each fermentation which generated values used to constrain the model system, allowed confirmation that the cultures were carbon limited and also gave a measure (through the acetate levels) of metabolic stress (Rahman and Shimizu, 2008).

It was suggested that as metabolic load increased so to would the level of acetate production. The experimental findings were seen to be broadly in line with this hypothesis with the highest rates of acetate production seen in the plasmid carrying strains after induction with IPTG. Differences seen however were not statistically significant in this case. A second trend was noticed and described in that maximum specific growth rate dropped as perceived metabolic load increased. Again, non

significant differences were calculated however a clear trend was seen with growth rates being highest in the plasmid free strains, lower when these strains were transformed with the plasmid vector, and lower still when the protein included in that vector was induced and was produced by the cell.

One significant difference was proved when a total protein level of the original strain was investigated. The wild type strain produced 1.91 mg/mL of protein after lysis and analysis using the Bradford method (see Chapter 2). This increased significantly to 2.30 mg/mL in the uninduced plasmid carrying strain and then increased (not significantly) to 2.41 mg/mL post induction. This was initially unexpected as it was thought the main difference would be after the induction and transcription of the recombinant protein GFP. It was thought that the cells containing puvGFP containing the AmpR gene would be constitutively producing β-lactamase conferring antibiotic resistance to the host. At approximately 96kDa (Paton et al., 1994), this is much larger than the 27kDa GFP and therefore it follows that its production should have a greater affect on cellular protein levels than the smaller GFP.

However when observing gels of cellular lysate and fluorescent data it was clear that in shake flask cultures, little, or no control was seen over GFP production with, or without induction using IPTG. It was then clear as both uninduced and induced cultures were producing GFP and β-lactamase there should be no expected difference between levels of intra cellular protein in either culture. This result was confirmed when fluorescence

profiles were generated showing similar levels of fluorescence for induced and uninduced cultures.

Growth parameters from the shake flask cultures were then compared with values recorded from 7L stirred tank reactors. With these faster growing cultures significant differences were seen in the fluorescence profiles suggesting transcription was under the control of the lac promotor. This level of control was then confirmed to be caused by faster growing cultures through the use of shake flask experiments using complex media rather than minimal M6 (see chapter 4 section 10).

Why this control of expression should be so greatly affected by growth rate of the host was unknown, however a theory was suggested that stated plasmid replication is not linked to cell division. At slow growth rates relative plasmid copy number was thought to increase, this increased the demand for the lac repressor up to a point where the control was not able to be generated by the cell. In faster growing cultures (such as those seen in the stirred tank reactors and LB shake flasks), this relative copy number increase did not occur as cell division was able to keep pace with plasmid replication. This phenomenon was seen with both the GFP plasmid and the MTOX expressing construct and was the reason for using LB in much of the MTOX shake flask growth experiments.

7.3. Evolution of the model system to express MTOX

As described above a plasmid vector had been developed, capable of expressing a foreign protein of choice. There were some issues over control of expression but these were thought to be understood and could be accounted for in future work. The developed method of parallel low volume shake flasks was then extended to express a co-factor dependent flavoprotein, MTOX.

It was discussed that as part of a wider group of flavoproteins this target gave a more industrially relevant purpose to the research. The theory being that any information gained about the expression of one member of the group may be universally applicable to all.

By definition flavoproteins require FAD as a cofactor, conferring activity when bound to the enzyme. This target protein allowed a directed approach for the selection of deletion mutants sourced from the Keio collection of non-lethal mutants developed by Mori (2000) specifically targeting the riboflavin pathway, responsible for the production and potentially regulation of the levels of FAD in the cell.

The peroxide assay initially showed the TG1 strain to be a much higher producer of MTOX than the strain developed in house and based on the K-12 strain MG1655 (see chapter 5.3). However once the assay was developed to include a measure of the active and inactive forms of the enzyme, data was generated to suggest the level of MTOX expression may be broadly similar in each strain but differences in activity may be due to

an ability of TG1 to supply greater amounts of free FAD for the activation of the enzyme product.

FAD measurements were therefore taken throughout the time course of a number of fermentations and it was shown that prior to induction the TG1 strain did demonstrate a significantly higher concentration of free FAD than the comparison strain MG1655.

This confirmed the assumption that investigating FAD cofactor supply by potentially altering parts of the riboflavin metabolic pathway could alter the amount of active enzyme produced by a host.

What it also suggested, however was the possibility that simply supplementing the growth media used for these experiments with riboflavin may increase the amount of active enzyme measured without the need for complex alterations to the production strain. A number of strains were therefore grown in excess riboflavin but no significant increase in enzyme activity was seen (chapter 5.5).

As riboflavin has no inherent biological activity (Vogl et al. 2007) it is known that it firstly needs to enter the cell and secondly needs to be converted to FMN or in this case FAD to have any affect on the host involved. Whether the result seen was caused by an inability of riboflavin to enter the cell in sufficiently high amounts or whether this riboflavin could not be converted into FAD/FMN rapidly enough to affect the activity

levels of the recombinant protein product is unknown. Some suggestions for future work are however discussed below.

Finally therefore mutants derived by Mori (the Keio collection) were used. This collection of non-lethal mutants contained a number of targets thought to affect the metabolism of riboflavin.

The three mutants chosen were cobB and cobT, both downstream of riboflavin the rationale behind their choice being that a decrease in the carbon flux downstream of riboflavin may affect the level of cellular riboflavin and ultimately the level of FAD available in the cell. And also appA, a gene coding for the enzyme governing the back conversion of FMN to riboflavin, again the rationale being the possibility of lowering carbon flux away from FAD may increase its relative abundance in the cell.

It was seen that all 6 strains initially displayed similar growth profiles and also produced largely similar amounts of the MTOX protein (see chapter 5.7). Two groups were discovered however in that MG1655 and BW25113 and the appA mutant generated noticeable lower amounts of active enzyme (roughly 30% of the total enzyme expressed), when compared with the cobT and cobB mutants and the TG1 strain (all displaying approximately 70% active protein product). This significant difference showed that optimisation for a higher level of enzyme activity was possible as there seemed to be a number of strains that could provide high levels of FAD for enzyme activation.

When investigating cellular FAD levels for these highly active strains it was observed that FAD dropped rapidly soon after induction (as it was being utilised by the recombinant protein product) but then rose sharply indicating a strong feedback response. It has been discussed earlier that little is known regarding the feedback mechanisms in place for control of riboflavin levels however some researchers have suggested the idea of a riboswitch (Vitreschak et al, 2004) and others have expanded on the theory to suggest transport proteins such as those coded for by ypaA may be expressed under the control of such switches and thus control the influx and efflux of riboflavin and its associated metabolites across the cell membrane (Vogl et al., 2007).

However also seen was a drop in the level of FAD towards the end of each fermentation in the stains showing a high level of enzyme activity suggesting the cell can adapt to a certain level of FAD demand but that adaptation draws on a finite pool of FAD (or its precursors) and the feedback mechanism will eventually either enable the cell to reach equilibrium or cause a failure to supply sufficient FAD for the demands imposed on each cell as a whole.

Similarly to the GFP work, a limited number of 5L stirred tank fermentations were undertaken in an attempt to replicate some of the findings seen in the small scale shake flasks. MG1655, TG1 and the BW25113 cobT mutant were used all carrying the designed plasmid pQR498. When compared with data generated from shake flask cultures it was clear similar patterns were obtained (see Chapter 5.10). Mean specific activity (stated in dimensionless units) of the MTOX protein for TG1 was the highest

observed returning a value of 10.55 followed by the cobT mutant which generated a value of 8.43. The 'wild type' MG1655 only produced a specific activity of 3.78. These values matched closely with those found in the earlier work with the strains returning values of 10.29, 8.05 and 3.69 respectively.

Cellular FAD levels also appeared to be higher in the stirred tank reactors for reasons that are discussed in section 5.10. It also seems that MTOX activity is not solely dependant on the amount of FAD available in the cell but the amount of that FAD that is available for binding to the protein. It would seem that this is strain specific but reasons for this are currently unknown.

The conclusion of the MTOX experimental work was that levels of active flavoproteins can be increased in cell cultures either by the selection of a strain known to maintain a high cellular FAD concentration or to engineer the metabolism of a host in such a way that pathways downstream of riboflavin are restricted and an increase in the supply of riboflavin, FMN and FAD is possible. Future work based on these observations is included at the end of this chapter.

7.4. Constraint based modelling and its relevance to the investigational work

To add a third dimension to the work carried out, a model was utilised consisting of a large number of reactions in the metabolism of *E. coli*. As stated previously a total of 695 genes and 720 reactions based on the originally chosen host strain MG1655 were included in this reaction network (Palsson et al., 1997).

Metabolic modelling is becoming more prevalent as it presents opportunities to screen a number of mutations or metabolic constraints *in silico* and thus generating results far quicker than possible with *in vitro* work whist also giving a deeper understanding of the interplay of metabolic reactions and its effect on metabolic parameters such as specific growth rate or recombinant protein yield.

With this in mind, the work sought to verify previously generated data and therefore provide credibility to predictions the model might make over other scenarios.

The model was adapted including known differences between MG1655 and BW25113 (see Table 2.1) and the stociometric equation for MTOX was also added as a constraint. After these additions linear programming was used to search for a solution space maximising either growth rate, MTOX production or FAD production (see Chapter 6.2).

Initially pairs of metabolic outputs representing carbon sinks were compared (MTOX vs specific growth rate; MTOX vs FAD) and then the data was combined to generate a solution space that described the possible interactions of each metabolites concentrations.

This ultimately produced a surface plot that was defined by the extreme pathways of maximum production of each variable and is shown in Figure 6.3.

Extra constraints were added to demonstrate how a reduction of flux downstream of riboflavin might affect cellular FAD levels. This was a rather simplistic view as no riboflavin regulation was included in the model, however it did demonstrate how cellular FAD concentrations would increase as flux downstream of riboflavin decreased.

Finally, in an attempt to make direct comparisons between the experimentally derived and modelled data mean flux over the course of the fermentations were calculated and compared with values generated from the model system.

MTOX production as seen in the BW25113 strain appeared to lie directly in the centre of the range predicted by the model system whereas FAD levels and specific growth rates seemed to be slightly overestimated by the model. It has been mentioned earlier that the model system was set up to be a steady state representation of metabolic processes and not a dynamic system with multiple inputs and feedback loops as found *in vivo*. The fact that this model cannot account for metabolic regulation may be a reason why it underestimates growth rate when the cell is subjected to the pressure of producing a recombinant protein. Similarly, although the FAD concentration seen in cultures is within the range predicted, the values obtained lie high on the predicted curve generated by the model (see figure 6.3). Again, with the model not including any replenishment of FAD or

feedback regulation it seems understandable that the model may underestimate the true concentration of this compound.

7.5. Future work

Possible extensions to this work could be based around three distinct goals. Initially the same experimental model (i.e. MTOX production using an *E.coli* host) would be used but over expression of a number of genes would be investigated rather than the use of deletion mutants. Developing strains with up regulated genes would present different challenges than have been encountered in previous work described but potential gene targets have already been discussed (see section 1.7). When deciding on deletion mutants it was seen that the production of riboflavin was controlled directly by the genes ribC and ribF. Deletions of these genes would be lethal and therefore are not included in the Keio collection previously mentioned here. However their position in the riboflavin pathway suggests that up regulation of these may directly affect the level of riboflavin (and potentially FAD) contained in the cell.

Another gene that has been mentioned in relation to riboflavin levels is *ypaA*, one of the transport proteins responsible for moving riboflavin across the cell membrane. It was discussed in chapter 5.5 that simply enriching the growth medium with riboflavin did very little to alter the level of intracellular FAD (as measured by HPLC and MTOX activity). Whether using an enriched growth medium and a strain with an up regulated *ypaA* gene would increase these intracellular concentrations is something that could be investigated with further work.

Secondly, part of the rationale for the move from GFP to MTOX was that the group of flavoproteins was thought to have more commercial potential that the fluorescent protein. An obvious extension to the work therefore would be to generate a number of similar plasmid vectors with the solA gene replaced by genes coding for a number of other flavoenzymes. Possible targets would be the closely related flavoprotein monomeric sarcosine oxidase (MSOX), also thioredoxin reductase (TrxR) shown to be involved in cell division and with a possible number of anti tumour uses. Also the flavoprotein cytochrome P450 reductase which would be an interesting target requiring both FMN and FAD for activity. Assays, similar to the peroxide assay for MTOX would need to be developed but a repeat of all the work carried out in chapter 5 could theoretically be done with a number of different protein targets, all requiring FAD for the activation of the enzyme product.

The second goal or area of further work would be research into the hypothesis put forward for the lack of control of expression displayed by the lac promoter in the plasmid constructs in slow growing cultures. It has been suggested that in slow growing cultures plasmid copy number increases as cell division does not keep pace with plasmid replication. A quantitative PCR assay with forward and reverse primers and a fluorescent probe (TaqMan ™) designed to be specific to part of the pQR498 plasmid could be developed. Samples from both M6 and LB media shake flask cultures could then be tested and the plasmid copy number estimated over the time course of the fermentation using a standard curve generated from dilutions of the purified plasmid DNA.

Finally, as discussed above it could be investigated whether a more defined model, potentially including some form of feedback mechanism could be developed. It is thought that far more information would be needed in terms of flux values into and out of this system, which may need to be determined experimentally. More information may also need to be known about regulation of the interconversion between riboflavin, FAD and FMN in the cell, and how this can be altered with as differing demands are put on the cell. Currently there is little information known about this and may be a project far greater than the scope of work suggested here.

7.6. Summary

In summary, it has been shown that a model system can be created generating an easily visualised recombinant protein which could be used for high throughput screening of strains/hosts/growth media or mutants designed to over produce various protein products.

This system was altered to focus on a specific protein (MTOX), especially interesting, as it is a member of a wider group of proteins, namely the flavoenzymes. As these enzymes require FAD to confer activity mutants and strains were investigated to determine the level of recombinant protein expression possible and also the level of free FAD available intracellular, which could be bound to the enzyme.

Finally a constraint based model system was used to generate comparative data describing the theoretical maximum levels of FAD, MTOX and rate of growth seen in the *E. coli* host. These values were then compared to experimentally derived figures and were shown to be broadly accurate within the known limitations of the model itself.

Appendices

Appendix 1:

Experimental matrix:

A copy of the experimental matrix used to track multiple cell cultures and assays on the resulting samples is included overleaf.

Strain	Plasmid	Available?	Growth in M6	Growth in LB	Total Protein	GFP	Peroxide	HPLC FAD
BW25113	WT	Yes	Yes	Yes	Yes	Yes		
	pGFPuv	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	pQR498	Yes	Yes	Yes		N/A	Yes	Yes
	pMEX8	N/A	N/A	N/A	N/A	N/A	N/A	N/A
BW25113 (appA)	WT	Yes	Yes	Yes		N/A	Yes	Yes
	pGFPuv	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	pQR498	Yes	Yes	Yes		N/A	Yes	Yes
	pMEX8	N/A	N/A	N/A	N/A	N/A	N/A	N/A
BW25113 (cobB)	WT	Yes	Yes	Yes		N/A	Yes	Yes
	pGFPuv	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	pQR498	Yes	Yes	Yes		N/A	Yes	Yes
	pMEX8	N/A	N/A	N/A	N/A	N/A	N/A	N/A
BW25113 (cobT)	WT	Yes	Yes	Yes		N/A	Yes	Yes
	pGFPuv	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	pQR498	Yes	Yes	Yes		N/A	Yes	Yes
	pMEX8	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MG1655	WT	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	pGFPuv	Yes	Yes	Yes	Yes	Yes	N/A	N/A
	pQR498	Yes	Yes	Yes	Yes	N/A	Yes	Yes
	pMEX8	Yes	Yes	Yes		N/A	Yes	
TG1	WT	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	pGFPuv	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	pQR498	Yes	Yes	Yes		N/A	Yes	Yes
	pMEX8	Yes	Yes	Yes		N/A	Yes	
W3110	WT	Yes	Yes	Yes		Yes	Yes	
	pGFPuv	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	pQR498	Yes	Yes				Yes	
	pMEX8	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Appendix 2:

Stociometric equation:

As mentioned previously one of the restraints imposed upon the reaction network model used in Chapter 6 was the stociometric equation for the solA gene (encoding the protein MTOX). This included all amino acids in the peptide and also the required phosphate for its synthesis. The equation as used, follows:

ala-L	-38
cys-L	-6
asp-L	-28
glu-L	-21
phe-L	-19
gly	-34
his-L	-13
ile-L	-21
lys-L	-16
leu-L	-33
met-L	-4
asn-L	-14
pro-L	-20
gln-L	-14
arg-L	-16
ser-L	-18
thr-L	-17
val-L	-22
tyr-L	-12
atp	-372
gtp	-372
MTOX	1
adp	372
gdp	372

Appendix 3:

Example Calculations:

1) **Maximum specific growth rate**

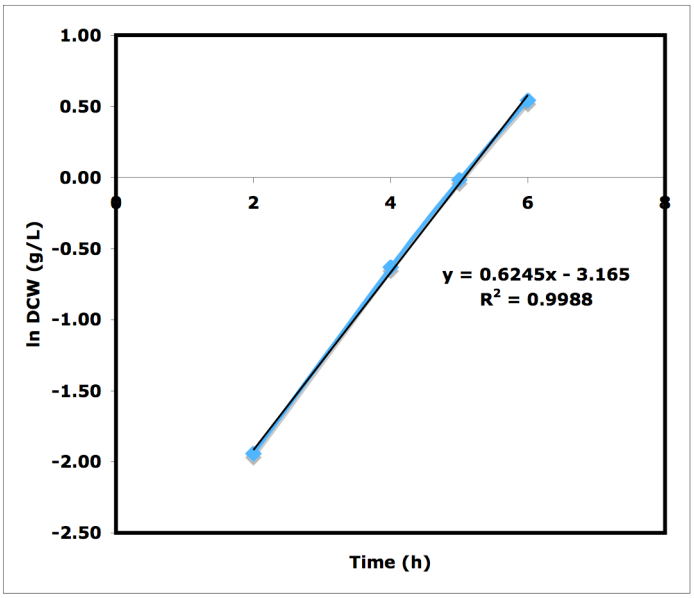


Figure A.1) Log plot of exponential phase growth of E.coli MG1655 WT, M6 in shake flask cultures.

Maximum specific growth rate was calculated using a plot of the natural log(ln) of D.C.W. against time of fermentation throughout the exponential phase of growth. Linear regression was used to fit a trendline with the slope defining $\mu_{max.}$ Clear outliers from each curve were removed to improve the coefficient of determination (R_2) >0.95.

2) **Yield calculations:**

Yield calculations were performed from a known glucose supply, (4 g/L) or a derived acetate final concentration (approx 2.1 g/L, culture dependent).

Both used the value of Xmax (maximum biomass concentration, g/L) depending on the culture in question.

Therefore:

$Y_{x/glu}$ = Xmax / Glucose concentration

for *E.coli* strain MG1655, WT, M6,

$Y_{x/glu}$ = 1.72 (g/L) / 4 (g/L) = 0.43 (g/g)

$Y_{ac/x}$ = Acetate concentration / Xmax

for *E.coli* strain MG1655, WT, M6,

$Y_{ac/x}$ = 2.01 (g/L)/ 1.72 (g/L) = 1.17 (g/g)

References

1. Agrawal, A. A., (2001), "Phenotypic Plasticity in the Interactions and Evolution of Species.", *Science,* **294**, (5541), 321 – 326.

2. Akimoto, M., Sato, Y., Okubo, T., Todo, H., Hasegawa, T., and Sugibayashi, K., (2006), "Conversion of FAD to FMN and Riboflavin in Plasma: Effects of Measuring Method", *Biol. Pharm. Bull.*, **29**, 1779 – 1782.

3. Alldrick, A. J., Smith, J. T., (1983), "R – plasmid effects on bacterial multiplication and survival.", *Antoine van Leeuwenhoek,* **49,** (2), 133 – 142.

4. Allen, T. E., Palsson, B. Ø., (2003), "Sequence – Based Analysis of Metabolic Demands for Protein Synthesis in Prokaryotes." *Journal of Theoretical Biology* **220**, 1 – 18.

5. Almaas, E., Kovacs, B., Vicsek, T., Oltvai, Z. N., Barabasi, A-L., (2004), "Global organisation of the metabolic fluxes in the bacterium *Escherichia coli.", Nature* **427**, 839 – 843.

6. Anderlei, T., Büchs, J., (2001), "Device for sterile online measurement of the oxygen transfer rate in shaking flasks", *Biochemical Engineering Journal,* **7,** 157 – 162.

7. Arnold, D., (1995), "Virus safety considerations for recombinant factor VIII (rFVIII, Kogenate)", *Haemophilia*, **1**, (2), 22 – 23.

8. Baba, Tomoya,Ara, Takeshi, Hasegawa, Miki, Takai, Yuki, Okumura, Yoshiko, Baba, Miki, Datsenko, Kirill A, Tomita, Masaru, Wanner, Barry L, Mori, Hirotada, (2006), "Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection", *Molecular Systems Biology* **2** 1744 - 1751

9. Bacher, A., Eberhardt, S., Eisenreich, W., Fischer, M., Herz, S., Illarionov, B., (2001), "Biosynthesis of Riboflavin, Vitamins and Hormones.", **61,** 1 – 49.

10. Báez-Viveros, J.L., Osuna, J., Hernández-Chavez, G., Soberón, X., Bolivar, F., Gosset, G., (2004), "Metabolic Engineering and Protein Directed Evolution Increase the Yield of L - Phenylalanine Synthesized from Glucose in *Escherichia coli*.", *Biotechnology and Bioengineering* **87** (4), 516 – 524.

11. Bailey, J.E., (1998), "Mathematical Modeling and Analysis in Biochemical Engineering: Past Accomplishments and Future Opportunities.", *Biotechnology Progress,* **14**, 8 – 20.

12. Baneyx, F**.**, Mujacic, M**.**, (2004), "Recombinant protein folding and misfolding in *Escherichia coli*", *Nat Biotechnology*, **22**, (11), 1399 – 1408.

13. Barile, M., Brizio, C., Virgilio, C., Delfine, S., Quagliariello, E., and Passarella, S., (1997), "Flavin Adenine Dinucleotide and Flavin Mononucleotide Metabolism in Rat Liver. The Occurrence of FAD Pyrophosphatase and FMN Phosphohydrolase in Isolated Mitochondria", *European Journal of Biochemistry* **249**, 777 – 785.

14. Barile, M., Passarella, S., Bertoldi, A., and Quagliariello, E., (1993), "Flavin Adenine Dinucleotide Synthesis in Isolated Rat Liver Mitochondria Caused by Imported Flavin Mononucleotide", *Archives of Biochemistry and Biophysics* **305**, 442 - 447.

15. Beard, D. A., Liang, S. D., Qian, H., (2002), "Energy Balance for Analysis of Complex Metabolic Networks.", *Biophysical Journal,* **83,** 79 – 86.

16. van Berkel, W. J. H., Kamerbeek, N. M., Fraaije, M. W., (2006), "Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts", *Journal of Biotechnology,* **124,** 670 – 689.

17. Berry, A., (1996), "Improving production of aromatic compounds in *Escherichia coli* by metabolic engineering.", *Trends in Biotechnology* **14**, 219 - 259.

18. Benita, Y., Wise, M. J., Lok, M. C., Humphery – Smith, I., Oosting, R. S., (2006), "Analysis of high throughput protein expression in *Escherichia coli*", *Mol Cell Proteomics*, **5**, (9), 1567 – 80.

19. Betenbaugh, M. J., diPasquantonio, V. M., Dhurjati, P., (1987), "Groeth kinetics of *Escherichia coli* containing temperature – sensitive plasmid pOU140.", *Biotechnology and Bioengineering,* **29,** (9), 1164 – 1172.

20. Blattner, F. R., Plunkett III, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., Song, J., (1997) "The Complete Genome Sequence of *Escherichia coli* K12.", *Science* **277**, 1453 – 1474.

21. Borodina, I., Krabben, P., Nielsen, J., (2005), *"*Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism*", Genome Res.* **15**, 820 – 829.

22. Bonarius, H. P. J., Schmid, G., Tramper, J., (1997) "Flux analysis of underdetermined metabolic networks: the quest for the missing constraints", *Trends in Biotechnology*, **15**, (8), 308 – 314.

23. Bradford, M. M., (1976), "A rapid and sensitive for the quantitation of microgram quantitites of protein utilizing the principle of protein-dye binding." *Analytical Biochemistry,* **72**, 248 – 254.

24. Brandsch, R., Bichler, V., (1989), "Covalent cofactor binding to flavoenzymes requires specific effectors.", European Journal of Biochemistry, 182 (1), 125 – 128.

25. Braud, S., Moutiez, M., Belin, P., Abello, N., Drevet, P., Zinn – Justin, S., Courçon, M., Masson, C., Dassa, J., Charbonnier, J. B., Boulain, J. C., Ménez, A., Genet, R., Gondry, M., (2005), "Dual expression system suitable for high – throughput fluorescence – based screening and production of soluble proteins.", *J Proteome Res,* **4,** (6), 2137 – 2147.

26. Britz – McKibbin, P., Markuszewski, M. J., Iyanagi, T., Matsuda, K., Nishioka, T., and Terabe, S., (2003), "Picomolar analysis of flavins in biological samples by dynamic pH junction-sweeping capillary electrophoresis with laser-induced fluorescence detection", *Analytical Biochemistry* **313**, 89 – 96.

27. Büchs, J. (2001). "Introduction to advantages and problems of shaken cultures.", *Biochemical Engineering Journal,* **7,** 91 – 98.

28. Burgess, C., O'connell – Motherway, M., Sybesma, W., Hugenholtz, J., van Sinderen, D., (2004), "Riboflavin production in *Lactococcus lactis*: potential for in situ production of vitamin-enriched foods.", *Appl Environ Microbiol*, **70**, (10), 5769 – 5777.

29. Cameron, D. C., Altaras, N. E., Hoffman, M. L. and Shaw, A. J. (1998), "Metabolic engineering of propanediol pathways.", *Advances in Biochemical Engineering*, **14,** 116 – 125.

30. Carrió, M. M., Cuatrecasas, S., Arís, A., Villaverde, A., (2005), "Bacterial inclusion bodies are cytotoxic in vivo in absence of functional chaperones DnaK or GroEL.", *J, Biotechnology*, **10**, 118 (4), 406 – 412.

31. Causey, T. B., Shanmugam, K. T., Yomano, L. P., Ingram, L. P., "Engineering *Escherichia coli* for efficient conversion of glucose to pyruvate.", *Proceedings of the National Academy of Science USA* **101, (**8), 2235 – 2240.

32. Chae, H. J., Delisa, M. P., Cha, H. J., Weigand, W. A., Rao, G., Bentley, W. E., (2000), "Framework for online optimization of recombinant protein expression in high-cell-density *Escherichia coli* cultures using GFP – fusion monitoring.", *Biotechnol Bioeng*, **5**, 69, (3), 275 – 285.

33. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., Prasher, D. C., (1994), "Green Fluorescent Protein as a marker for gene expression.", *Science,* **263**, 802 – 805.

34. Chang, D. E., Shin, S., Rhee, J. S., Par, J. G., (1999), "Acetate metabolism in a pta mutant of *Escherischia coli* W3110: Importance of maintaining acetyl Coenzyme A flux for growth and survival", *J. Bact.,* **181,** (21), 6656 – 6663.

35. Chen, J. Q., Zhang, H. T., Hu, M. H., Tang, J. G., (1995), "Production of human insulin in an *E. coli* system with Met – Lys – human proinsulin as the expressed precursor.", *Appl Biochem Biotechnology* **55,** (1), 5 – 15.

36. Chen, P., Ailion, M., Weyand, N., Roth, J., (1995), "The End of the cob Operon: Evidence that the Last Gene (cobT) Catalyzes Synthesis of the Lower Ligand of Vitamin B12, Dimethylbenzimidazole.", *Journal of Bacteriology*, **177**, (6), 1461 – 1469.

37. Christie, J. M., Reymond, P., Powell, G. K., Bernasconi, P., Raibekas, A. A., Liscum, E., Briggs, W. R., (1998), "Arabidopsis NPH1: a flavoprotein with the properties of a photoreceptor for phototropism", *Science*, **282** (5394), 1698 – 1701.

38. Cherch, B., Rivera, E., Llagostera, M., (1996), "Identification of a pKM101 region which confers a slow growth rate and interferes with susceptibility to Quinolone in *Escherichia coli* AB1157.", *J. Bact.,* **178,** 5568 – 5572.

39. Cordell, H. J., (2002), "Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans.", *Hum. Mol. Genet.,* **11,** 2463 – 2468.

40. Cui, D. F., Li, M. Y., Zhang, Y. S., Feng, Y. M., (2001), "Monomeric destetrapeptide human insulin from a precursor expressed in *Saccharomyces cerevisiae*", *Journal of Peptide Research*, **57,** (3) 188.

41. Cusano, A. M., Parrilli, E., Marino, G., Tutino, M. L., (2006), "A novel genetic system for recombinant protein secretion in the Antarctic *Pseudoalteromonas haloplanktis* TAC125.", *Microb Cell Fact*., **5**, 40.

42. Edwards, J., Covert, M., and Palsson, B., (2002), "Metabolic modeling of microbes: the flux-balance approach.", *Environmental Microbiology*, **4**, 133 – 140.

43. Edwards, J., Ibarra, R., and Palsson, B., (2001), "In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data", *Nature Biotechnology,* **19**, (2), 125 – 130.

44. Edwards, J., and Palsson, B., (2000), "The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities", *Proceedings of the National Academy of Sciences*, **97**, (10), 5528 – 5533.

45. Edwards, J., and Palsson, B., (1999), "Systems Properties of the *Haemophilus influenzae* Rd Metabolic Genotype", *J. Biol. Chem.,* **274**, 17410 – 17416.

46. Endo, Y., Sawasaki, T., (2004), "High – throughput, genome-scale protein production method based on the wheat germ cell-free expression system", *Journal of Structural and Functional Genomics*, **5**, (1), 45 – 57.

47. Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., Palsson, B. Ø., (2007), "A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information", *Molecular Systems Biology*, **3,** 121.

48. Feliu, J. X., Cubarsi, R., Villaverde, A., (1997), "Optimized Release of Recombinant Proteins by Ultrasonication of *E. coli* Cells", *Biotechnology and Bioengineering*, **58**, 5.

49. Ferrer-Miralles, N., Domingo-Espin, J., Corchero, J. L., Vazquez, E., Villaverde, A., (2009), "Microbial factories for recombinant pharmaceuticals.", *Microbial Cell Factories,* **8, (**17) , 1475 – 2859.

50. Fischer, M., Romisch, W., Illarionov, B., Eisenreich, W., Bacher, A., (2005), "Structures and reaction mechanisms of riboflavin synthases of eubacterial and archaeal origin.", *Biochem Soc Trans* **33**, 780 – 784.

51. Fischer, M., Haase, I., Feicht, R., Richter, G., Gerhardt, S., Changeux, J. P., Huber, R., Bacher, A., (2002), "Biosynthesis of riboflavin: 6,7 – dimethyl – 8 – ribityllumazine synthase of Schizosaccharomyces pombe.", *Eur J Biochem* **269**, 519 – 526.

52. Förster, J., Famili, I., Fu, P., Palsson, B. O., Nielsen, J., (2003) "Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network." *Genome Res* **13**, 244 – 253.

53. Fraaije, M. W., Wu, J., Heuts, D. P. H. M., van Hellemond, E. W., Spelberg, J. H., Janssen, D. B., (2005), "Discovery of a thermostable Baeyer – Villiger monooxygenase by genome mining.", *Appl. Microbiol. Biotechnol,* **66,** (4), 393 – 400.

54. Frederick, R., Westler, W. M., Song, J., Zornetzer, G., Markley, J., (2004) "Protein Production using *E.coli,* Yeast (*Pichia pastoris*) and Cell Free (Wheat Germ and Bacterial) Systems.

55. Furusawa, C., Kaneko, K., (2008), "A generic mechanism for adaptive growth rate regulation.", *PLoS Computational Biology*, **4,** (1), 35 – 42.

56. Gad, S. C., (2007), "Handbook of Pharmaceutical Biotechnology", John Wiley and Sons.

57. Galan, B, Diaz, E., Garcia, J. L., (2000), "Enhancing desulphurization by engineering a flavin reductase – encoding gene cassette in recombinant biocatalysts.", *Environmental Microbiology,* **2,** 687 – 694.

58. Garcia-Arrazola, R., Chau Siu, S., Chan, G., Buchanan, I., Doyle, B., Titchener - Hooker N. J., Baganz, F., (2005), "Evaluation of a pH-stat feeding strategy on the production and recovery of Fab' fragments from *E. coli*.", *Biochemical Engineering Journal,* **23,** 221 – 230.

59. Gray, K. A., Mrachko, G. T., Squires, C. H., (2003), "Biodesulphurization of fossil fuels.", *Current Opinion in Microbiology.,* **6,** 229 – 235.

60. Hahm, D. H., Pam, J., Rhee, J. S., (1994), "Characterization and evaluation of a pta (phosphotransacetylase) negative mutant of *Escherichia coli* HB101 as a production host of foreign lipase", *Applied Microbiology and Biotechnology, **42,** (1), 100 – 107.*

61. Hatfield, G. W., Roth, D. A., (2007), "Optimizing scaleup yield for protein production: Computationally Optimized DNA Assembly (CODA) and Translation Engineering", *Biotechnol Annu Rev,* **13,** 27 – 42.

62. Hassan – Abdallah, A., Bruckner, R. C., Zhao, G., Schuman – Jorns, M., (2005), "Biosynthesis of Covalently Bound Flavin: Isolation and in vitro Flavination of the Monomeric Sarcosine Oxidase Apoprotein", *Biochemistry*, **44**, 6452 – 6462.

63. Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S., Ohtsubo, E., Baba, T., Wanner, B. L., Mori, H., Horiuchi, T., (2006), "Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110." *Mol. Syst. Biol.* **2**, E1 - E5.

64. Heim, R., Cubitt, A., Tsien, R., (1995). "Improved green fluorescence". *Nature* **373**, (6516) 663 – 664.

65. Hewitt, L**.**, McDonnell, J. M**.**, (2004), "Screening and optimizing protein production in *E.coli*", *Methods Mol Biol*, **278**, 1 – 16.

66. Hoffmann, F., and Rinas, U., (2004), "Plasmid amplification in *Escherichia coli* after temperature upshift is impaired by induction of recombinant protein synthesis", *Biotechnology Letters*, **23**, (22) 1819 – 1825.

67. Hustad, S., Ueland, P. M., Schneede, J., (1999), "Quantification of Riboflavin, Flavin Mononucleotide and Flavine Adenine Dinucleotide in Human Plasma by Capillary Electrophoresis and laser induced Fluorescence Detection.", *Clinical Chemistry*, **45**, (6), 862 - 868.

68. Ito, Y., Suzuki, M., Husimi, Y., (1999), "A Novel Mutant of Green Fluorescent Protein with Enhanced Sensitivity for Microanalysis at 488 nm Excitation.", *Biochemical and Biophysical Research Communications,* **264**, 556 – 560.

69. Job, V., Marcone, G. L., Pilone, M. S., and Pollegioni, L., (2002), "Glycine Oxidase from *Bacillus subtilis", Biol. Chem.*, **277**, (9), 6985 – 6993.

70. Joyce, A. R., Reed, J. L., White, A., Edwards, R., Osterman, A., Baba, T., Mori, H., Lesely, S. A., Palsson, B., Agarwalla, S., (2006), "Experimental and computational assessment of conditionally essential genes in *E. coli*.", J. *Bacteriol.* **10**, 1128.

71. Kang, S**.**, Kang. K**.**, Lee, K**.**, Back, K., (2007), "Characterization of rice tryptophan decarboxylases and their direct involvement in serotonin biosynthesis in transgenic rice", *Planta*, Epublished ahead of print

72. Kauffman, K. J., Prakash, P., Edwards, J. S., (2003) "Advances in flux balance analysis.", *Curr Opin Biotech* **14**, 491 – 496.

73. Kearney, E. B., (1952), "The interaction of yeast flavokinase with riboflavin analogues.", *J. Biol. Chem.* **194**, 747 – 754.

74. Keio collection (The), (http://ecoli.aist-nara.ac.jp/)

75. Khanna, P., and Jorns, M. S., (2001) "N-Methyltryptophan Oxidase from *Escherichia coli*: Reaction Kinetics with N-Methyl Amino Acid and Carbinolamine Substrates", *Biochemistry* **40**, 1451 - 1459.

76. Khanna, P., and Jorns, M. S., (2001) "Characterization of the FAD-Containing *N*-Methyltryptophan Oxidase from *Escherichia coli"*, *Biochemistry* **40**, 1441 – 1450.

77. Khanna, P., and Jorns, M. S., (2003) "Tautomeric Rearrangement of a Dihydroflavin Bound to Monomeric Sarcosine Oxidase or *N*-Methyltryptophan Oxidase", *Biochemistry,* **42**, 864 – 869.

78. King, T. E., Howard, R. L., Wilson, D. F., Li, J. C. R., (1962), "The Partition of Flavins in the Heart Muscle Preparation and Heart Mitochondria", *J.Biol.Chem.*, **237**, 2941 – 2946.

79. Kirkpatrick, C., Maurer, L. M., Oyelakin, N. E., Yoncheva, Y. N., Maurer, R., Slonczewski, J. L., (2001), "Acetate and formate stress: opposite responses in the proteome of *Escherichia coli"*, *Journal of Bacteriology,* **183**, (21), 6466 – 6477.

80. Koyama, Y., Ohmori, H., (1996), "Nucleotide sequence of the *Escherichia coli* solA gene encoding a sarcosine oxidase - like protein and characterization of its product.", *Gene,* **28**, (181), 1-2, 179 – 183.

81. Koffas, M., Roberge, L., Lee, K., Stephanopoulos, G., (1999), "Metabolic engineering", *Annual review of Biomedical Engineering*, **1**, 535 – 557.

82. Kohara, Y., Akiyama, K., Isono, K., (1987), "The Physical map of the whole *E.coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library.", *Cell*, **50**, 495 – 508.

83. Labarca, C., Paigen, K., (1980) "A simple, rapid, and sensitive DNA assay procedure", *Analytical Biochemistry*, **102,** (2), 344 – 352.

84. Leonhartsberger, S., (2006) "*E. coli* Expression System Efficiently Secretes Recombinant Proteins into Culture Broth", *Bioprocess international*

85. Mack, M., va Loon, A. P. G. M., Hohmann, H. P., (1998), "Regulation of Riboflavin biosynthesis in *Bacillus subtilis* is affected by the activity of the flavokinase/flavin adenine dinucleotide synthetase encoded by ribC.", *J. Bact.,* **180,** (4), 950 – 955.

86. Massey, V., (2000), "The Chemical and Biological Versatility of Riboflavin", *Biochem. Soc. Trans*., **28**, 283 – 296.

87. de Marco, A., (2007), "Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in *Escherichia coli", Nature Protocols,* **2**, (10), 2632 – 2639.

88. González – Montalban, N., García – Fruitós, E., Ventura, S., Arís, A., Villaverde, A., (2006), "Comparative analysis of *E. coli* inclusion bodies and thermal protein aggregates", *Microbial Cell Factories,* 5 (1)

89. Mahadevan, R., Doyle III, F. J., (2003), "On – line optimization of recombinant product in a fed – batch bioreactor.", *Biotechnol Prog.*, **19**, (2), 639 – 646.

90. Martinez – Julvez, M., Frago, S., Medina, M., (2003), "Structural and functional analysis of the FAD synthetase from *Corynebacterium ammoniagenes", European Journal of Biochemistry*, **1** (1).

91. Massey, V., (2000), "The chemical and biological versatility of riboflavin.", *Biochem Soc Trans*., **28,** (4), 283 – 296.

92. Merten, O. W., Mattanovich, D., Cole, J., Lang, C., Larsson, G., Neubauer, P., Porro, D., Postma, P., Teixeira de Mattos, J., (2001), "Production of recombinant proteins with prokaryotic and eukaryotic cells.", *Kluwer Academic Publ*, 339 – 346.

93. Michel Jr. F. C., Grulke, E. A., Reddy, C. A., (1990), "Development of a stirred tank reactor system for the production of lignin peroxidases by Phanerochaete chrysosporium.", *Journal of Industrial Microbiology,* **5**, 103 – 112.

94. Mori, H., Isono, K., Horiuchi, T., Miki, T., "Functional genomics of *Escherichia coli* in Japan.", *Research in Microbiology,* **151**, 121 – 128.

95. Naudi, A., Caro, P., Jove, M., Gomez, J., Boada, J., Ayala, V., Portero – Otin, M., Barja, G., Pamplona, R., (2007), "Methionine Restriction Decreases Endogenous Oxidative Molecular Damage and Increases Mitochondrial Biogenesis and Uncoupling Protein 4 in Rat Brain", *Rejuvenation Research.*, **10**,(4), 473 – 484.

96. Palsson, B. O., (1997), "What lies beyond bioinformatics?", *Nature Biotechnology,* **15**, 3 – 4.

97. Panke, S., Witholt, B., Schmid, A., Wubbolts, M. G., (1998), "Towards a biocatalyst for (S) – styrene oxide production: characterization of the styrene degredation pathway of *Pseudomonas* sp. strain VLB120.", *Applied Environmental Microbiology,* **64,** 2032 – 2043.

98. Paton, R., Miles, R. S., Amyes, S. G., (1994), "Biochemical properties of inducible beta-lactamases produced from *Xanthomonas maltophilia*." *Antimicrob. agents chemother.,* **38,** (9), 2143 – 2149.

99. Petersson, L., Carrió, M. M., Vera, A., Villaverde, A., (2004), "The impact of dnaKJ overexpression on recombinant protein, solubility results from antagonistic effects on the control of protein quality.", *Biotechnol Lett.,* **26**, (7), 595 – 601.

100. Petushkov, V. N., Lee, J., (1997), "Purification and Characterisation of Flavoproteins and Cytochromes from the Yellow Bioluminescence Marine Bacterium *Vibrio Fischeri* Strain Y1", *FEBS Journal,* **245**, (3), 790 – 796.

101. Pramanik, J., and Keasling, J., (1997), "Stoichiometric model of *Escherichia coli* metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements.", *Biotechnology and Bioengineering*, **56**, (4), 398 – 421.

102. Prasher, D., Eckenrode, V., Ward, W., Prendergast, F., Cormier, M., (1992) "Primary structure of the *Aequorea victoria* green – fluorescent protein". *Gene,* **111**, (2), 229 – 233.

103. Rahman, M., and Shimizu, K., (2008), "Altered acetate metabolism and biomass production in several *Escherischeria coli* mutants lacking rpo-S-dependant metabolic pathway genes.", *Molecular Biosystems*, **4**, 160 – 169.

104. Reed, J. L., Famili, Thiele, I., Palsson, B. O., (2006), "Towards multidimensional genome annotation.", *Nat. Rev. Genet.,* **7,** 130 – 141.

105. Reed, J. L., Palsson B. O., (2003), "Thirteen years of building constraint – based in silico models of *Escherischia coli*.", *J. Bact,* **185**, (9), 2692 – 2699.

106. Read, S. M., and Northcote, D. H., (1981), "Minimization of variation in the response to different proteins of the Coomassie blue G dye-binding assay for protein", *Analytical Biochemistry*, **116**, 53 – 64.

107. Record of GMO and GM Product Dealings under Section 138 of the *Gene Technology Act,* (2000)*,* Dealings Involving GM products – therapeutics, http://www.ogtr.gov.au/pdf/gmorec/theraprod.pdf

108. Remold, S. K., Lenski, R. E., (2004), "Pervasive joint influence of epistasis and plasticity on mutaional effects in *Escherichia coli*.", *Nature Genetics*, **36**, (4), 423 – 426.

109. 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**, (7), 909 – 915.

110. Sambrook, J., Fritsch, F. E., and Maniatis, T., (1989), "Molecular Cloning, A Laboratory Manual", *Cold Spring Harbour Laboratory, Cold Spring Harbour, NY*

111. Seech, A. G., Trevors, J. T., (1991), "Environmental variables and evolution of xenobiotic catabolism in Bacteria.", *Trends in Ecology and Evolution,* **6,** (3), 79 - 83

112. Segre, D., Vitkup, D., and Church, G. M., (2002), "Analysis of optimality in natural and perturbed metabiolic networks", *PNAS*, **99**, (23), 15112 – 15117.

113. Sharma, S. S., Blattner, F. R., Harcum, S. W., (2006), "Recombinant protein production in an *Escherichia coli* reduced genome strain.", *Metabolic Engineering*, **9**, (2) 133 – 141.

114. Shimomura, O., Johnson, F. H., Saiga, Y., (1962), "Extraction, purification and properties of aequorin, a bioluminescent protein from luminous hydromedusan, *Aequorea.", J.Cell Comp Physiol*, **59**, 223 – 239.

115. Shivananda, V. S., Srikanth, B. S., Mohan, M., Kulkarn, P. S., (2006), "Comparison of Two Hepatitis B Vaccines (GeneVac-B and Engerix-B) in Healthy Infants in India.", *Clinical and Vaccine Immunology*, **13**, (6), 661 – 664.

116. Shumyantseva, V. V., Bulko, T. V., Petushkova, N. A., Samenkova, N. F., Kuznetsova, G. P., Archakov, A. I., (2003) "Fluorescent assay for riboflavin binding cytochrome P450 2B4", *Journal of Inorganic Biochemistry*, **98**, 365 – 370.

117. Škulj, M., Okršlar, V., Jalen, S., Jevševar, S., Slanc, P., Štrukelj, B., Menart, V., (2008), "Improved determination of plasmid copy number using quantitative real-time PCR for monitoring fermentation processes.", *Microbial Cell Factories*, **7**, 6.

118. Solovieva, I. M., Tarasov, K. V., Perumov, D. A., (2003), "Main physicochemical features of monofunctional flavokinase from Bacillus subtilis.", *Biochemistry*, **68**, (2), 177 – 181.

119. Steinborn, G., Boer, E., Scholz, A., Tag, K., Kunze, G., Gellissen, G., (2006), "Application of a wide- range yeast vector (CoMed) system to recombinant protein production in dimorphic *Arxula adeninivorans*, methylotrophic *Hansenula polymorpha* and other yeasts." *Microb Cell Fact*, **5**, 33.

120. Stephens, T. D., Bunde, C. J., Fillmore, B. J., (2000), "Mechanism of action in thalidomide teratogenesis.", *Biochemical Pharmacology,* **59,** (12), 1489 – 1499.

121. Terpe, K., (2006), "Overview of bacterial expression systems for heterologous protein production from molecular and biochemical fundamentals to commercial systems.", *Applied Microbiology and Biotechnology,* **72,** (2), 211 – 222.

122. Torres, N. V., Voit, E. O., Gelz-Alcon, C., Rodriguez, F., (1997), "An indirect optimization method for biochemical systems: Description of method and application to the maximization of the rate of ethanol, glycerol, and carbohydrate production in *Saccharomyces cerevisiae*." *Biotechnology and Bioengineering,* **55,** (5), 758 – 772.

123. Trezzani, I., Nadri, M., Dorel, C., Lejeune, P., Bellalou, J., Lieto, J., Hammouri, H., Longin, R., Dhurjati, P., (2003), "Monitoring of Recombinant Protein Production Using Bioluminescence in a Semiautomated Fermentation Process." *Biotechnology Progress,* **19**, 1377 – 1382.

124. Tsien, R. Y., (1998), "The Green Fluorescent Protein.", *Annual Review of Biochemistry*, **67**, 509 – 544.

125. Tsien Laboratory Website, University of California, San Diego, USA., http://www.tsienlab.ucsd.edu/

126. Tu, B. P., Weissman, J. S., (2002), "The FAD and O_2^- Dependent Reaction Cycle of Ero1 – Mediated Oxidative Protein Folding in the Endoplasmic Reticulum.", *Molecular Cell,* **10,** (5), 983 – 994.

127. U.S. Food and Drug Administration (FDA), (2003), Information for the Patient, HUMULIN®R, http://www.fda.gov/cder/foi/label/2003/18780scm080_humalog_lbl.pdf

128. Varma, A., and Palsson, B. O., (1994), "Metabolic Flux balancing: basic concepts, scientific and practical use." *Bio/Technology* **12**, 994 – 998.

129. Venci, D., Zhao, G., Schuman Jorns, M., (2002), "Molecular characterisation of NikD, a New Flavoenzyme Important in the Biosynthesis of Nikkomycin Antibiotics.", *Biochemistry*, **41**, 15795 – 15802.

130. Vera, J., DeAtauri, P., Cascanti, M., Torres, N. V., (2003), "Multicriteria optimization of biochemical systems by linear programming: Application to production of ethanol by *Saccharomyces cerevisiae.*" *Biotechnology and Bioengineering,* **83,** (3), 335 – 343.

131. Vinkler, C., Rosen, G., and Boyer, P. D., (1978), "Light-driven ATP formation from 32Pi by chloroplast thylakoids without detectable labeling of ADP, as measured by rapid mixing and acid quench techniques.", *Journal of Biological Chemistry*, **253**, (8), 2507 – 2510.

132. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., Gelfand, M. S., (2002), "Regulation of riboflavin biosyntheis and transport genes in bacteria by transcriptional and translational attenuation.", *Nucleic acids Research*, **30**, (14) 3141 – 3151.

133. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., Gelfand, M. S., (2004), "Riboswitches, the oldest mechanism for regulation of gene expression?", *Trends in Genetics*, **20**, (1), 44 –50.

134. Wagner, M. A., Khanna, P., and Jorns, M. S., (1999), "Structure of the Flavocoenzyme of Two Homologous Amine Oxidases: Monomeric Sarcosine Oxidase and N-Methyltryptophan Oxidase", *Biochemistry* **38**, 5588 – 5595.

135. Walsh, G., (2002), Proteins, *Biochemistry and Biotechnology*, Chapter 2, Pub: Wiley.

136. Wilback, S. J., Madadevan, R., Palsson, B. O., (2004), "Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: The "*E. coli"* spectrum.", *Biotechnology and Bioengineering,* **86,** (3), 317 – 331.

137. Yang, Y., Biedendieck, R., Wang, W., Gamer, M., Malten, M., Jahn, D., Deckwer, W. D., (2006), "High yield recombinant penicillin G amidase production and export into the growth medium using *Bacillus megaterium*.", *Microb Cell Fact*, **5**, 36.

138. Yoon, S. A., Yong-Ho, H., Kyuboem, H., (2001), "Enhancement of recombinant erythropoietin production in CHO cells in an incubator without CO_2 addition.", *Cytotechnology*, **37**, (2), 119 – 132.