The expression and purification of GlgE and GlgB, two enzymes involved in the synthesis of α-glucan in mycobacteria

and

Large subunit ribosomal protein 41 is preferentially found in *Drosophila melanogaster* larval midgut stem cells



A thesis submitted to
the
University of
Birmingham for the
award of
Masters of Research in Molecular and Cellular Biology

Sarah Maria Taylor

School of Biosciences, September 2014

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

This thesis consists of two projects which were carried out within the School of Biosciences at the University of Birmingham.

Project 1:

The four step α-glucan synthesis GlgE pathway is involved in capsular layer production in *Mycobacterium tuberculosis* and two of the enzymes in the pathway, GlgE and GlgB, which are essential to cell viability have been determined as potential targets for new drugs to treat TB. This project involved expressing and purifying *M. tuberculosis* and *M. smemgatis* GlgE and GlgB using Ni-NTA chromatography, ion exchange chromatography and size exclusion chromatography in order to be able to characterise the enzymes. An activity assay indicated both proteins are active and small integrated crystals of *M. tuberculosis* GlgE were grown by vapour diffusion.

Project 2:

Ribosomal proteins have an essential role in ribosome biogenesis and function but research is beginning to show they can also have extra ribosomal functions while not part of the ribosome. This project characterised a number of transgenic *Drosophila melanogaster* strains expressing a number of ribosomal proteins tagged with YFP at their endogenous loci in salivary gland, fat body and midgut cells. Within the midgut RpL41-YFP showed strong expression in adult midgut precursor (AMP) stem cells. An immunostain using an antibody against the endogenous RpL41 confirmed higher expression in AMP stem cells and also in a small area within the nucleus of differentiated cells, but the specificity of the antibody used has yet to be confirmed. Visualisation of these proteins was also carried out during larval development and the expression within the enterocyte nuclei was detected only during early second larval instar.

Project One

The expression and purification of GlgE and GlgB: two enzymes involved in the synthesis of α-glucan in mycobacteria

A thesis submitted in partial fulfillment for the award of

MRes in Molecular and Cellular Biology

Supervisor: Dr Klaus Fütterer

Institute of Microbiology and Infection
School of Biosciences
University of Birmingham

September 2014

Acknowledgements

I would first like to thank my supervisor, Dr Klaus Fütterer, for offering me the opportunity to carry out my research project within his lab and for his support and advice throughout the project. I would also like to thank Ali Kermani for all of his help and advice while working in the lab and Dr Luke Alderwick for his help with conducting the enzyme assays.

Abstract

Due to increasing resistance to antibiotics, TB is becoming difficult to cure and new drugs are required that can treat it successfully. The four step α-glucan synthesis GlgE pathway is involved in capsular layer production in *Mycobacterium tuberculosis* and two of the enzymes in the pathway, GlgE and GlgB, which are essential to cell viability have been determined as potential targets for new drugs to treat TB. This project focussed on purifying and characterising the recombinant proteins from *M. tuberculosis* and *M. smegmatis* in *E. coli* using Ni-NTA chromatography, ion exchange chromatography and size exclusion chromatography. *M. tuberculosis* and M. smegmatis GlgE were successfully purified and shown to be active using an enzyme assay. *M. tuberculosis* GlgE was concentrated to 6.20 mg/mL and used to set up crystal trays and *M. smegmatis* GlgE was concentrated to 9.67 mg/mL. Only one condition from JCSG plus gave a few small integrated crystals. Size exclusion chromatography of both proteins showed they are behaving as monomers in solution. Size exclusion chromatography of *M. smegmatis* GlgE at pH 6.0 did not indicate the presence of higher oligomers. The *glgB* gene from *M. tuberculosis* and *M. smegmatis* was amplified and ligated into plasmids and GlgB was expressed in pilot scale cultures of *E. coli. M. tuberculosis* GlgB did not express well but *M. smegmatis* GlgB expressed very well and was concentrated to 54 mg/mL.

Contents

1.0 Introduction 1
1.1 The *Mycobacterium tuberculosis* complex 1
1.2 *Mycobacterium tuberculosis* evolution 2
1.3 *Mycobacterium tuberculosis* infection 3
1.4 Active TB 5
1.5 TB and HIV 6
1.6 Diagnosis 6
1.7 Prevention 7
1.8 TB treatment 7
1.9 Multi-drug and extensively drug resistant TB 9
1.10 New drug discovery to treat TB 10
1.11 The mycobacterial cell envelope as a drug target 12
1.12 The mycobacterial capsule layer 14
1.13 The GlgE pathway produces capsular α-glucans 14
1.14 Potential novel anti-TB drug candidates in the GlgE pathway 17
1.15 Discovering the mechanisms of the GlgE pathway 18
1.16 Research project aims 20
2.0 Materials and methods 21
2.1 *Mycobacterium tuberculosis* GlgE expression 21
2.2 Cell lysis 21
2.3 Ni-NTA chromatography 22
2.4 Ion exchange chromatography 22
2.5 Size exclusion chromatography 22
2.6 Crystallisation 23
2.7 *Mycobacterium smegmatis* GlgE expression 23
2.8 *Mycobacterium smegmatis* GlgE chromatography 24
2.9 Western blot 24
2.10 *Mycobacterium smegmatis* plasmid screen 25
2.10.1 Thermal shift assay 26
2.11 *Mycobacterium* smegmatis GlgE activity assay 28
2.12 GlgE:Pep2 size exclusion chromatography 28
2.13 Detecting GlgE higher oligomers 28
2.14 *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* GlgB expression and purification 29

3.0 Results 31
3.1 *Mycobacterium tuberculosis* GlgE purification 31
3.1.1 Ni-NTA chromatography 31
3.1.2 Ion exchange chromatography 32
3.1.3 Size exclusion chromatography 33
3.1.4 Crystal trays 35
3.2 *Mycobacterium smegmatis* GlgE purification 35
3.2.1 Ni-NTA chromatography 35
3.2.2 Ion exchange chromatography 39
3.2.3 Size exclusion chromatography 42
3.2.4 Western Blot 44
3.2.5 Plasmid Screen 45
3.2.6 *Mycobacterium smegmatis* GlgE activity assay 49
3.2.7 Size exclusion chromatography to investigate complex formation between *Mycobacterium smegmatis* GlgE and Pep2 50
3.2.8 Size exclusion chromatography of GlgE at pH 7.0 51
3.2.9 Size exclusion chromatography of GlgE at pH 6.0 52
3.3 Cloning and expression of *M. tuberculosis* and *M. smegmatis* GlgB 53
3.3.1 PCR 53
3.3.2 Ligation 54
3.3.3 Ni-NTA chromatography 55
3.3.4 Ion exchange chromatography 58
3.3.5 Size exclusion chromatography 59
4.0 Discussion 61
4.1 *Mycobacterium tuberculosis* GlgE purification 61
4.2 *Mycobacterium tuberculosis* GlgE characterisation 64
4.3 *Mycobacterium smegmatis* GlgE purification 64
4.4 *Mycobacterium smegmatis* GlgE characterisation 68
4.5 *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* GlgB purification 71
4.6 Conclusions 73
5.0 References 74
6.0 Appendices 82

1.0 Introduction

In recent years there has been an ever increasing need to discover new drugs for the treatment of tuberculosis (TB). TB is an infectious disease that has been declared a global health emergency since 1993 and still today is a major health problem. In 2013, there were 8.6 million new reported cases of TB and 1.3 million deaths caused by it worldwide (World Health Organisation, 2013). Although mortality rates have fallen by 45% since 1993, recent and emerging factors such as antibiotic resistance and the HIV epidemic pose problems in future success of prevention and treatment of the disease.

1.1 The *Mycobacterium tuberculosis* complex

The main causative organism of TB in humans is the bacillus *Mycobacterium tuberculosis.* There are also a number of other *Mycobacterium* species that cause TB in humans and animals which together make up the mycobacterium tuberculosis complex (MTBC). *M. africanum* affects humans and is mainly found in West Africa (De Jong *et al*., 2010) while *M. cannetti* is a smooth tubercle bacillus also found to infect humans but is rare (Koeck *et al*., 2011). *M. bovis* mainly causes bovine TB in cattle and affects humans as well as a range of mammalian hosts, it also gave rise to the strain *M. bovis* BCG, an attenuated version that is the most widely used vaccine to prevent TB (Garnier *et al*., 2003). Other strains that make up the MTBC are *M. microti* which affects voles, *M. pinnipeddi* which infects seals and *M. mungi* which was recently identified to infect mongoose (Alexander *et al*., 2010).

The bacillus *Mycobacterium tuberculosis* was first identified in 1882 by Robert

Koch. DNA/DNA hybridisation studies identified the other mycobacterium species which make up the MTBC and found DNA between the species was related by at least 90% (Imaeda, 1985) and 16S rRNA sequence comparisons were discovered to be identical (Böddinghaus *et al.,* 1990). Although similar, each species differs by their host range, their pathogenicity and by phenotypes (Brosch *et al*., 2002).

1.2 *Mycobacterium tuberculosis* evolution

The first recorded account of TB in humans was in early Chinese and Egyptian manuscripts (Gillespie, 2007) and paleopathologic evidence has indicated the disease affected humans as long ago as 3700 BC in Egypt and also in pre-Columbian North and South America and Europe (Kapur *et al*., 1994). DNA analysis of a 1000 year old pre-Columbian Peruvian mummy was the first specific evidence of TB infection (Salo *et al*., 1994).

Before the availability of whole genome sequencing, it was first hypothesised that *Mycobacterium tuberculosis* evolved from *M. bovis* with the establishment of agriculture during the Neolithic period (Stead *et al*., 1995). Once the whole genome of *Mycobacterium tuberculosis* was sequenced however, comparative genomic studies discovered *M. bovis* has many deletions in its sequence relative to *M. tuberculosis* and therefore has a smaller genome which contradicted the original view (Cole *et al*., 1998).

It was also hypothesised that 15,000-20,000 years ago the speciation of the MTBC underwent an evolutionary bottleneck in that there was at some point a sharp reduction in the mycobacterium population and therefore a reduction in the gene

pool as it was found that many structural genes throughout the complex were conserved (Sreevatsan *et al*., 1997). This was also proved incorrect as comparative genomics in the MTBC indicated many variable regions (Brosch *et al*., 2002) and it has recently been discovered that human tuberculosis first emerged from Africa, 70,000 years ago (Comas *et al*., 2013).

1.3 *Mycobacterium tuberculosis* infection

Mycobacterium tuberculosis is considered an acid-fast Gram-positive bacillus as it has no outer cell membrane but does have a cell wall which contains a high lipid content and this inhibits Gram staining. The bacilli are able to spread easily through the air such as by coughing or sneezing and droplets of mucus containing just one bacillus is all that is required for TB infection (Riley *et al*., 1995). The bacilli are highly aerobic and divide slowly (every 15-20 hours) and infection is most commonly local within the lungs but can also be extra-pulmonary. Viable bacilli can infect individuals with no disease development and the person can be asymptomatic, this is known as latent TB and *M. tuberculosis* has evolved with the human hosts' immune system to develop evolutionary novel survival mechanisms.

When *M. tuberculosis* is breathed in the bacilli enter the airways and travel down into the lungs. They are then phagocytosed by alveolar macrophages via a number of receptor molecules (Pieters, 2008). Alveolar macrophages play a critical role in host defence and are able to respond to foreign materials by engulfing them and initiating an immune response, they also control homeostatis within tissue and have a role in tissue modelling (Lambrecht, 2006). While some macrophages are able to activate the immune response and interact with acidic, hydrolytically active lysosomes (Rohde *et al*., 2007) that degrade them along with their toxic and

inflammatory contents, others containing viable *M. tuberculosis* can evade the immune response.

To prevent the activation of the immune response *M. tuberculosis* is able to induce granuloma formation (Flynn and Chan, 2003). The granuloma is not entirely understood and has proven difficult to study in mouse models due to the fact that mice are unable to form the complex granuloma structures that humans do and do not exhibit necrosis and caseation (Russell *et al*., 2009).

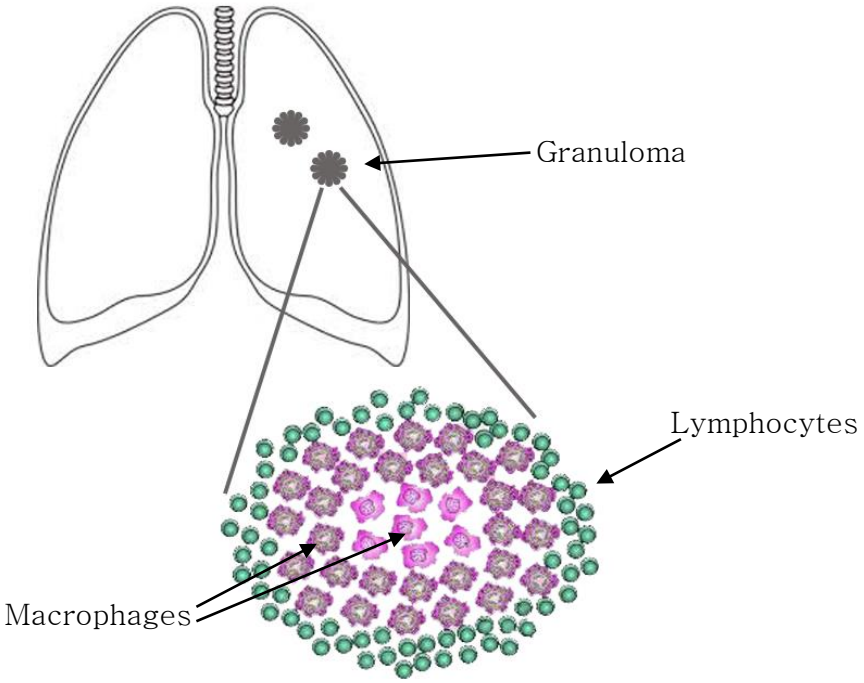


Figure 1. *M. tuberculosis* bacilli are inhaled into the lungs and engulfed by alveolar macrophages. The organism avoids an immune response by forming a granuloma. Non-activated macrophages remain in the centre of the granuloma while activated macrophages recruit T lymphocytes which secrete chemokines that maintain activation of macrophages and immune cell recruitment. [Image adapted from (Pieters, 2008)].

It is thought that the granuloma is able to contain the infection within non-activated macrophages in the centre of the granuloma which are surrounded by more macrophages and T lymphocytes (Pieters, 2008). Activated macrophages are speculated to surround the non-activated macrophages where T lymphocytes can be activated by mycobacterial antigens via T cell receptors (Kaufmann, 1993). With this activation, chemokines are secreted which in turn maintain macrophage activity and immune cell recruitment (Russell, 2007). The granuloma is able to maintain itself in this way for a whole life time within the host. It is not yet known whether the *M. tuberculosis* within the granuloma are actively dividing or dormant (Saunders and Britton, 2007).

1.4 Active TB

When a host's immune system becomes compromised the disease is able to infect the host and become life-threatening active TB. The granuloma caseates when the immune system is compromised and mycobacteria are freed and can spread by irritating the airways and causing the host to cough (Russell, 2007).

In the 18th and 19th centuries TB was a major cause of death, as the industrial revolution brought cramped living conditions and poor sanitation. As a result the disease was able to spread quickly (Cole *et al*., 1998). As the western world developed new drugs to treat TB and a vaccine to prevent it, TB became curable. In the developing world however TB was still a huge burden with limited access to treatment and poor living conditions which is still true throughout the developing world today. Old age is another factor that causes the immune system to deteriorate and TB that has been latent for years can become active (Lillebaek *et al*., 2002).

1.5 TB and HIV

HIV is a disease that by itself has high mortality rates especially in the developing world. As the immune system is compromised with HIV infection, the risk of catching TB or latent TB becoming active TB increases greatly, which makes the two together a 'cursed duet' which is especially prevalent in African regions (Murray, 1990). Having HIV increases the chances of developing active TB by 12 to 20 times that of a person without HIV (WHO, 2013). Last year, 1.1 million new cases of TB were reported in individuals with HIV (World Health Organisation, 2013).

1.6 Diagnosis

Latent TB although asymptomatic requires treatment to minimise risk of it becoming active TB. Also, it is important to estimate the burden of TB so as to assess the most high risk areas (Sharma and Mohan, 2013). There are two commonly used diagnostic tools used to detect latent TB: the tuberculin skin test and interferon-gamma release assays (Rangaka *et al.,* 2012) although in high burden countries it has been found the incidence of false positives is quite high (Sharma and Mohan, 2013).

Active TB is commonly detected using sputum culture but diagnosis can prove difficult as there has to be a threshold concentration of TB bacilli for a smear positive result. Culture methods have improved over the years and what once took six to eight weeks to culture now can take around 10 days (Rodrigues *et al.,* 2009). The microscopy used to detect bacilli has been upgraded to fluorescence LED microscopy which has improved sensitivity (Dheda *et al*., 2013). A range of imaging techniques are also used such as chest x-rays or magnetic resonance

imaging (MRI) in order to assess the extent of infection.

The nucleic acid amplification test (NAAT) for TB is a fast tool that can detect mycobacterial nucleic acid in three to six hours and is a more sensitive method that uses the polymerase chain reaction (PCR). It is however expensive and difficult to use in the developing world as it requires infrastructure and scientific knowledge to run the tests (Ling *et al*., 2008).

In developing countries a technique called loop mediated isothermal amplification (LAMP) is a promising molecular diagnostic test with high sensitivity and is easy to use (Boehme *et al*., 2007). Breath tests are also being investigated to detect volatile organic compounds that are produced by *Mycobacterium tuberculosis* (Phillips *et al.*, 2012).

1.7 Prevention

The only vaccine currently available is the Bacillus Calmette-Guérin (BCG) vaccine but resistance to TB is variable and may not protect against more virulent strains of TB (Fine, 1995). At this time, there are a number of new candidate vaccines in different stages of development. MVA85A for example which is a recombinant strain of Vaccinia Ankara virus expressing the immunodominant *M tuberculosis* protein, antigen 85A is in phase II trials and has been designed to boost the effectiveness of the BCG vaccine (Tameris *et al.,* 2013).

1.8 TB treatment

It has only been since 1944 that a cure for TB was made available in the form of the antibiotic streptomycin. Streptomycin was used on its own initially to treat TB but

it was soon noticed that the organism was becoming resistant to the drug (Silva and Palomino, 2011). This highlighted the need for a multidrug therapy to combat resistance and streptomycin was then used in combination with para-amino salicylic acid (PAS) over a regimen of eighteen months ("Treatment of Pulmonary Tuberculosis with Streptomycin and Para-Amino-Salicylic Acid," 1950). Following this ethambutol and rifampicin were discovered and began to be used for treatment. Rifampicin replaced the use of PAS which allowed treatment regimens to be reduced from six to nine months (Sharma and Mohan, 2013). Since this initial upsurge in antibiotics discovery, the last few decades have yielded no new drugs to treat TB and it is only very recently new drugs are being put through clinical trials.

Treatment today is by directly observed treatment that is short course (DOTS) and consists of the use of four antibiotics; isoniazid, rifampicin, pyrazinamide and ethambutol for two months. After this initial phase isoniazid and rifampicin are continued for another four months. Using multi-drug treatment regimens increases the chances of successfully treating TB by targeting the organism in different ways and lowering the risk of drug resistance.

Isoniazid acts by inhibiting the synthesis of mycolic acid, a fatty acid that makes up a large part of the mycobacterial cell wall (Sethi *et al*., 2010) while ethambutol also targets the cell wall by obstructing cell wall formation (Takayama *et al*., 1979). Rifampicin is a lipophilic ansamycin which inhibits the β-subunit of RNA polymerase disrupting mRNA (Rattan *et al*., 1998). Pyrazinamide is a pro-drug that gets converted to pyrazinoic acid by pyrazinamidase (PZase) which is produced by *M. tuberculosis* and then inhibits the bacterial membrane disrupting transport (Silva and Palomino, 2011). Most people will be cured using this four-drug regimen

but there is an increasing amount of multidrug resistance which poses an ever increasing threat to health.

Second line drugs such as fluoroquinolones and injectable drugs such as amikacin, kanamycin and capreomycin are also used to treat TB but this treatment is expensive, harder to administer and can have toxic effects (Silva and Palomino, 2011).

1.9 Multi-drug and extensively drug resistant TB

In 2012, there were 450,000 new cases of multi-drug resistant TB (MDR-TB) and 170,000 deaths (WHO, 2013). MDR-TB is resistant to at least both isoniazid and rifampicin and rates of cure are 50-70% which is a lot lower than drug treatable TB which is ~95% (Koul *et al.*, 2011). There was at least one case of MDR-TB in 92 countries by the end of 2012 with India and China making up the majority of MDR cases with more than 50% of the total cases being from these countries. The treatment regime for MDR-TB is increased to 20 months and includes second line drugs (WHO, 2013).

Extensively drug resistant TB (XDR-TB) is an even more worrying threat to successful TB treatment and is resistant to not only the first line antibiotics but also to fluoroquinolones and other second line drugs (Gandhi *et al*., 2010). Treatment of XDR- TB involves the use of third line drugs such as clofamizine, linezolid and amoxicillin plus clavulanate which are used when all else fails due to the fact they are expensive and have serious side effects (Palomino and Martin, 2013). Survival rates of XDR-TB are very low and involve vigorous treatment regimes. 9.6% of MDR cases are thought to be XDR-TB (WHO, 2013).

Advances in molecular biology and genome sequencing techniques has

allowed investigations into drug resistance and has detected many different mutations that make the bacilli resistant to the range of drugs available (Silva and Palomino, 2011).

M. tuberculosis contain no plasmids so most mutations must occur spontaneously mainly at chromosomal genes (Gillespie, 2007). Non-compliance by patients to the long treatment regimens also increases the risk of drug resistance.

In the last few years cases of strains of *M. tuberculosis* have been reported to be totally resistant to any drug that is available (Velayati *et al*., 2009) (Velayati *et al*., 2013) (Migliori *et al*., 2007). There has however been hesitation on whether to label this as 'totally drug-resistant tuberculosis' as current emerging drugs in clinical trials may be able to treat this (Palomino and Martin, 2013).

1.10 New drug discovery to treat TB

Treatment options for X/MDR-TB are decreasing over time with increasing resistance to current drugs and there is an urgent requirement for new drugs to treat TB successfully. An ideal drug would have a short treatment regime that is easy for patients to adhere to and be able to treat MDR and XDR-TB strains. A new drug would ideally be able to be taken along with anti-retroviral therapy medications for HIV without any drug-drug interactions and reduce the amount of pills needed as to reduce compliance issues (Koul *et al*., 2011).

With investment and molecular biology technology advancements there has been growth in the anti-TB drug pipeline in recent years with some drugs even in phase II and III clinical trials (Zumla *et al.* 2013). In December 2012, the US Food

and Drug Administration (FDA) approved the use of the drug Bedaquiline to treat MDR-TB but only as part of a combinatory regime with other drugs where there are no other treatment options available (Mahajan, 2013). The drug delamanid is also a new drug that is in phase III clinical trials which inhibits mycolic acid synthesis for use to treat MDR-TB (Gler *et al*., 2012).

A number of screening strategies have been utilised to screen for potential anti-mycobacterial targets. Target based approaches have been tried based on the *M. tuberculosis* genome so far with little success. This is due to the fact identifying target genes essential to the organism does not ensure it will be a druggable target and it has been found inhibitors cannot be identified for a lot of screening hits (Payne *et al*., 2007). Whole cell phenotypic screens have proved more successful and have produced a number of new drug candidates such as benzothiazines and diarylquinolines (Koul *et al.*, 2007) (Makarov *et al*., 2009). Even so, the mechanisms of action remains unknown therefore analysis is hampered and many hits have been found to produce non-specific or cytotoxic effects (Koul *et al*., 2011). Pathway inhibitor screens using high throughput screening technologies and compound libraries have been utilised to search for inhibitors of pathways that have already been analysed (Weinstein *et al*., 2005).

There has also been an interest to review and reengineer drugs that are already used to treat TB. With mechanisms of drugs being slowly being deduced using analytical tools, drugs such as rifampicin and linezolid which are now known to target the *M. tuberculosis* cell wall have been reengineered to increase their potency. Re- engineering drugs by adding or removing chemical groups to the drugs scaffold will hopefully improve their efficacy, resistance and bactericidal activity which would also hopefully mean the drug could be administered at a

lower dosage in turn reducing cytotoxic effects (Fischbach and Walsh, 2009). Linezolid, a second line drug to treat TB has been re-engineered to improve its performance as a drug in multi-drug regimens (Williams *et al*., 2009). Re-engineering existing drugs may help to ease the burden of TB but increasing and emerging resistance shows that is especially important that new and novel drugs are found to combat resistance and ensure TB treatment has a successful future.

1.11 The mycobacterial cell envelope as a drug target

The mycobacterial cell envelope is unique and complex in relation to other bacterial cells. It is thick, waxy, and hydrophobic and contains large amounts of mycolic fatty acids giving the organism a permeable barrier which is beneficial to its survival and virulence within the host. Linked covalently to the plasma membrane is the cell wall core, known as the mycolyl-arabinogalactan-peptidoglycan complex (mAGP). It is made up of a peptidoglycan layer covalently linked to arabinogalacatan which is covalently attached to long and short chain mycolic acids (Kaur *et al*., 2009).

The outer section of the cell envelope is linked covalently to the mAGP and is made up of free glycolipids which are interspersed throughout such as trehalose dimycolates (TDM, also known as cord factor), trehalose monomycolates (TMM), sulfolipids, phosphatidylinositol mannosides (PIMs), pthiocerol dimycocerosate (PDIM), phenolic glycolipids (PGLs) and the pthiocerol containing lipids; lipomannan (LM) and lipoarabinomannan (LAM) (Brennan and Crick, 2007). The architecture of the cell envelope is illustrated in figure 3.

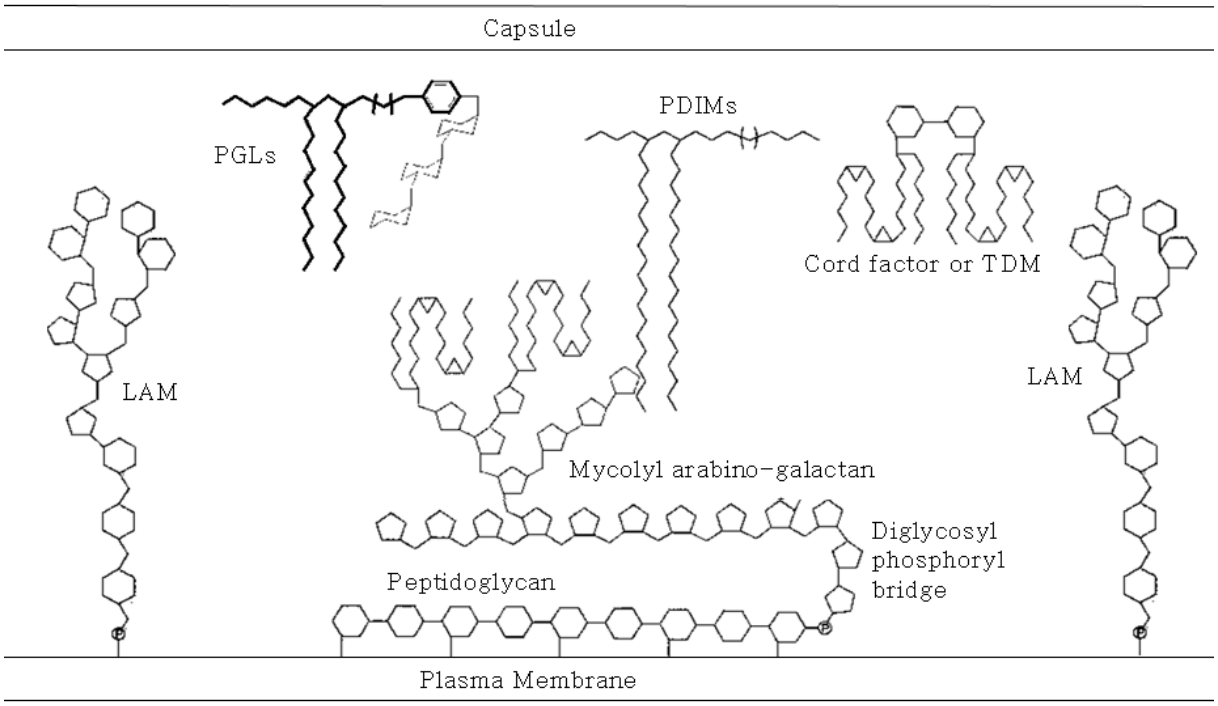


Figure 3. The mycobacterial cell wall [Figure adapted from (Favrot *et al.,* 2012)].

It is the outer glycolipid layer that gives *M. tuberculosis* a permeable barrier and allows the organism to survive in harsh surroundings and prevents entry to many solutes. This permeable barrier also hinders drugs which may have been found to inhibit enzyme activity within the cell but are unable to pass the permeable barrier to have any advantageous effect in curing TB (Koul *et al*., 2011).

When the mechanisms of action of existing drugs used to treat TB for many years were analysed it was found that many targeted the cell wall such as the antibiotics isoniazid and ethambutol (Chopra and Brennan, 1998). Today, the cell envelope is still an attractive target to search for novel therapeutics.

1.12 The mycobacterial capsule layer

The outermost layer of the cell envelope is a non-covalently linked, extractable capsule made up of polysaccharides, proteins and lipids some of which are found to be key to the organisms pathogenicity (Chapman *et al*., 1959; Daffé and Etienne, 1999) (Figure 3). Although it was known that the capsule layer of mycobacteria existed, it was not until 2010 that its presence was confirmed experimentally using plunge freezing cryo- electron microscopy due to the fact that the capsule layer is easily shed in detergent containing liquid culture (Sani *et al*., 2010). The capsule layer is made up predominantly of α-glucan, arabinomannan, oligomannsoyl-capped glycolipids and proteins from the ESX-1 secretion system (Sani *et al*., 2010).

1.13 The GlgE pathway produces capsular α-glucans

Eighty to ninety percent of the capsule layer is made up of α-glucans which are composed of α(1->4) linear glucans and α(1->6) branched glucans. In mycobacteria, three different metabolic pathways are known to produce α-glucans. These are the classical GlgC-GlgA pathway, the Rv3032 pathway and the more recently identified GlgE pathway, all of which can be visualised in figure 4 (Chandra *et al*., 2011). As of yet it is still unknown exactly how much each pathway contributes to α-glucan production.

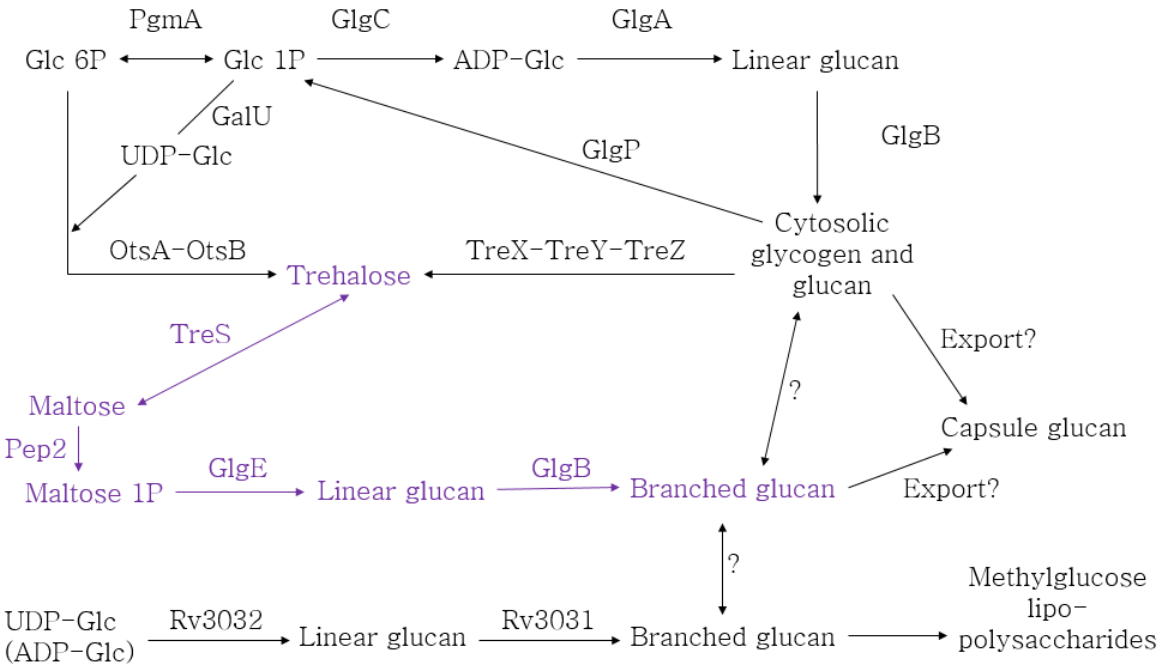


Figure 4. The α-glucan pathways of *M. tuberculosis* consist of the GlgC-GlgA pathway, the GlgE pathway (highlighted in purple) and the Rv3032 pathway [Image adapted from (Chandra *et al.*, 2011)].

A *glgA* knockout in a mouse model indicated that α-glucans are a requirement for the capsular layer production as *M. tuberculosis* persistence was impaired when capsular glucan levels decreased as a result of the knockout (Sambou *et al*., 2008).

The GlgE pathway was only recently discovered in 2010 and begins at the disaccharide trehalose which is converted to the disaccharide maltose and is catalysed by the enzyme trehalose synthase (TreS). The maltokinase enzyme Pep2 then catalyses the phosphorylation of maltose to maltose-1-phosphate which is then converted into an α(1,4)-glucan linear chain by the glycosyltransferase GlgE. The linear α-glucan becomes branched α-glucan (α(1,6)-glucan) by the glycosyltransferase GlgB (Kalscheuer *et al*., 2010). The GlgE pathway is visualised in figure 5.

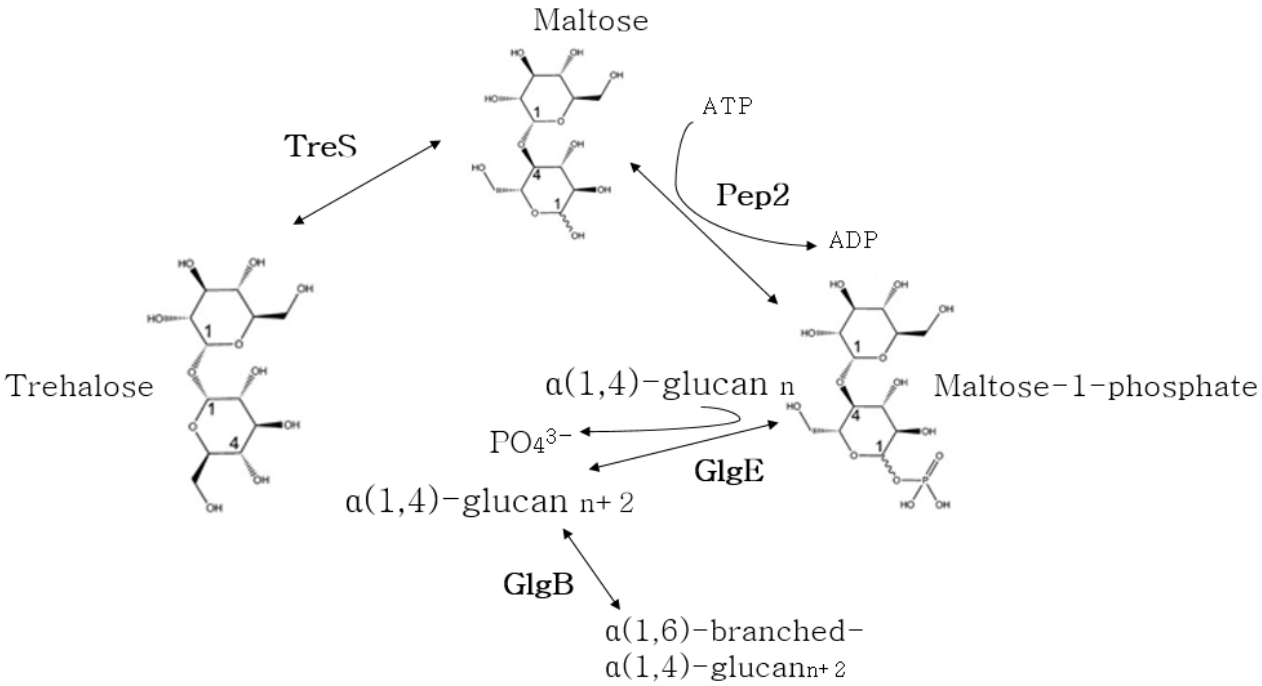


Figure 5. The GlgE Pathway of mycobacterial α-glucan synthesis [Image adapted from (Roy *et al.* 2013)].

Comparative genome analysis of the three pathways found that fourteen percent of the genomes of diverse bacteria contained genes for the GlgE pathway indicating that the newly discovered pathway is half as common as the classical GlgC-GlgA pathway in contributing to the capsular layer (Chandra *et al*., 2011).

Traditional and chemical reverse genetics found that TreS and RV3032 in the GlgE and Rv3032 pathways respectively are involved in a synthetic lethal pathway, as inactivation of either alone has no effect on persistence but both inactivated led to mycobacterial cell death (Kalscheuer *et al*., 2010).

1.14 Potential novel anti-TB drug candidates in the GlgE pathway

Inactivating either *glgE* or *glgB* genes in the GlgE pathway led to the accumulation of maltose-1-phosphate which is toxic and lethal to the cell if it is over produced. Interestingly the organism was found to self-amplify maltose-1-phosphate production due to a stress response in a mouse infection model (Kalscheuer *et al*., 2010). *glgB* was also shown to be an essential gene after a genome wide transposon mutagenesis study demonstrated it is essential for normal *M. tuberculosis* growth (Sassetti *et al*., 2003).

With the involvement of GlgE in a synthetic pathway with Rv3032 and the essentiality of the *glgE* and *glgB* genes, the GlgE pathway has been highlighted as a pathway within which GlgE and GlgB have been genetically validated as potential novel anti-TB drug target candidates. GlgE is thought to be the better of the two candidates due to the fact that GlgE homologues are not present in mammals therefore the mycobacterium protein can be targeted without interfering with humans.

GlgE, the defining enzyme of the GlgE pathway, is an α-maltose 1-phosphate: (1→4)- α-D-glucan 4-α-D-maltosyltransferase of the CAZy (http://cazy.org) glycoside hydrolase subfamily GH13_3. This is known as although the crystal structure of *M. tuberculosis* GlgE has to date not been confirmed using x-ray crystallography or NMR, Syson et al. have previously confirmed the crystal structure of *Streptomyces coelicolor* isoform I GlgE which has a similar gene sequence to *M. tuberculosis* GlgE (Syson *et al*., 2011). It was discovered that *S. coelicolor* isoform I GlgE and *M. tuberculosis* GlgE have similar kinetic and catalytic properties which has allowed the *S. coelicolor* GlgE model to be used as a framework for understanding *M. tuberculosis* GlgE mechanisms and allow for

development of potential inhibitors for therapeutic effect (Sengupta *et al.,* 2014; Syson *et al*., 2014, 2011). Very recently, two papers have reported finding the first inhibitors of GlgE (Veleti *et al*., 2014a, 2014c).The *M. tuberculosis* glycogen branching enzyme, GlgB, also known as amylo(α1→4- 6)transglycosylase belongs to carbohydrate binding module family 48 and glycoside hydrolase family 13 (GH13) (Cantarel *et al*., 2009). *M. tuberculosis* GlgB was first expressed and characterised in 2006 (Garg *et al*., 2007) and the crystal structure for full length *M. tuberculosis* GlgB was elucidated in 2010 (Pal *et al*., 2010). To date no inhibitors have been discovered for this enzyme.

1.15 Discovering the mechanisms of the GlgE pathway

To aid the discovery of molecules which inhibit the identified potential anti-TB drug target candidates *M. tuberculosis* GlgE and GlgB, it is important to understand the GlgE pathway mechanisms on a molecular level.

By confirming the crystal structures of the enzymes involved in the GlgE pathway, a better insight into the how the reaction steps of the pathway take place can be formed and molecular models of the proteins can be analysed in depth and understood. So far in the GlgE pathway, *M. tuberculosis* GlgB (Pal *et al*., 2010) and TreS (Roy *et al*., 2013) have both had their crystal structures elucidated while GlgE has had its structure designed based on the crystal structure of *S. coelicolor* isoform I GlgE (Sengupta *et al*., 2014; Syson *et al*., 2011) but the true *M. tuberculosis* crystal structures of GlgE and Pep2 have yet to be confirmed.

A recent study by Roy *et al*. (2013) which determined the crystal structure of *M. tuberculosis* TreS also demonstrated that TreS forms tetramers in solution by using small angle x-ray scattering. This paper also showed data that was indicative

of TreS and Pep2 forming a hetero-octamer which is built upon the TreS tetramer and this complex formation was found to accelerate the activity of Pep2 (Roy *et al*., 2013). This study also suggested that the GlgE pathway could be controlled as a 'multicatalytic machine' that may involve all of the enzymes involved in the pathway.

In order to continue research with the aim to characterise the enzymes involved in the GlgE pathway and understand how they may function within the pathway as well as to grow protein crystals in order to determine protein structure, a suitable amount of fresh and pure protein must be obtainable.

1.16 Research project aims

The aims of this project were:

- To express, purify and characterise the recombinant protein GlgE from *Mycobacterium tuberculosis*.
- To express, purify and characterise recombinant GlgE from *Mycobacterium smegmatis*.
- To clone and ligate the *M. tuberculosis* and *M. smegmatis glgB* gene into a suitable plasmid before expressing and purifying the GlgB proteins on a pilot scale.

2.0 Materials and methods

2.1 *Mycobacterium tuberculosis* GlgE expression

Plasmids (pET28A, Novagen) containing the *M. tuberculosis glgE* and *M. Smegmatis glgE* gene inserts were provided by Ali Kermani (School of Biosciences). The plasmids were heat transformed (45s at 42 °C in a water bath) into BL21(DE3) competent cells (Life Technologies) and cultured on agar plates (LB/kanamycin 25 μgmL^{-1}). A single colony was used to inoculate LB broth (25 gL^{-1}) containing kanamycin (25 μgmL^{-1}) and incubated overnight at 37 °C, with shaking. The liquid culture was centrifuged and re- suspended in fresh Terrific Broth media (see appendices) with kanamycin (50 μgmL^{-1}) and incubated at 37 °C with shaking until OD_{600} = 0.4 – 0.6. The temperature was lowered to 16°C for 3h and protein expression was then induced using 1 mM isopropyl β-D-1- thiogalactopyranoside (IPTG) (Bioline). Cultures were incubated overnight at 16 °C with shaking. Cells were harvested using centrifugation (7000 g for 10 mins at 4 °C), washed with phosphate buffered saline (PBS) and re-suspended in lysis buffer (20 mM Na_2HPO_4 pH 7.2, 500 mM NaCl, 10% (v/v) glycerol). The pellets were frozen at -80°C until required.

2.2 Cell lysis

Pellets were thawed and re-suspended with 0.5 $mgmL^{-1}$ lysozyme (from chicken egg white, Sigma Aldrich), 10 µg/mL DNase I, 10 mM $MgSO_4$, 1 mM 2-Mercaptoethanol, protease inhibitor cocktail (Roche) and 1 mM PMSF. The pellet was passed three times through a French Press (Thermo Spectronic FA-078). The crude lysate was centrifuged at 48,000 g for 30 mins at 4°C and the supernatant (cleared lysate) collected.

2.3 Ni-NTA chromatography

The cleared lysate from a 5 L culture was diluted 3-fold with buffer A (20 mM Na_2HPO_4 pH 7.2, 500 mM NaCl, 10% (v/v) glycerol), filtered (0.45 µm pore size) and loaded on a pre-equilibrated Ni-NTA column (1 mL, GE Healthcare Life Sciences). The column was washed with buffer A (5 column volumes) and then buffer A plus 20 mM, 40 mM, 80 mM, 100 mM, 150 mM, 200 mM, 300 mM and 500 mM imidazole, respectively. Fractions were analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and fractions containing GlgE were pooled and used step 2.4.

2.4 Ion exchange chromatography

The pooled fractions were concentrated down to ~10 mL using Amicon Ultra-4 centrifugal filter units (50 kDa) and diluted 10-fold using Buffer B (100 mM Bis-tris propane pH 7.0, 10% (v/v) glycerol) to reduce the salt concentration to 50 mM NaCl and then filtered (0.45 µm). The eluate was applied to a pre-equilibrated HiTrap Q- column (1 mL, GE Healthcare Life Sciences). The column was washed using buffer B (5 column volumes) followed by buffer B plus 100 mM, 150 mM, 200 mM, 250 mM, 300 mM, 400 mM and 500 mM NaCl, respectively. Fractions were analysed using SDS- PAGE and fractions containing GlgE were pooled and concentrated using Amicon Ultra-4 centrifugal filter units (50 kDa) to <1 mL.

2.5 Size exclusion chromatography

Concentrated fractions were loaded onto a HiPrep Sephacryl 26/60 S-200HR column (GE Healthcare Life Sciences) pre-equilibraited with buffer (20 mM bis-tris

propane pH 7.0, 10% (v/v glycerol, 150 mM NaCl) using ÄKTAprime plus. Size exclusion chromatography was ran using ÄKTAprime plus and 5 mL fractions collected. Fractions were analysed by SDS-PAGE and fractions containing GlgE concentrated using Amicon Ultra-4 centrifugal filter units (50 kDa) to <1 mL. Protein concentration was measured using the Beer-Lambert law using a spectrophotometer.

2.6 Crystallisation

A structure screen to grow crystals of GlgE was carried out using Structure Screen, JCSG plus, MIDAS and MORPHES (Molecular Dimensions) screening solutions in 96- well plates. A mosquito liquid handling system (TTM Labtech) set up crystallization drops containing 100 nL of GlgE (6.20 mg/mL). Trays were kept at 18°C for all crystallisation attempts.

2.7 *Mycobacterium smegmatis* GlgE expression

Plasmids (pET28a, Novagen) containing the *M. smegmatis glgE* gene were provided by Ali Kermani (School of Biosciences). The plasmids were heat transformed into BL21(DE3) and also BL21(DE3)pLysS competent cells (Life Technologies) using the same method as for *M. tuberculosis* GlgE. Cell culture, protein expression induction and cell lysis was carried out using the same methods as section 2.1 and 2.2.

2.8 *Mycobacterium smegmatis* GlgE chromatography

Ni-NTA chromatography was repeated using the same method as before for *M. tuberculosis* GlgE (section 2.3) but substituting the clear lysate for *M. smegmatis*. Ion exchange was also carried out using the same method (section 2.4) but flow through was collected and applied again through the HiTrap Q Column (1 mL, GE Healthcare Life Sciences). Following this the flow through that was collected was also applied through a Hi Trap SP Column (1 mL, GE Healthcare Life Sciences) also using the same method. Fractions containing GlgE were pooled and concentrated using Amicon Ultra-4 centrifugal filter units (50 kDa).

2.9 Western blot

After Ni-NTA chromatography, fractions containing 40 mM to 500 mM imidazole were run on a 10% SDS-polyacrylamide gel and then transferred to a 0.2 µM polyvinylidene fluorise (PVDF) membrane (Amersham, GE Healthcare) by electrophoresis (105 mAmp, 180 mins) in western buffer (see appendices). The membrane was blocked using 3% bovine serum albumin (BSA) in 1x phosphate buffered saline with 0.1% Tween® 20 (v/v) (PBST) and incubated overnight at room temperature with shaking. Primary antibody anti-His6 (1:1000) (Sigma Aldrich) was added with 3% BSA and 1x PBS and incubated at room temperature with shaking for 1.5 hrs. The membrane was then incubated with the secondary antibody (Antimouse IgG alkaline phosphatase, Sigma Aldrich) (1:25000) with 3% BSA and 1X PBS for 1.5 hrs at room temperature with shaking. The membrane was then developed using BCIP/NBT (Sigma Fast™).

2.10 *Mycobacterium smegmatis* plasmid screen

The ligation reaction mixture from when the PCR amplified *M. smegmatis glgE* gene was ligated into plasmids (pET28a) was supplied by Ali Kermani (School of Biosciences) and heat transformed into Top 10 cells (Life Technologies). Cells were cultured on agar plates (LB/kanamycin 25 μgmL^{-1}) and 6 colonies were picked to inoculate 6x 5 mL LB broth (25 gL^{-1}) containing kanamycin (50 μgmL^{-1}) which were incubated overnight at 37 °C with shaking. Plasmids from each over-night culture were purified (QIAprep spin Mininprep kit) according to the manufacturer's instructions and each heat transformed into BL21(DE3) cells and cultured on agar plates (LB/kanamycin 25μgmL^{-1}). A single colony from each of the 6 plates was then used to inoculate 6xLB broth (25 gL^{-1}) containing kanamycin (25 μgmL^{-1}) and incubated overnight at 37 °C with shaking). The liquid cultures were centrifuged and the 6 pellets were each re-suspended and cultured in fresh TB broth (6 x 1 L) (see appendices) containing kanamycin (50 μgmL^{-1}). From this point culture followed the same method as previous (section 2.1) apart from 3 of the pellets were re-suspended in a different lysis buffer consisting of 25 mM HEPES pH 7.0, 500 mM NaCl, 10% (v/v) glycerol. 3 of the clear lysates were applied to a pre- equilibrated Ni-NTA column and proteins eluted using the same method as for section 2.3. The remaining 3 clear lysates were also followed the same method apart from a different buffer A (25 mM HEPES pH 7.0, 500 mM NaCl, 10% (v/v) glycerol) was used. SDS-PAGE gels were used to assess the protein content of each of the samples.

2.10.1 Thermal shift assay

This experiment was carried out by Ali Kermani (School of Biosciences). *M. smegmatis* GlgE of concentration 35 μgmL^{-1} was added to a range of buffers (150 mM) (see table at high (250 mM) and low (100 mM) NaCl concentrations. 10 μL of each assay solution containing SYPRO orange dye stain (Life Technologies) was added to the wells of a 96 well plate. The assay was ran from 25 °C to 85 °C at 1 °C intervals for 1 min at each temperature using a Q-PCR machine (Mx3005P, Agilent Technologies) and was repeated twice. Melting temperatures (T_m) were determined by the software.

Buffer (150 mM)	pH
Sodium citrate/citric acid	4.0
Sodium acetate	4.2
Sodium acetate	4.8
Sodium citrate/citric acid	5.0
Sodium acetate	5.4
MES	5.7
Sodium citrate/citric acid	6.0
Bis-Tris	6.0
MES	6.1
Imidazole	6.4
MES	6.5
Bis-Tris	6.5
Bis-Tris	7.0
HEPES	7.0
Imidazole	7.1
Tris	7.2
HEPES	7.5
Imidazole	7.6
HEPES	8.0
Tris	8.0
Tris	8.6
CHES	8.6
CHES	9.1
CHES	9.6

Table 1. A table showing all of the buffers used along with their pH used in the activity assay with *M. smegmatis* GlgE. Each buffer was at a concentration of 150 mM.

2.11 *Mycobacterium smegmatis* GlgE activity assay

Reaction mixtures probing *M. smegmatis* GlgE activity were in a final volume of 200 µL, containing 20 mM buffer (same as used in SEC) (pH 7.0), 25 µL 1 mM MESG, 4 µL PNP (1U), 0.06 $mgmL^{-1}$ GlgE and were initiated by the addition of 10, 20 and 30 µL Maltose-1- phosphate (1x). Reactions were all at 37 °C and run for 334 s using a BMG PHERastar FS microtiter plate reader. Two control runs were carried out, one without GlgE and the other without maltose-1-phosphate.

2.12 GlgE:Pep2 size exclusion chromatography

Ali Kermani (School of Biosciences) carried out size exclusion chromatography of GlgE and Pep2 which were concentrated and loaded onto a HiPrep Sephacryl 26/60 S- 200HR column (GE Healthcare Life Sciences) pre-equilibraited with buffer (20 mM bis- tris pH 6.0, 10% (v/v glycerol, 150 mM NaCl) using ÄKTAprime plus and absorbance at 280 nm measured.

2.13 Detecting GlgE higher oligomers

A 10 L culture of BL21(DE3) competent cells transformed with plasmids containing the *M. smegmatis glgE* gene was grown, lysed and was purified following the same methods as used previously (sections 2.1, 2.2, 2.3 and 2.4). Concentrated protein was then aliquoted into two and dialysed into two buffers: 50 mM Bis-tris propane pH 7.0, 200 mM NaCl, 10% (v/v) glycerol and 20 mM Bis-tris pH 6.0, 200 mM NaCl, 10% (v/v) glycerol. Size exclusion chromatography was then carried out using the same method as previously used (section 2.5) but using the relevant buffers just mentioned. Fractions were analysed using SDS-PAGE.

2.14 *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* GlgB expression and purification

The gene sequences for *M. tuberculosis* and *M. smegmatis* GlgB were identified using the J. Craig Venter Comprehensive Microbial Resource (now offline) and confirmed using EnsemblBacetria. Using the ExPASy ProtParam tool, parameters for both genes with an added his-tag added were identified. The ClonTech Primer Design tool was used to design primers for PCR containing restriction sites for HindIII and NdeI for gene insertion in a pET28a plasmid vector. Once designed the primers were ordered from Eurofins and can be viewed in table 2.

Gene	Primer	Sequence
M. tuberculosis GlgB	Forward	CGCGCGGCAGCCATAATGAGTCGATCCGAGAAAC
	Reverse	GTGCGGCCGCAAGCTCTAGGCGGGCGTCAGCCA
M. smegmatis GlgB	Forward	CGCGCGGCAGCCATAATGACGAGAAGCAGCAATC
	Reverse	GTGCGGCCGCAAGCTCTAAGCCGGCTCGAACCA

Table 2. Sequence of primers for gene cloning for *M. tuberculosis GlgB* and *M. smegmatis GlgB* genes.

The full length genes of *M. tuberculosis* and *M. smegmatis* GlgB were amplified by polymerase chain reaction (PCR) from *M. tuberculosis* H37Rv and *M. smegmatis* genomic DNA (Supplied by Prof. Besra Lab, University of Birmingham). The PCR reaction mixtures consisted of: 1 µL genomic DNA, 10 µM forward primer, 10 µM reverse primer, 10 µM dNTPs, 10 µL 5x GC buffer, 2.5 µL 3% DMSO v/v, 0.5 µL 1x DNA polymerase v/v and 33 µL sterile water. A negative control was also ran which contained no genomic DNA.

PCR was carried out in a thermo cycler at the following conditions: Initial

denaturation was at 98 °C for 2 mins, denaturation was at 98 °C for 30s, annealing was at 72 °C for 30s, extension was at 72 °C for 1.5 mins and denaturation, annealing and extension was ran for 35 cycles followed by a final extension of 72 °C for 10 mins.

PCR products were then ligated into *NdeI* and *Hind*III sites of plasmid pET28a (Novagen) and plasmids from successful ligation reactions were purified (QIAprep spin miniprep kit) according to manufacturer's instructions.

3x pET28A plasmids (Novagen) containing the *M. smegmatis glgE* gene and 2x pET28A plasmids (Novagen) containing the gene for *M. tuberculosis glgB* were heat transformed into BL21(DE3) competent cells following same method as for section 2.1 and each cultured in 1 L fresh terrific broth with kanamycin (50 μgmL^{-1}) as done before and pellets resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% (v/v) glycerol) and frozen at -80 °C until required.

Each of the pellets were lysed in the lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% (v/v) glycerol) following step 2.2 and the collected clear lysate was then diluted 2-fold in buffer A (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% (v/v) glycerol) and applied to a pre-equilibriated Ni-NTA column (1 mL, GE Healthcare Life Sciences) (following section 2.3) and collected fractions analysed by SDS-PAGE.

Ion-exchange chromatography was carried out for only one of the cultures containing *M. smegmatis* GlgB which expressed well using the same protocol (section 2.4). Size exclusion chromatography followed ion-exchange also following the previous method but using a different buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% (v/v) glycerol). All fractions were assessed using SDS-PAGE.

3.0 Results

3.1 *Mycobacterium tuberculosis* GlgE purification

3.1.1 Ni-NTA chromatography

The first step of protein purification of *M. tuberculosis* GlgE was Ni-NTA chromatography and was carried out using a step gradient of 20 mM, 40 mM, 80 mM, 100 mM, 150 mM, 200 mM, 300 mM and 500 mM imidazole (5 mL per fraction). Fractions were analysed using an SDS-PAGE gel which can be viewed in figure 6. In the fractions containing 20 mM to 80 mM imidazole there are many bands present along the gel representing different sized proteins that have eluted from the column. The band at the size expected for GlgE, 80 kDa, is weak. For the fraction containing 100 mM imidazole there is a band indicative of GlgE but there is also a stronger band present indicating protein at around 25 kDa. Fractions containing 150 to 500 mM imidazole had strong bands at the expected size of 80 kDa and appear quite pure apart from the presence of some much weaker bands present at smaller molecular weights. These were the fractions that were pooled for ion-exchange chromatography.

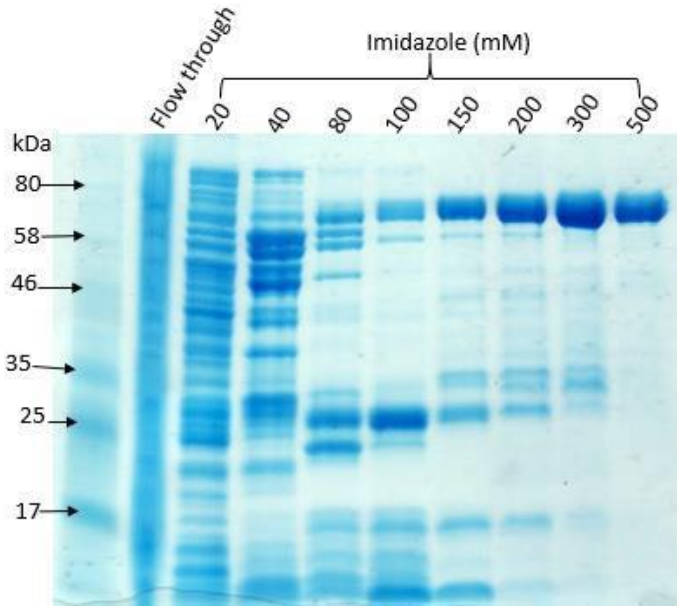


Figure 6. An SDS-PAGE analysis of step gradient imidazole fractions (5 mL per fraction) containing *M. tuberculosis* GlgE after Ni-NTA chromatography. Molecular weight standards are given in units of kDa and the first lane indicates the protein marker. The next lane is the flow through followed by the step gradient concentrations of 20, 40, 80, 100, 150, 200, 300 and 500 mM imidazole.

3.1.2 Ion exchange chromatography

Ion exchange chromatography of *M. tuberculosis* GlgE was carried out using a step gradient of NaCl in steps of 100 mM, 150 mM, 200 mM, 250 mM, 300 mM and 500 mM (5 mL per fraction). Collected fractions were analysed by SDS-PAGE (figure 7). It can be seen that anion exchange chromatography was unable to bind all of the GlgE to the column as there is still a band present in the flow through at 80 kDa (indicated by an arrow). The fractions containing 100 to 200 mM NaCl show a band for GlgE at 80 kDa which is relatively pure. Most GlgE protein eluted from the column at 250 mM NaCl. At this fraction there is also another band directly under the protein band which can also be seen in the following fractions of 300 and 500 mM NaCl. Fractions 100 to 300 mM NaCl were pooled and concentrated for size exclusion chromatography.

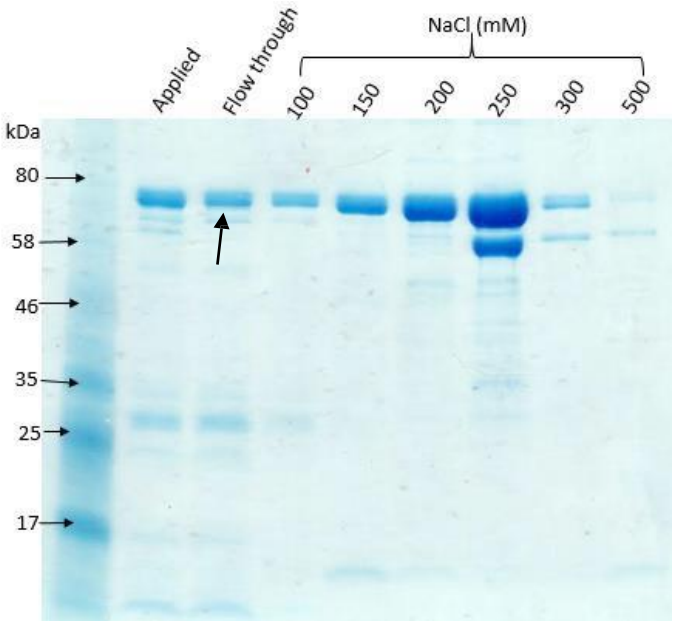


Figure 7. An SDS-PAGE analysis of step gradient imidazole fractions (5mL per fraction) containing *M. tuberculosis* GlgE after anion exchange chromatography. Molecular weight standards are given in units of kDa and the first lane indicates the protein marker. The second lane is the applied followed by the flow through. Following this are step gradient fractions containing 100, 150, 200, 250, 300 and 500 mM NaCl. The black arrow indicates GlgE present in the flow through.

3.1.3 Size exclusion chromatography

In figure 8 is the graph of the elution volume vs. normalised absorbance at A_{280} along with the corresponding SDS-PAGE which confirmed where GlgE eluted from the column. The SDS-PAGE indicates that the GlgE protein which eluted from the column is of high purity and eluted between 155 and 210 mL with the strongest band at 180 mL. These were the fractions pooled and concentrated in preparation for crystallisation attempts. The single peak on the absorbance graph indicates the protein is a monomer in solution. *M. tuberculosis* GlgE was concentrated to 6.20 mg/mL.

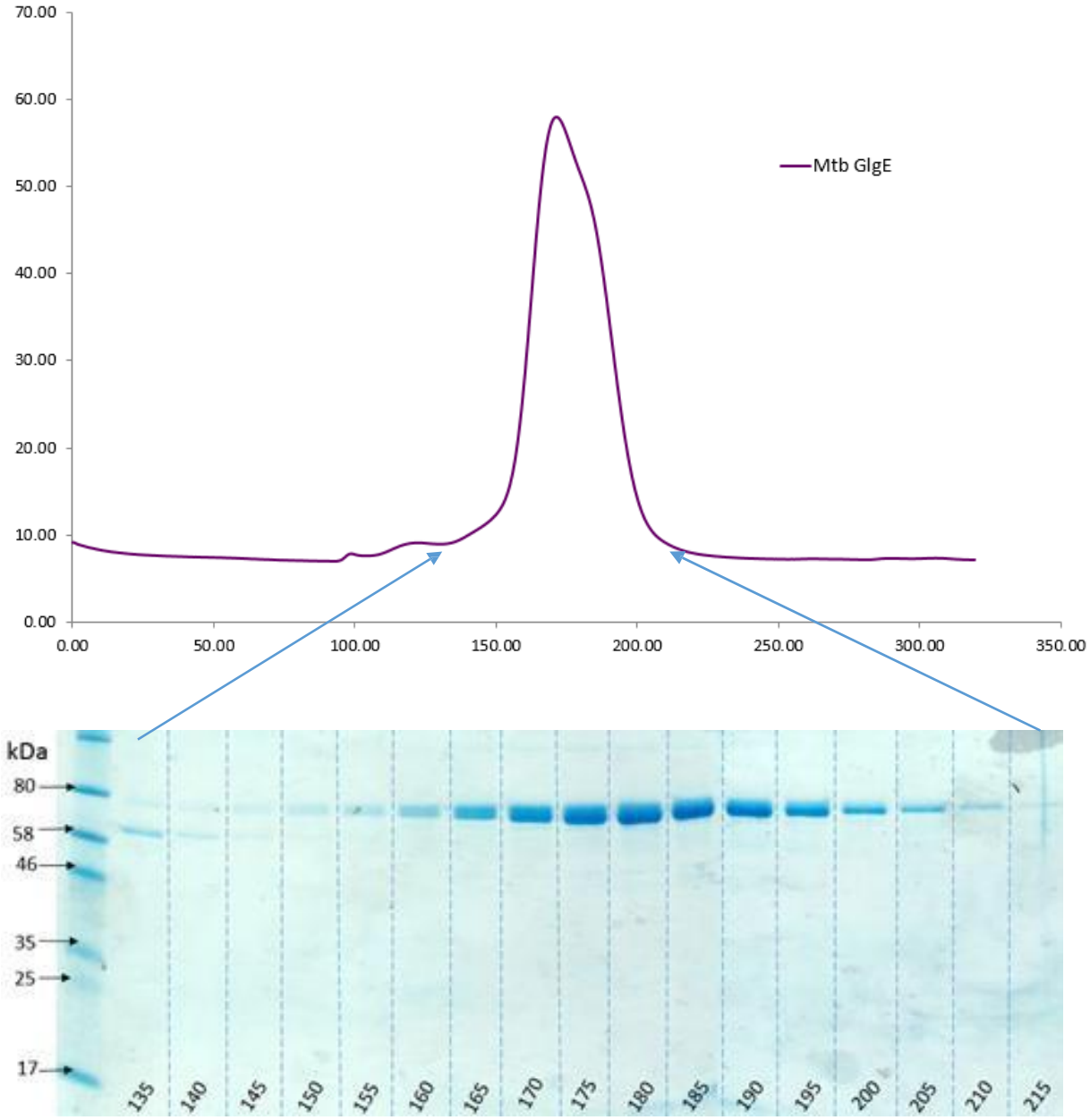


Figure 8. Elution volume (mL) vs. normalised absorbance A_{280} of *M. tuberculosis* GlgE elution from a gel filtration column and confirmed by SDS-PAGE analysis of fractions (5 mL per fraction). Molecular weight standards are given in units of kDa and the first lane is the protein marker.

3.1.4 Crystal trays

Crystallisation trials used vapour diffusion and commercially available sparse matrix screens to identify initial hits. Small integrated crystals grew in JCSG plus at the conditions of 0.5 M NaCl, 0.1 M Na citrate, 4% (v/v) polyethyleneimine, pH 5.6 and can be viewed in figure 9.



Figure 9. Small integrated crystals of *M. tuberculosis* GlgE grown in the matrix screen for JCSG plus at the conditions of 0.5 M NaCl, 0.1 M Na citrate, 4% (v/v) polyethyleneimine, pH 5.6.

3.2 *Mycobacterium smegmatis* GlgE purification

3.2.1 Ni-NTA chromatography

5 L cultures of two different *E. coli* competent cells: BL21(DE3) and BL21(DE3)pLySs) were grown to express *M. smegmatis* GlgE and Ni-NTA chromatography was used to assess which cell type best expressed GlgE. Ni-NTA chromatography was carried out using the cell extract from both cell types using a step gradient of imidazole in the same manner as in part 3.1.1. Figure 10 shows the SDS-PAGE analysis carried out using samples from each of the 5 mL step

gradient imidazole fractions following Ni-NTA chromatography for BL21(DE3) cell extract. Figure 11 shows the same but for the BL21(DE3)pLySs cell extract.

Looking at figure 10, most GlgE (79 kDa) can be seen to have eluted from the column in the fraction containing 80 mM imidazole and continued to elute from the column for the remaining increasing imidazole concentrations. Fractions containing 20 and 40 mM imidazole showed weak bands at the size of GlgE but there are also a lot of other bands in each fraction at a range of sizes indicating nonspecific proteins. Fractions containing 80 to 200 mM imidazole saw that some of the smaller protein bands were just as strong as the expected band for GlgE such as at around 25 kDa and below.

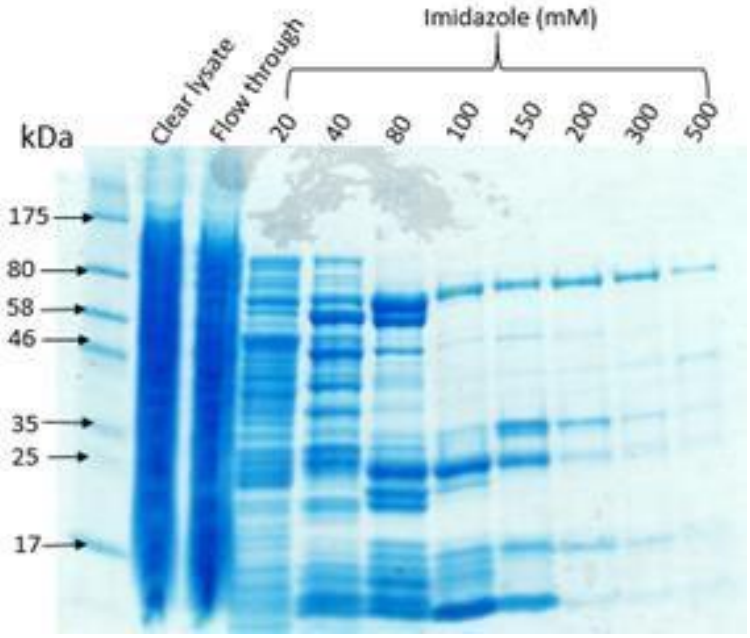


Figure 10. An SDS-PAGE analysis of step gradient imidazole fractions (5 mL per fraction) containing *M. smegmatis* GlgE after Ni-NTA chromatography of cell extract from a 5 L cell culture using BL21(DE3) cells. Molecular weight standards are given in units of kDa and the first lane is the protein marker. Next along is the clear lysate then the flow through followed by the step gradient concentrations of 20, 40, 80, 100, 150, 200, 300 and 500 mM imidazole.

After purifying the protein from the BL21(DE3)pLySs cell extract using Ni-NTA chromatography, it can be seen from the SDS-PAGE analysis (figure 11) that after eluting using an imidazole concentration gradient the protein elutes across all of the concentrations but the stongest band was at 80 mM imidazole. At this concentration however, bands were present at smaller protein sizes that were stronger than for the expected size of 79 kDa. Fractions containing 100 to 300 mM imidazole had weak bands around 79 kDa which decreased in size with increasing imidazole concentrations and finally no protein was eluted at 500 mM imidazole. Across these imidazole concentrations there are also strong bands at around 35 kDa and below which were also present in figure 5 but in this SDS-PAGE gel the bands which indicate impurities are stronger.

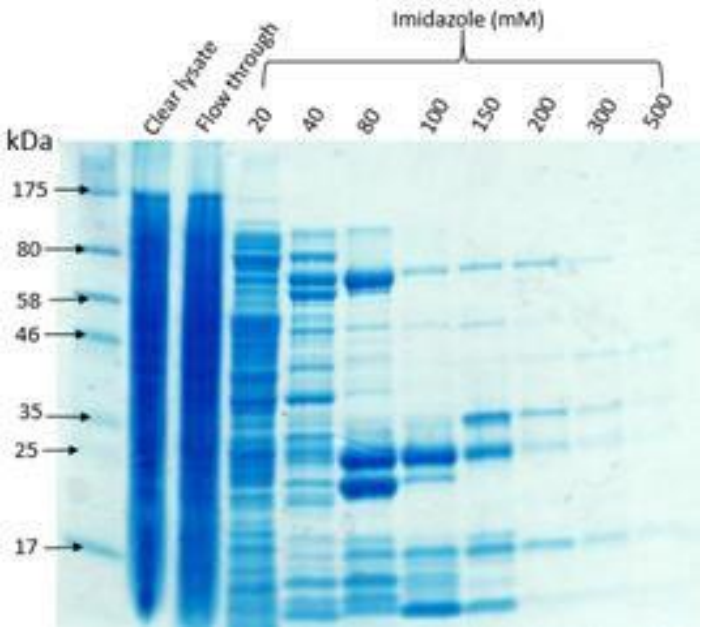


Figure 11. An SDS-PAGE analysis of step gradient imidazole fractions (5 mL per fraction) containing *M. smegmatis* GlgE after Ni-NTA chromatography of cell extract from a 5 L cell culture using BL21(DE3)pLySs cells. Molecular weight standards are given in units of kDa and the first lane is the protein marker. Next along is the clear lysate then the flow through followed by the step gradient concentrations of 20, 40, 80, 100, 150, 200, 300 and 500 mM imidazole.

Protein was purified from a cell extract of a 10 L culture of BL21(DE3) cells using Ni- NTA chromatography. Figure 12 shows the SDS-PAGE analysis of the imidazole step gradient fractions collected and found that in comparison to figure 10 which used a 5 L culture of the same cells, GlgE elutes at the same imidazole concentrations but in figure 12 the bands are stronger indicating more GlgE is present. There are also the same bands that were seen previously at around 35 kDa and below but in figure 12 the bands are also stronger. Fractions containing 80 to 500 mM imidazole were pooled for ion exchange chromatography.

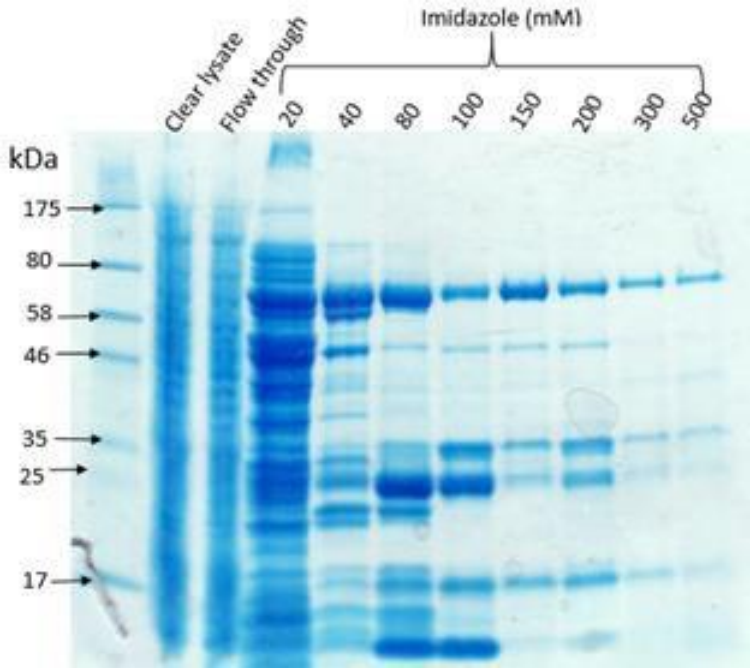


Figure 12. An SDS-PAGE analysis of step gradient imidazole fractions (5 mL per fraction) containing *M. smegmatis* GlgE after Ni-NTA chromatography of cell extract from a 10 L cell culture using BL21(DE3) cells. Molecular weight standards are given in units of kDa and the first lane is the protein marker. Next along is the clear lysate then the flow through followed by the step gradient concentrations of 20, 40, 80, 100, 150, 200, 300 and 500 mM imidazole.

3.2.2 Ion exchange chromatography

Following Ni-NTA chromatography, ion exchange chromatography was carried out on the previously pooled fractions and a step gradient of NaCl was used to elute proteins from the column. Fractions were analysed using SDS-PAGE and the resultant gel can be viewed in figure 13. In the flow through there is a visible band at 79 kDa indicating that not all GlgE was bound to the column resulting in protein loss (black arrow). Fractions containing 100, 150 and 200 mM NaCl eluted most GlgE from the column as strong bands are present. Just below 79 kDa for the fractions containing 100 and 150 mM NaCl there is also another strong band present. There are also strong bands present indicating impure proteins of smaller sizes of 35 kDa and below. There are also bands present above 79 kDa which were not present in Ni-NTA chromatography.

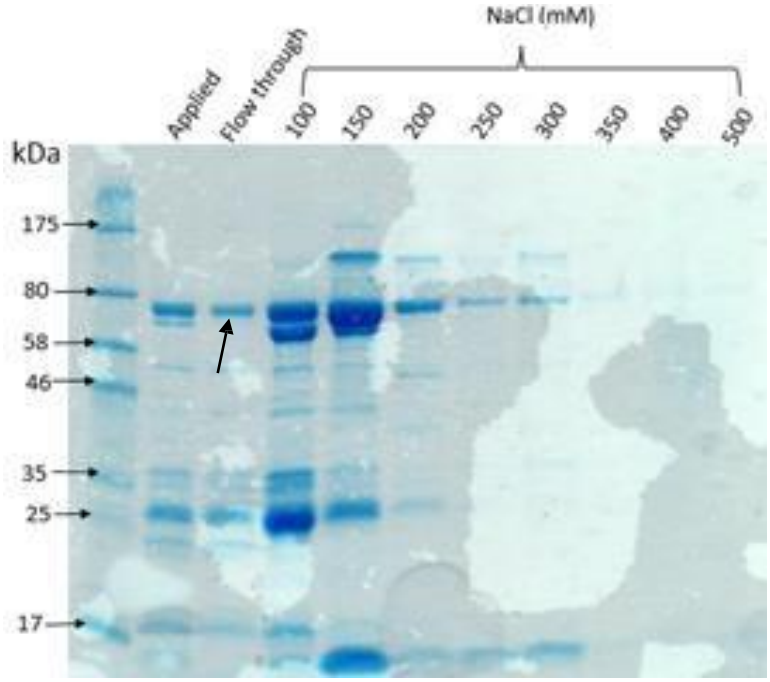


Figure 13. An SDS-PAGE analysis of step gradient NaCl fractions (5 mL per fraction) containing *M. smegmatis* GlgE after anion exchange chromatography. Molecular weight standards are given in units of kDa and the first lane is the protein marker. Next along is the applied then the flow through followed by the step gradient concentrations of 100, 150, 200, 250, 300, 350, 400 and 500 mM NaCl. An arrow shows that GlgE is still present in the flow through.

The flow through that was determined to still have GlgE present in figure 13 was run again over the anion exchange column and fractions containing the same NaCl step gradient were used to elute protein from the column. Collected fractions were analysed again by SDS-PAGE (figure 14). It can be seen that again not all GlgE bound the column as there is still a band at 79 kDa in the flow through. Protein was only eluted from the column when the fractions containing 100 and 150 mM NaCl were applied. There are strong bands at the expected 79 kDa but there are also strong bands at aroung 35 kDa and below.

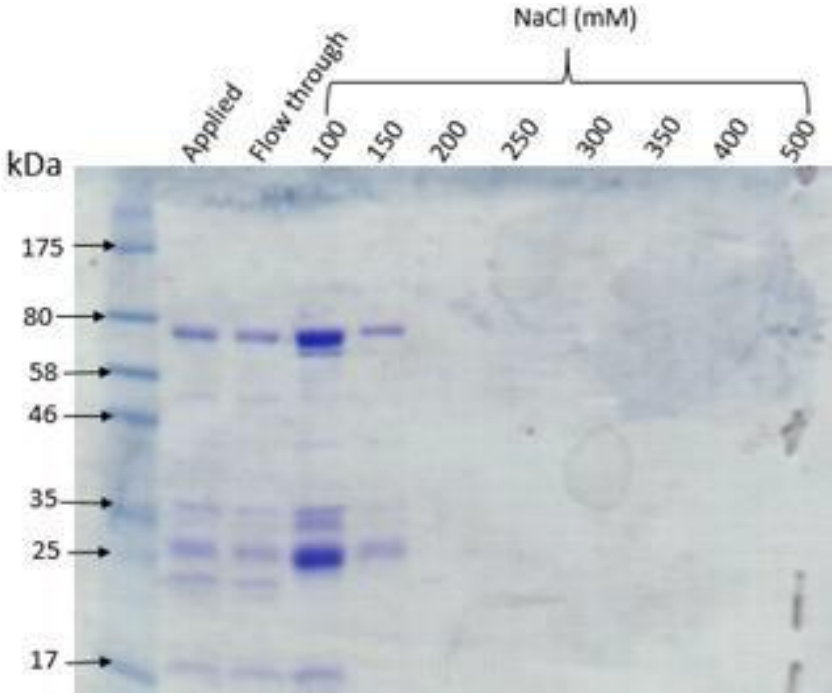


Figure 14. An SDS-PAGE analysis of step gradient NaCl fractions after anion exchange chromatography of the flow through from figure 13 which contained *M. smegmatis* GlgE. Molecular weight standards are given in units of kDa and the first lane is the protein marker. Next along is the applied and flow through followed by the step gradient concentrations of 100, 150, 200, 250, 300, 350, 400 and 500 mM NaCl.

Using the flow through which was analysed in figure 14, cation chromatography was carried out and a step gradient of NaCl was used to elute protein from the column. Fractions were analysed by SDS-PAGE and the results can be viewed in figure 15. There is still a weak band at 79 kDa present in the flow through. Bands are weaker than in the previous ion exchange chromatography runs but GlgE can be seen to elute here across all fractions at 79 kDa. Proteins of smaller sizes also elute across all of the fractions with some bands stronger than those at the correct size for GlgE.

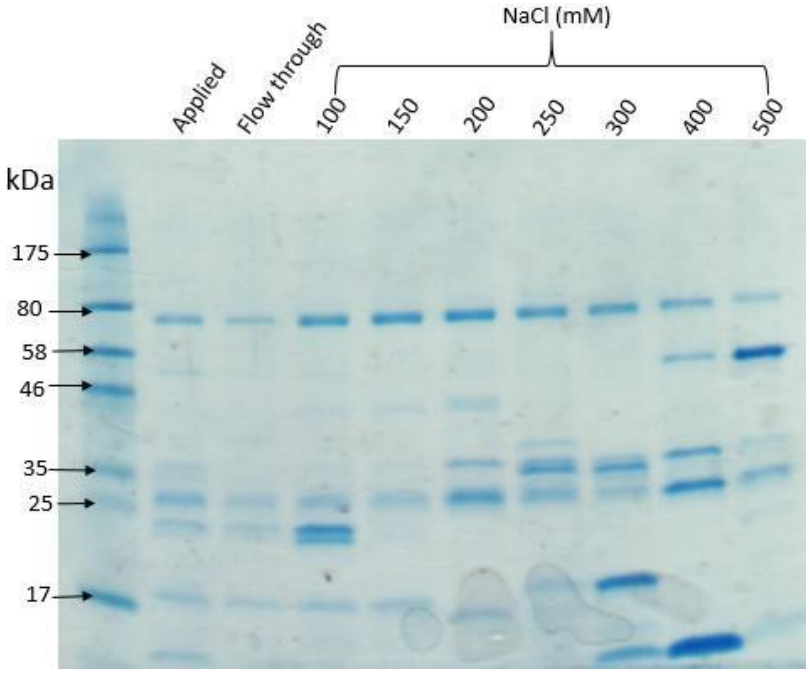


Figure 15. An SDS-PAGE analysis of step gradient NaCl fractions (5 mL per fraction) containing *M. smegmatis* GlgE after cation exchange chromatography. Molecular weight standards are given in units of kDa and the first lane is the protein marker. Next along is the applied then the flow through followed by the step gradient concentrations of 100, 150, 200, 250, 300, 400 and 500 mM NaCl.

3.2.3 Size exclusion chromatography

The eluted fractions containing 100, 150, 200, 250 and 300 mM NaCl from figure 13, the fractions containing 100, 150 and 200 mM NaCl from figure 14 and all of the fractions from figure 15 were pooled and concentrated to be used in size exclusion chromatography. Figure 16 shows the elution volumes at which GlgE eluted from the column and the absorbance at 280 nm which was confirmed by an SDS-PAGE gel. GlgE could be seen to elute from the column between 120 and 190 mL with most eluting at ~168 mL. Looking at the SDS-PAGE gel bands of smaller sized proteins between 25-35 kDa were separated out and mostly eluted between 185 and 205 mL although there are faint bands present between 140 and 180 mL. Between 165 and 180 mL there are quite strong bands indicating

proteins which are around 17 kDa. The single peak on the graph indicates the protein is a monomer in solution. *M. smegmatis* GlgE was concentrated to 9.67 mg/mL.

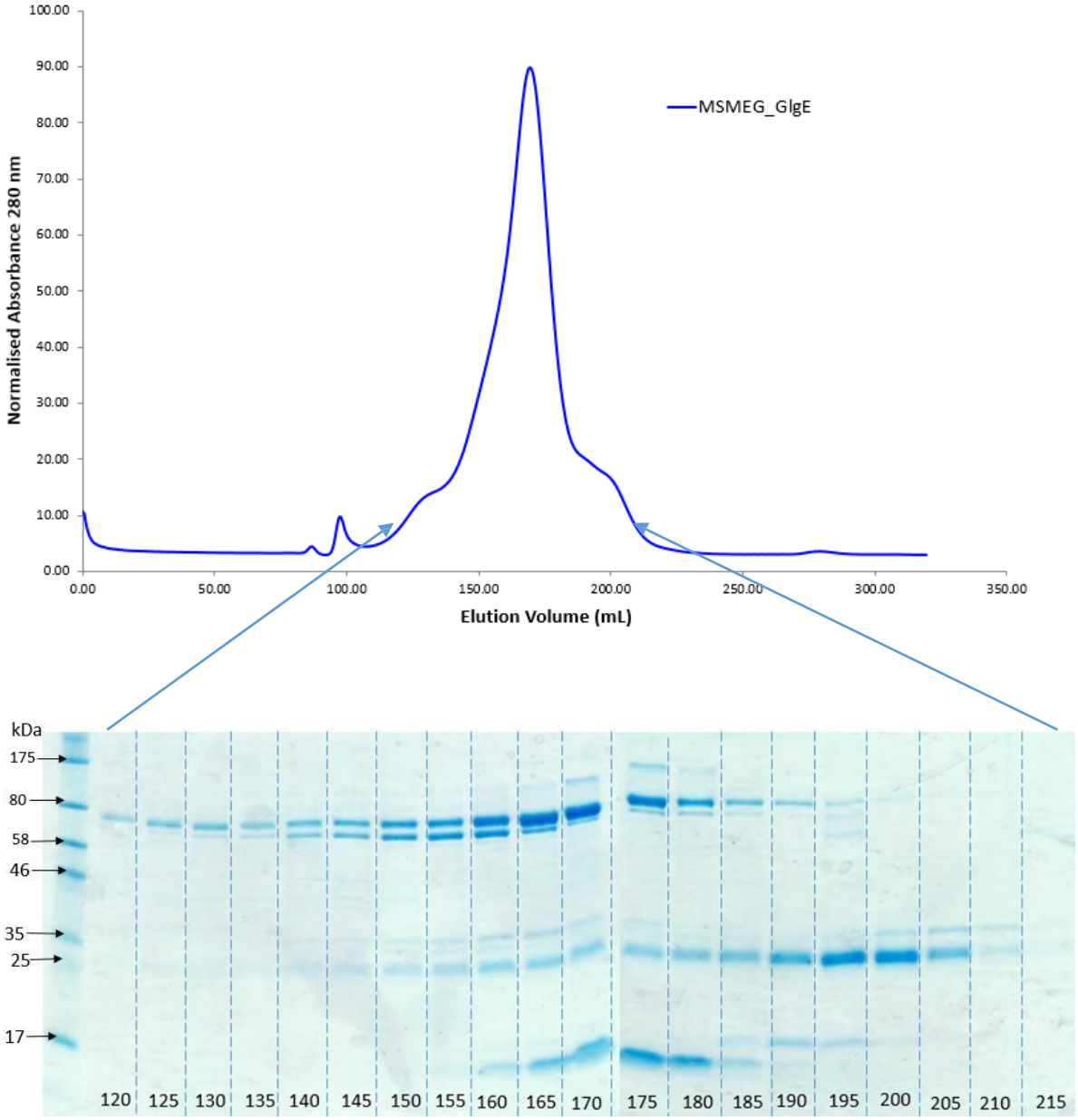


Figure 16. Elution volume (mL) vs. normalised absorbance A_{280} of *M. smegmatis* GlgE elution from a gel filtration column and was confirmed by SDS-PAGE analysis of eluted fractions (5 mL per fraction). Molecular weight standards are given in units of kDa and the first lane is the protein marker.

3.2.4 Western blot

Figure 17 displays the result following a western blot analysis which was carried out using an SDS-PAGE gel of the step gradients of 20, 40, 80, 100, 150, 200, and 300 mM imidazole following Ni-NTA chromatography. Anti-6xHis antibodies were used to detect the 6xHis tag on recombinant GlgE. Strong bands can see across all fractions at the expected size of 79 kDa. There are no other clear bands present on this western blot at other protein sizes.

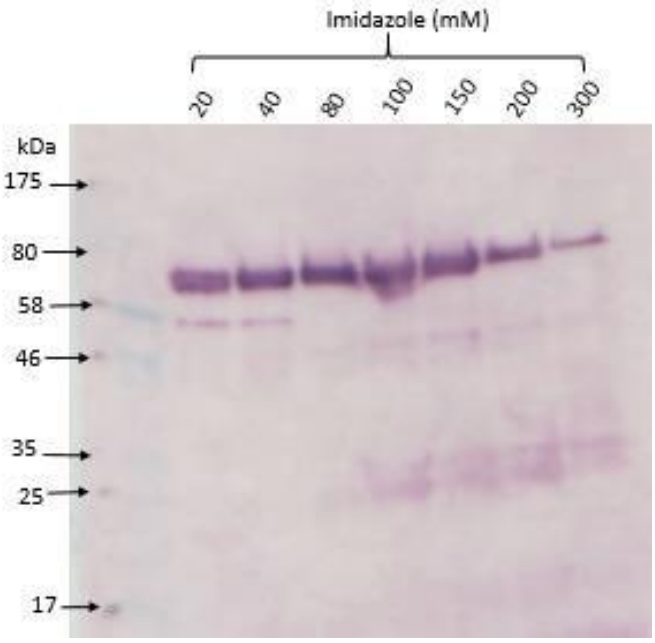


Figure 17. Western blot analysis of the step gradient imidazole fractions containing *M. smegmatis* GlgE after Ni-NTA chromatography using anti-6xHis antibodies to detect the 6xHis tag on recombinant GlgE. Molecular weight standards are given in units of kDa and the first lane is the protein marker followed by the step gradient concentrations of 20, 40, 80, 100, 150, 200, 300 and 500 mM imidazole.

3.2.5 Plasmid screen

Figure 18 shows three SDS-PAGE analyses (A, B and C) of the step gradient concentrations of imidazole used to elute protein following Ni-NTA chromatography after three plasmids from a previously carried out ligation reaction containing the *M. smegmatis* glgE gene were transformed into BL21(DE3) cells and cultured in 1 L media. From studying each SDS-PAGE it is clear that none of these plasmids expresses GlgE any better than the plasmid currently used as bands at the expected size for GlgE were not as strong and there was still a lot of nonspecific proteins present.

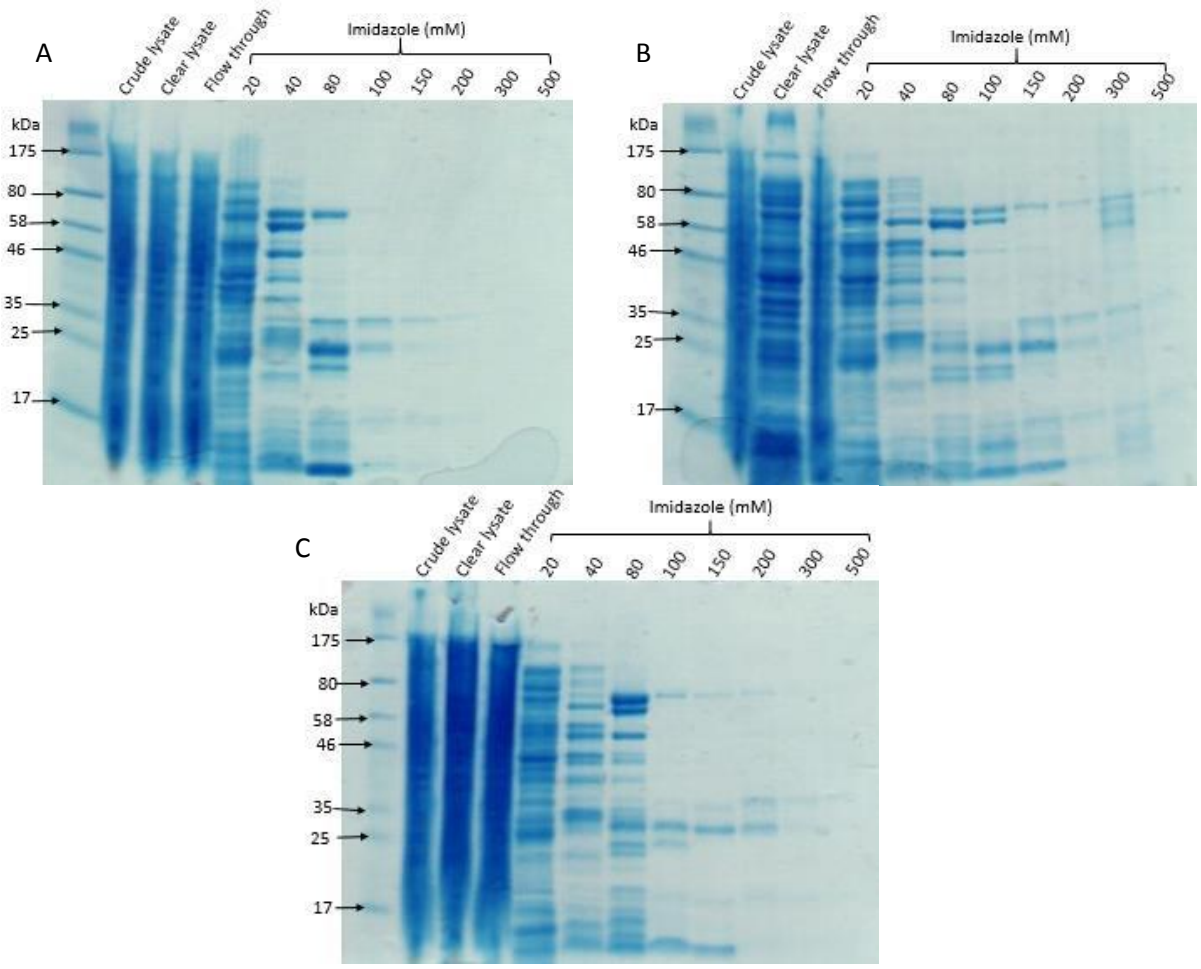


Figure 18. Three plasmids were used from a previous ligation reaction that contained the *M. smegmatis glgE* gene insert and were transformed into BL21(DE3) cells and each cultured in 1 L media. Ni-NTA chromatography was carried out for each and the collected step gradient imidazole fractions (5 mL per fraction) were analysed by SDS-PAGE (A, B and C). For each gel the first lane contains the protein marker and molecular weight standards are given in kDa. The next lanes contain the crude lysate, the clear lysate and then the flow through. Following these are step gradient concentrations of 20, 40, 80, 100, 150, 200, 300 and 500 mM imidazole.

Figure 19 shows a thermal stability assay that was carried out for *M. smegmatis* GlgE in different buffers at a range of pH's and at high and low salt concentrations (100 mM and 250 mM) by another student within the lab. SYPRO orange dye which is present in the assay solution with GlgE and the buffers gives increased fluorescent signal when the protein denatures with increasing temperature meaning hydrophobic amino acids on GlgE are exposed. A higher melting temperature corresponds to increased protein stability. At pH 7.0 and 250 mM NaCl using the HEPES buffer the thermal stability of the protein is highest.

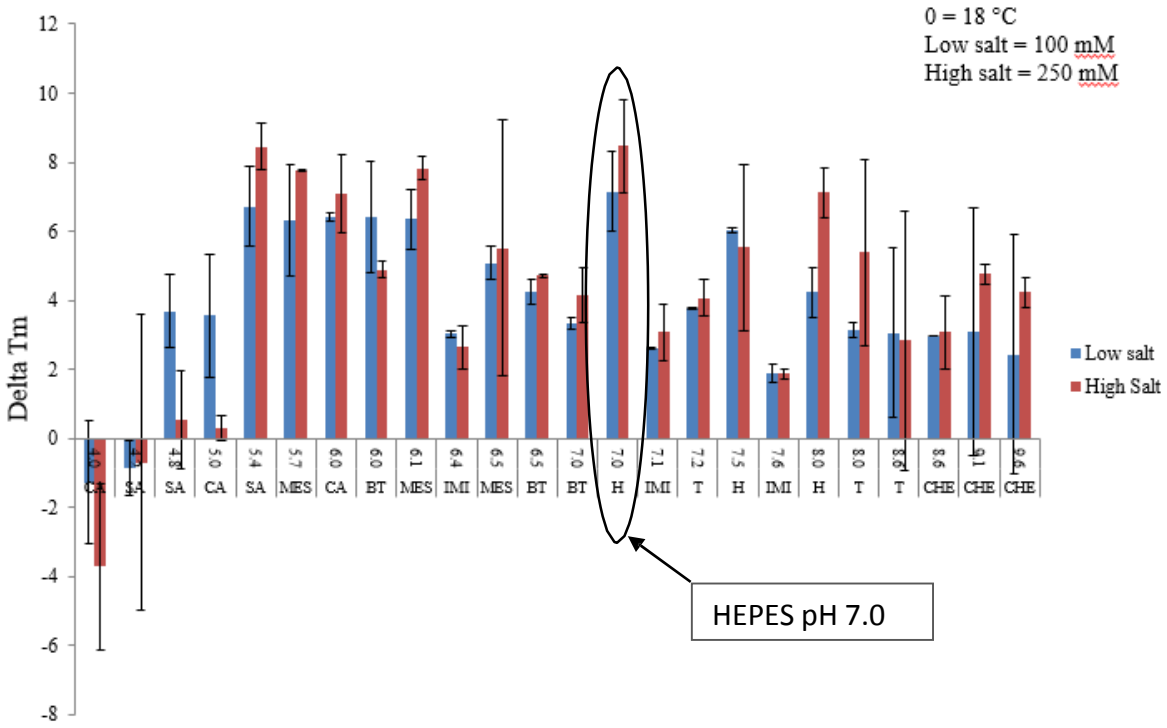


Figure 19. Thermal stability assay results of *M. smegmatis* GlgE. The assay was carried out in a range of buffers (all 150 mM) at increasing pH and at low (100 mM) and high (250 mM) NaCl concentrations. Displayed is a bar chart showing the different melting temperatures (T_m) for GlgE in all of the buffers tested (Results supplied by Ali Kermani, School of Biosciences).

Considering the result from figure 19, HEPES buffer at pH 7.0 with 250 mM NaCl was tested as the buffer for Ni-NTA chromatography in order to try and improve purification. Figure 20 shows three more SDS-PAGE analyses (A, B and C) carried out using step gradients of imidazole used to elute protein following Ni-NTA chromatography after three plasmids from a previous ligation reaction containing the *M. smegmatis* glgE gene were each transformed into BL21(DE3) cells and cultured in 1 L media. During each Ni-NTA chromatography run, buffer A was substituted for a buffer containing HEPES pH 7.0, 250 mM NaCl to test if this different buffer had a pronounced positive effect on purification. From studying the three gels it is clear that there is no clear improvement on purification. Bands indicated GlgE at around 80 kDa are present but there are also a lot of nonspecific proteins still and expression appears to not be as good as the plasmid used in the original experiments.

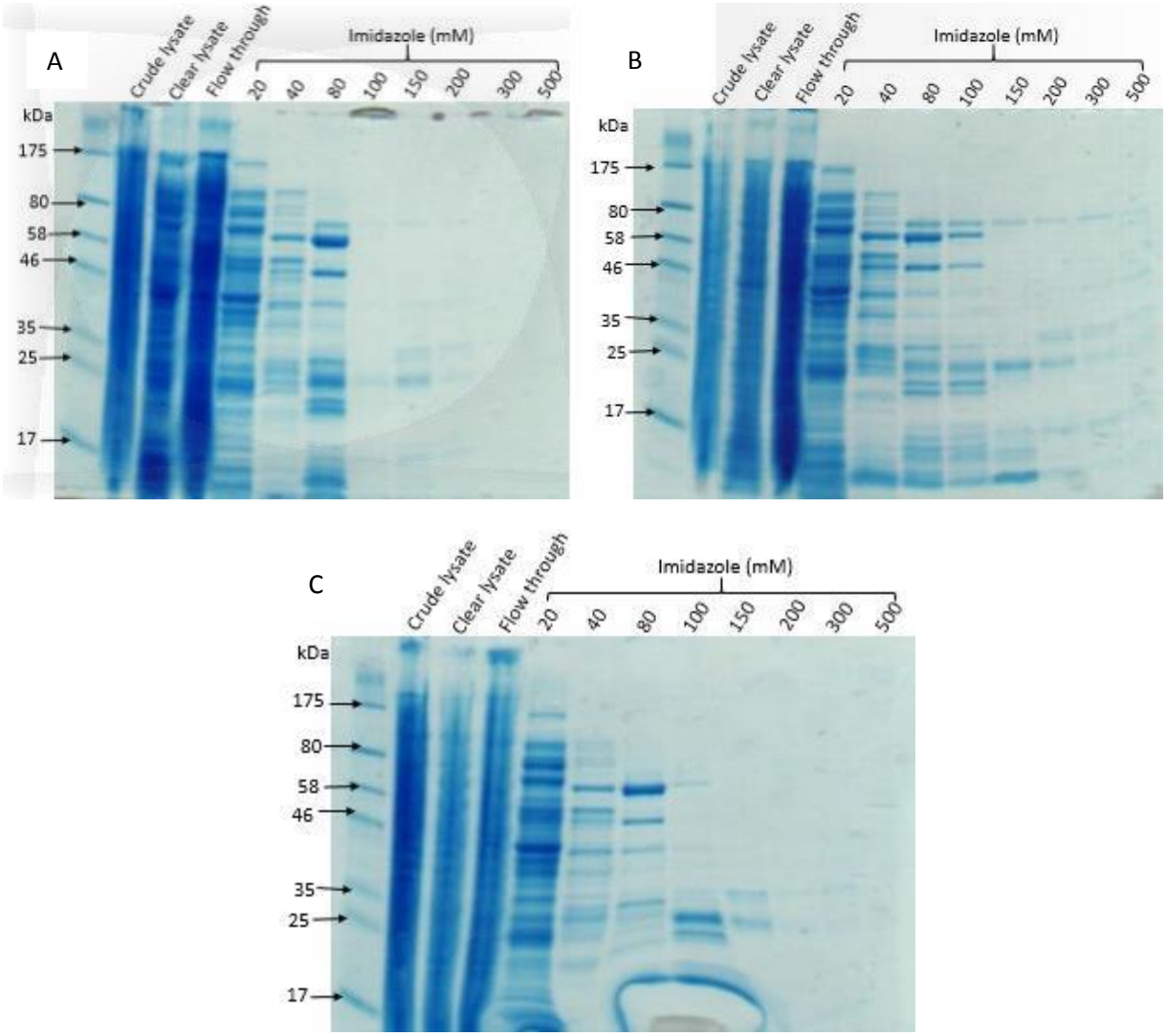


Figure 20. Three plasmids were used from a previous ligation reaction that contained the *M. smegmatis glgE* gene insert and were transformed into BL21(DE3) cells and cultured to 1 L. Ni-NTA chromatography was carried out for each plasmid (A, B and C) and the resultant step gradient imidazole fractions (5 mL per fraction) were analysed by SDS-PAGE. For each gel the first lane contains the protein marker and molecular weight standards are given in kDa. The next lanes contain the crude lysate, the clear lysate and then the flow through. Following these are step gradient concentrations of 20, 40, 80, 100, 150, 200, 300 and 500 mM imidazole.

3.2.6 *Mycobacterium smegmatis* GlgE activity assay

Figure 21 shows the activity assay carried out for *M. smegmatis* GlgE. The control of maltose-1-phosphate on its own showed no activity while the control using just GlgE showed an unexpected small increase in absorbance over time. Using 10

µL GlgE increased absorbance over time and 20 µL gave an even larger increase in absorbance. Adding 30 µL GlgE further increased absorbance but the increase was not as much as between 10 and 20 µL. *M. tuberculosis* GlgE activity was also confirmed within the lab (data not shown).

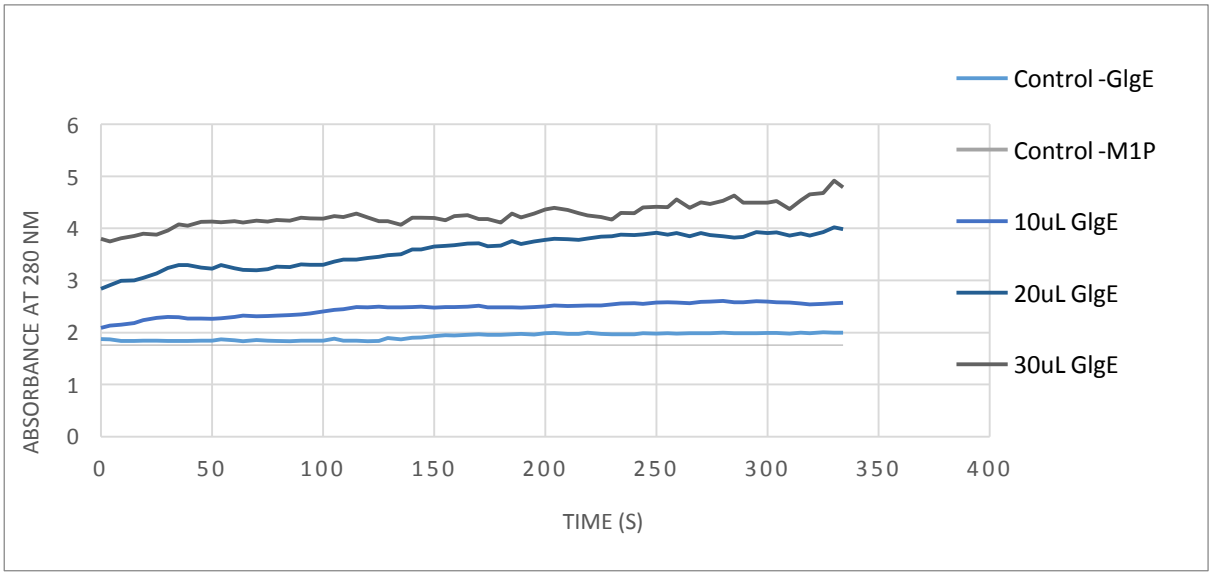


Figure 21. Activity assay for *M. smegmatis* GlgE using increasing amounts of GlgE (10, 20 and 30 µL) and two controls; one without GlgE and another without the substrate maltose-1-phosphate. Results are displayed over time vs. absorbance at 280 nm.

3.2.7 Size exclusion chromatography to investigate complex formation between *Mycobacterium smegmatis* GlgE and Pep2

Figure 22 was provided by another student within the lab and shows an analysis after size exclusion chromatography of GlgE at pH 7.2 which can be seen to elute at around 170 mL as a monomer in solution (purple line). When GlgE is ran through the gel filtration column at pH 6.0 along with Pep2 (blue line), a peak for Pep2 can be seen at ~190mL. At around 140 mL there is a small peak and no peak where GlgE was expected to elute at ~175 mL indicating GlgE may elute in solution at pH 6 as a higher oligomer.

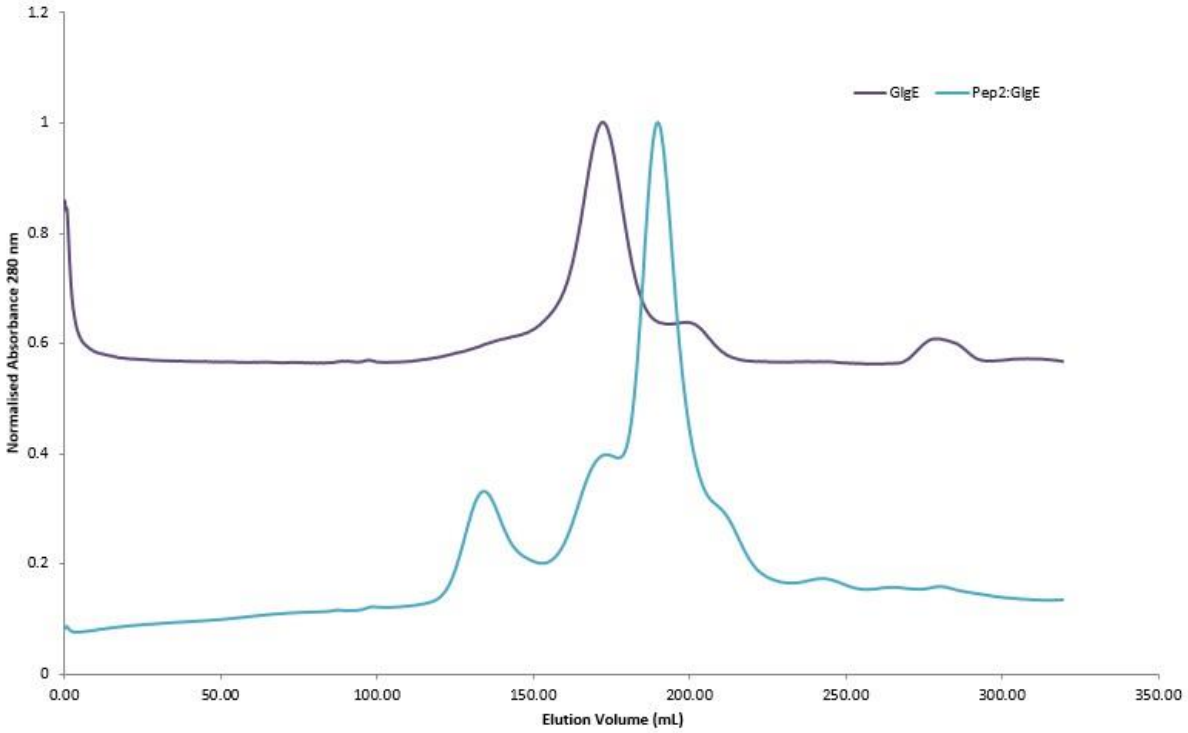


Figure 22. Elution volume (mL) vs. normalised absorbance A_{280} of *M. smegmatis* GlgE elution from a gel filtration column at pH 7.2 (purple line) and elution volume (mL) vs. normalised absorbance A_{280} of *M. smegmatis* GlgE and Pep2 at pH 6.0 (blue line). Data provided by Ali Kermani (School of Biosciences).

3.2.8 Size exclusion chromatography of GlgE at pH 7.0

Pooled fractions from ion exchange chromatography were dialysed into a buffer of pH 7.0 and size exclusion chromatography was carried out. The resultant fractions were analysed by SDS-PAGE and the gel can be viewed in figure 23. Very faint bands can be seen at the elution volumes 165 to 192 mL at 79 kDa and there are no other bands present which would suggest the protein is pure. Due to mechanical failure where the pump of the AKTA stopped functioning correctly, a new machine was located but lacked equipment to measure absorbance therefore absorbance readings are unfortunately not present in this section.

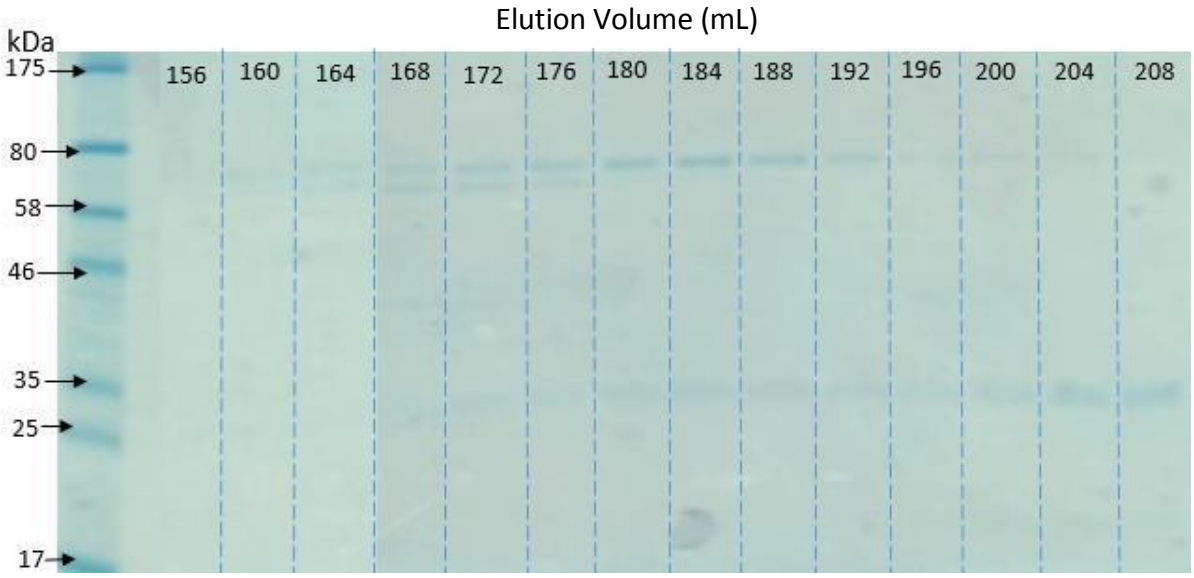


Figure 23. An SDS-PAGE analysis of fractions (5 mL per fraction) containing *M. smegmatis* GlgE elution in a buffer of pH 7.0. The first lane represents the protein marker and all molecular weight standards are given in kDa following by elution fractions from 156 mL to 208 mL.

3.2.9 Size exclusion chromatography of GlgE at pH 6.0

Size exclusion was also carried out in the same way as in 3.2.8 only this time a buffer of pH 6.0 was used instead. Fractions of the volumes in which there had shown to be possible formation of higher oligomers in part 3.2.7 were analysed by SDS-PAGE and can be viewed in figure 24, A. For gel A, 116 to 156 mL, no bands can be seen across any of the fractions. Unexpectedly, faint bands are present for all of the fractions on gel B from 204 to 244 mL. Due to mechanical failure absorbance could not be read during experimentation.

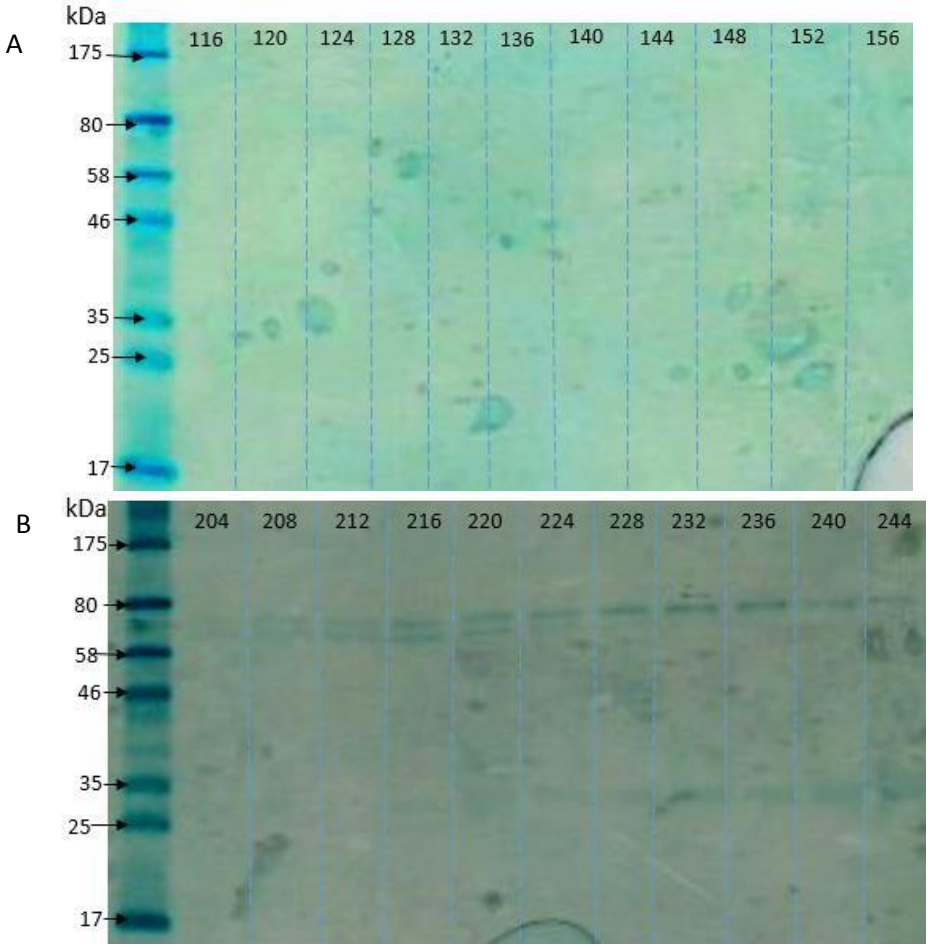


Figure 24. Two SDS-PAGE analyses of fractions (5 mL per fraction) containing *M. smegmatis* GlgE elution in a buffer of pH 6.0. The first lane on both gels represents the protein marker and all molecular weight standards are given in kDa following by A) elution fractions from 116 mL to 156 mL and for B) 204 mL to 244 mL.

3.3 Cloning and expression of *M. tuberculosis* and *M. smegmatis* GlgB

3.3.1 PCR

Figure 25 shows the outcome of the PCR reactions carried out to amplify the genes for

M. tuberculosis glgB (A) and *M. smegmatis glgB* (B). It can be seen that PCR was successful for both reactions as bands are present at the expected sizes of 2.196 kb for *M. tuberculosis* GlgB and 2.211 kb for *M. smegmatis* GlgB. The

negative control (lane 1 for both A and B) produced no bands as expected.

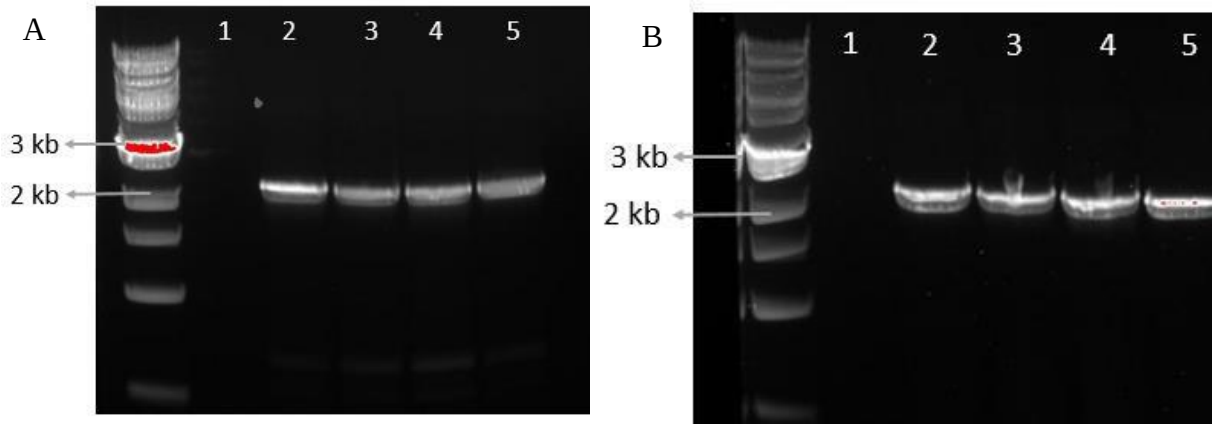


Figure 25. Gel electrophoresis following PCR to amplify A) *M. tuberculosis glgB* gene and B) *M. smegmatis glgB* gene. After the 1 kb DNA ladder, lane 1 in both gels is a negative control containing no genomic DNA followed by four lanes of amplified PCR products.

3.3.2 Ligation

To confirm the ligation reaction was successful. Plasmids were purified and then PCR was carried out for each and the resultant gel can be viewed in figure 26. The first 6 lanes indicate PCR reactions to detect plasmids containing the *M. tuberculosis glgB* gene insert and lanes 7-12 show PCR reactions for detection of plasmids containing the *M. smemgmatis glgB* gene insert. Only lane 2 produced no fragment indicating ligation for that plasmid was unsuccessful yet the rest of the lanes produced a fragment after PCR showing that those plasmids contained the correct inserts of the expected sizes.

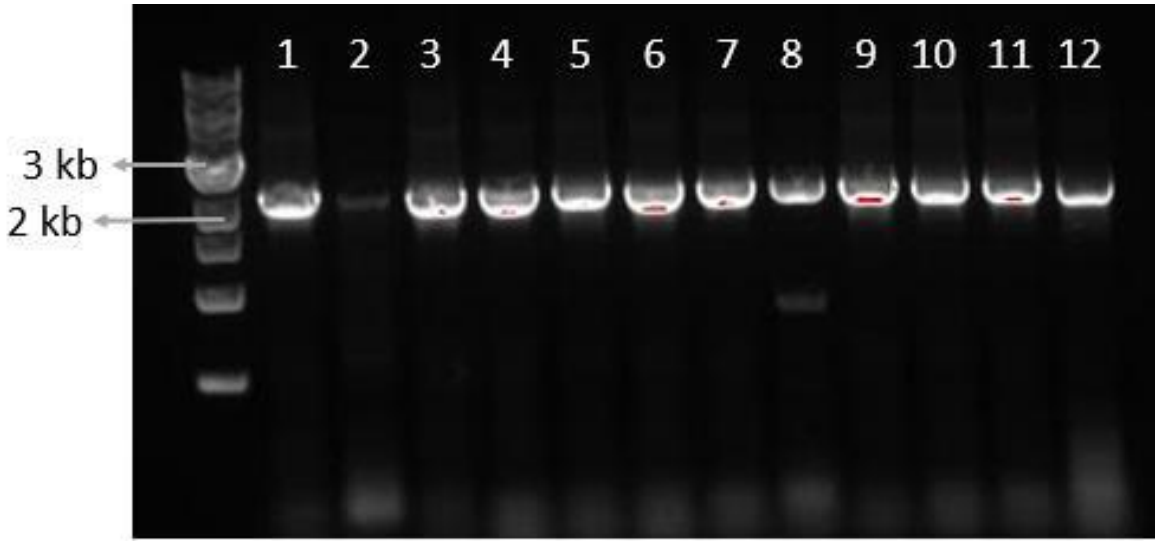


Figure 26. An electrophoresis gel after PCR to detect which plasmids contained the gene insert following a ligation reaction. After the 1 kb DNA ladder, lanes 1-6 represent the PCR amplification for detection of plasmids containing the *M. tuberculosis glgB* insert and lanes 7-12 show PCR amplification for detection of plasmids containing the *M. smegmatis glgB* insert.

3.3.3 Ni-NTA chromatography

Mycobacterium tuberculosis GlgB

Two of the resultant recombinant plasmids from the ligation reaction (Figure 26, lanes 3 and 4) of the *M. tuberculosis glgB* gene were transformed and expressed in BL21(DE3) cells. Ni-NTA chromatography was carried out using cell extracts and a step gradient of imidazole in steps of 20, 40, 80, 100, 150, 200, 300 and 500 mM (5 mL per fraction) were used to elute protein from the column. Results were analysed by SDS-PAGE and the gels (A and B) can be viewed in figure 27. There are bands present at the expected protein size of 81 kDa indicating GlgB was expressed and mostly eluted in the fractions containing 40 and 80 mM imidazole although bands are still relatively weak indicating the protein was not expressed in abundance. There are also many other bands present at smaller and larger protein sizes indicating the presence of nonspecific proteins in both gels.

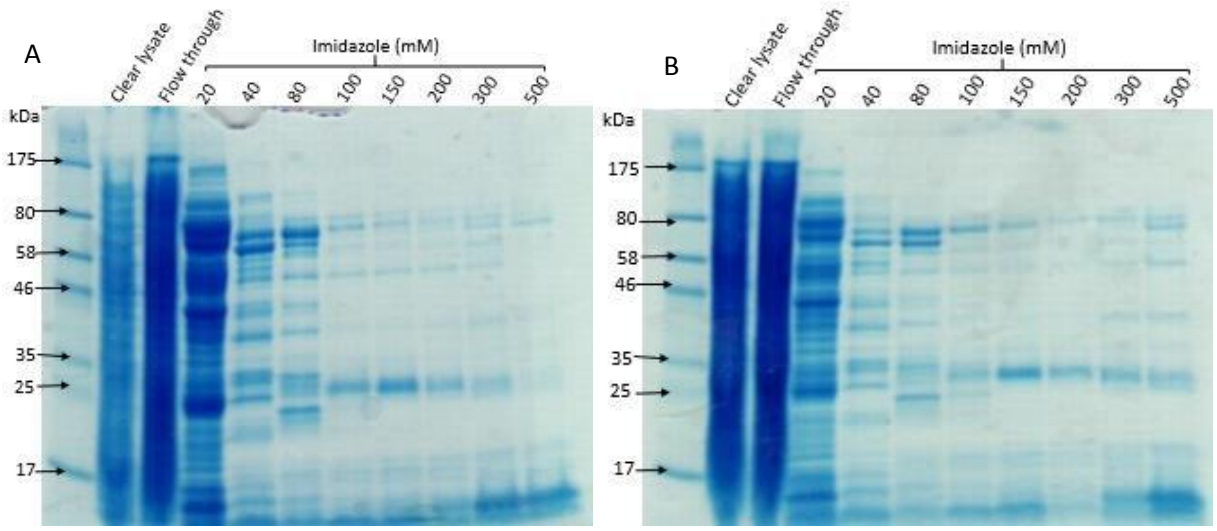


Figure 27. Two SDS-PAGE analyses (A and B) of step gradient imidazole fractions (5 mL per fraction) containing *M. tuberculosis* GlgB after Ni-NTA chromatography to compare two plasmids expression. Molecular weight standards are given in units of kDa and the first lane is the protein marker. Next along is the clear lysate then the flow through followed by the step gradient concentrations of 20, 40, 80, 100, 150, 200, 300 and 500 mM imidazole.

Mycobacterium smegmatis GlgB

For *M. smegmatis* GlgB, three of the recombinant plasmids from the ligation reaction (figure 26, lanes 9, 10 and 11) were transformed into BL21(DE3) cells and the protein expressed. Ni-NTA chromatography was used for each of the cell extracts and proteins were eluted from the column using the same imidazole step gradient as for *M. tuberculosis* GlgB. Fractions were analysed using SDS-PAGE and the resultant gels can be viewed in figure 28 (A, B and C). It is clear that by comparing the three gels, one in particular, gel A, had very strong bands at 84 kDa across all imidazole fractions indicating that the plasmid used gave good expression of GlgB. The other two gels, B and C had bands at 84 kDa but were weak in comparison to A. There are still bands of nonspecific proteins present on gel A, so fractions containing 40 to 500 mM imidazole were pooled and ion exchange was used to further purify the protein.

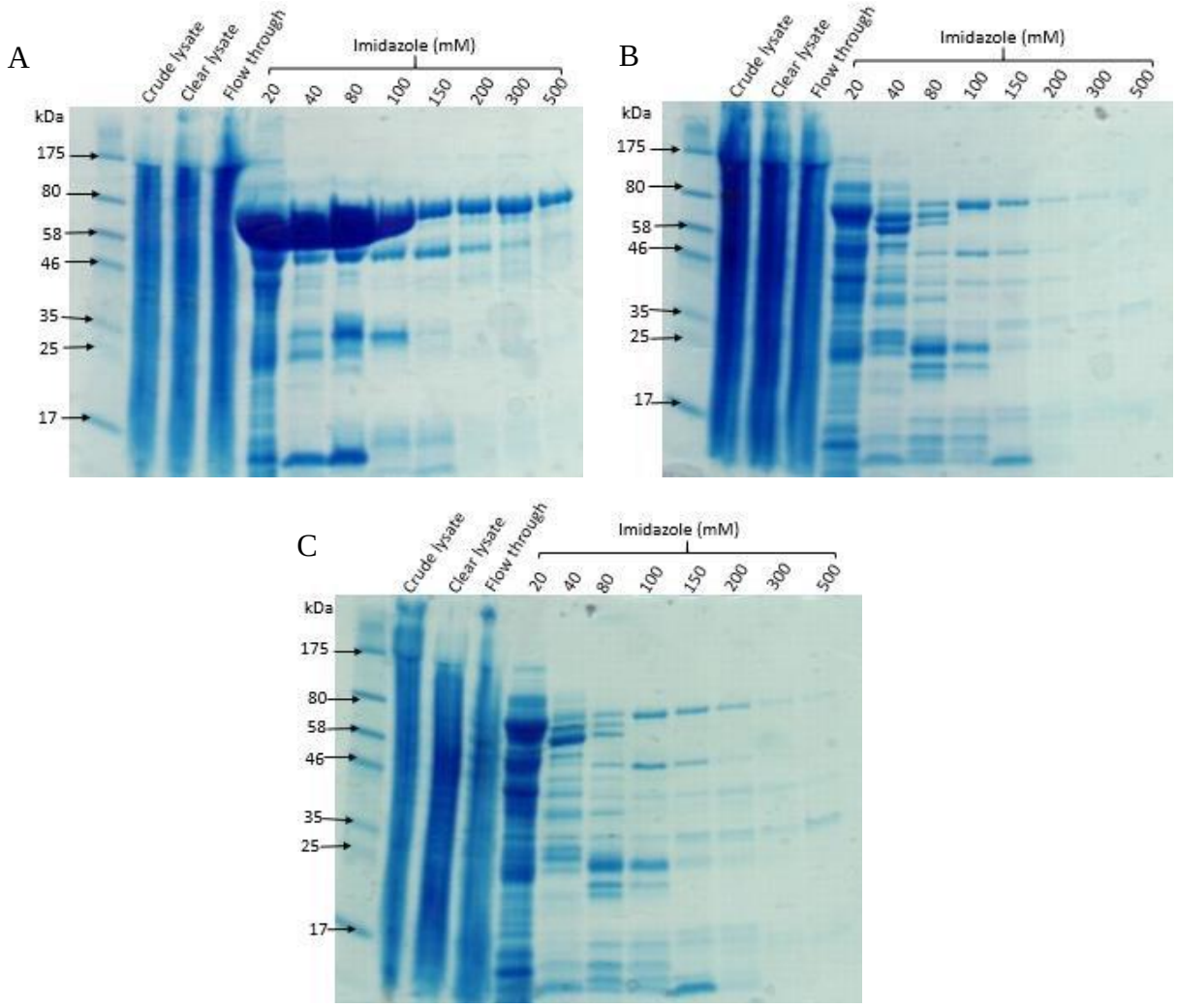


Figure 28. Three SDS-PAGE analyses of step gradient imidazole fractions (5 mL per fraction) containing *M. smegmatis* GlgB after Ni-NTA chromatography to compare three plasmids expression. Molecular weight standards are given in units of kDa and the first lane is the protein marker. Next along is the crude lysate and the clear lysate and the flow through followed by the step gradient concentrations of 20, 40, 80, 100, 150, 200, 300 and 500 mM imidazole.

3.3.4 Ion exchange chromatography

Using the pooled fractions from the previous step, anion exchange chromatography was carried out and a step gradient of NaCl in steps of 100, 150, 200, 250, 300, 350, 400 and 500 mM was used to elute protein from the column (5 mL per fraction). Fractions were assessed using SDS-PAGE and results can be viewed in figure 29. It is clear that all of the GlgB protein was bound to the column as there is no clear band present in the flow through. There are strong bands across all of the fractions at the expected size of 84 kDa yet there are also some bands which are weaker at nonspecific protein sizes. All of the fractions were pooled and used in the next size exclusion purification step.

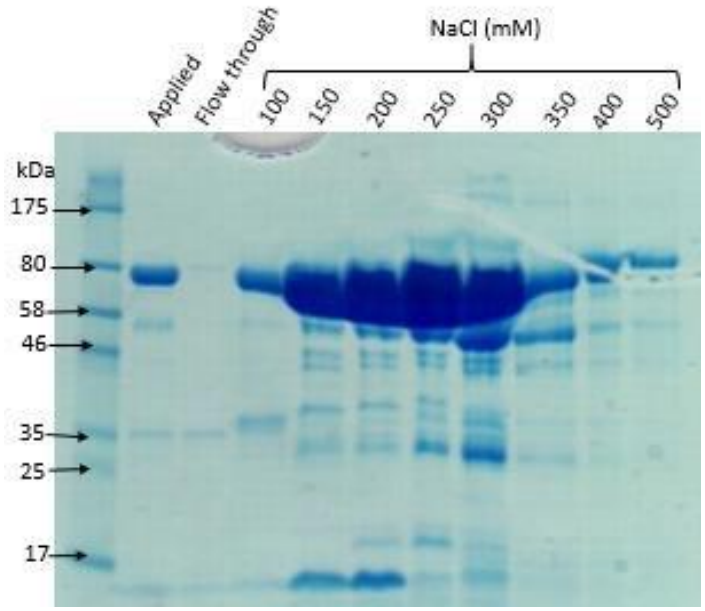


Figure 29. An SDS-PAGE analysis of step gradient NaCl fractions (5 mL per fraction) containing *M. smegmatis* GlgB after ion exchange chromatography. Molecular weight standards are given in units of kDa and the first lane is the protein marker. Next along is the applied and then the flow through followed by the step gradient concentrations of 100, 150, 200, 250, 300, 350, 400 and 500 mM NaCl.

3.3.5 Size exclusion chromatography

A gel filtration column was used to pass through the pooled fractions from ion exchange chromatography and a trace was produced to analyse UV absorbance at 280 nm against elution volume to detect where protein eluted from the column. This was verified also by an SDS-PAGE, both of which can be viewed in figure 30. GlgB protein eluted from the column between 160 and 210 mL with most eluting at 190 mL. Bands were very strong at 84 kDa and there are also much weaker bands at smaller protein sizes. *M. smegmatis* GlgB was concentrated to 54 mg/mL.

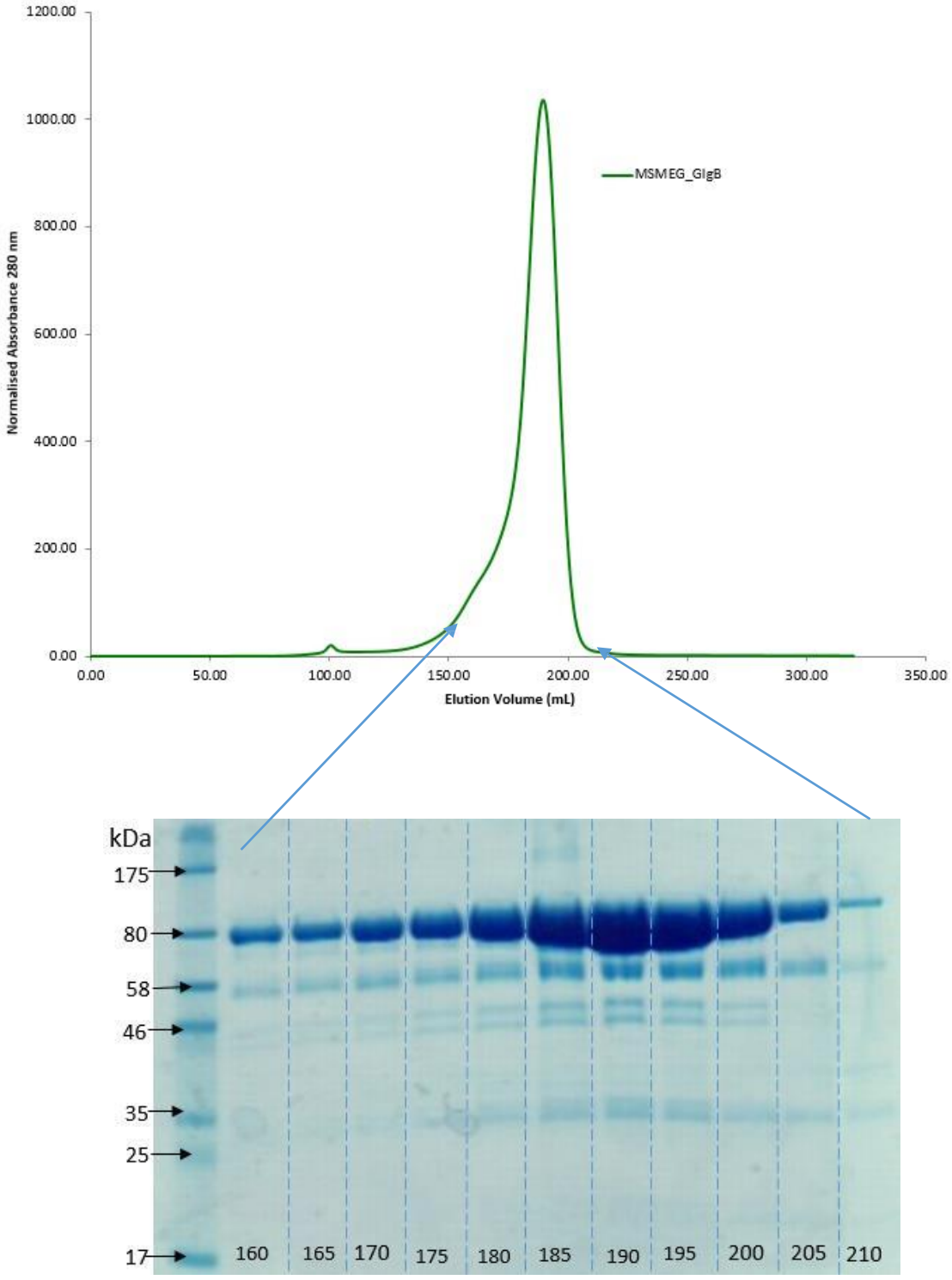


Figure 30. Elution volume (mL) vs. normalised absorbance A_{280} of *M. smegmatis* GlgB elution from a gel filtration column. This was confirmed by SDS-PAGE analysis of fractions (5 mL per fraction). Molecular weight standards are given in units of kDa and the first lane is the protein marker.

4.0 Discussion

4.1 *Mycobacterium tuberculosis* GlgE purification

In order to determine the structure of *M. tuberculosis* GlgE and therefore be able to determine the proteins function on a molecular level and also characterise it, a high concentration of the protein is required that is high in purity.

This project used a number of protein purification techniques in order to try and optimise protein expression and purification of *M. tuberculosis* GlgE in order to attempt to grow protein crystals for structure analysis.

To express the protein, the *M. tuberculosis glgE* gene was inserted into a pET28a plasmid vector. The pET28a vector was chosen as it has a number of factors that make it an attractive option for protein expression in this instance. The first is that it contains a T7 promoter sequence which is taken from a bacteriophage and has a high affinity for T7 RNA polymerase which gives a higher level of expression. Secondly, the plasmid has a multiple cloning site which can undergo restriction digests to allow easy insertion of the gene, in this case the restriction sites of BamHI and NdeI were used. Thirdly, a kanamycin coding sequence is also present on the plasmid which allows for selection of plasmids containing the gene insert after the ligation reaction. A Final reason for selecting this plasmid is that it contains an N-terminal His-tag coding sequence which is required for Ni-NTA chromatography: the first step of GlgE purification.

Ni-NTA chromatography was carried out using a nickel affinity column using the same protocol as Kalscheuer et al. who first expressed *M. tuberculosis* GlgE in *E. coli* (Kalscheuer *et al.,* 2010). The 6xHis-tag expressed on the end of GlgE has an affinity for the nickel ions on the surface of the column leading to the

protein binding to the column. Protein was eluted from the column using a step gradient of imidazole, which has a very similar structure to histidine and displaced the protein from the column. This method successfully eluted GlgE from the column as strong bands can be seen on the SDS-PAGE at 80 kDa but further protein purification was necessary in order to increase protein purity.

In the paper published by Kalscheuer *et al*. (2010) only size-exclusion chromatography was carried out after Ni-NTA chromatography. In this present study, ion exchange chromatography was used after Ni-NTA chromatography in order to test if it would be an appropriate method to use in combination with the previously used purification techniques to increase protein purity.

The technique of ion exchange chromatography is based on the net surface charge of a protein. This can be altered by knowing the proteins isoelectric point (the pH at which a molecule carries no net charge) and in this case the isoelectric point of *M. tuberculosis* GlgE is at pH 5.38. By passing the protein through an anion exchange column in a buffer at pH 7.0 the protein will in theory have a negative charge and should bind to the positive anion column.

Indeed using anion exchange chromatography, some of the nonspecific proteins present after Ni-NTA chromatography were removed, but a band of GlgE at 80 kDa is visible in the flow through indicating that not all of the protein was bound to the column. This resulted in a loss of protein after using this technique which may have been due to the isoelectric point and the fact that it is theoretical. The isoelectric point is calculated by taking the mean of the acid dissociation constants and does not take into account protein folding where some charges may not be present on the surface of the protein.

Even with GlgE loss in the flow through, there are quite strong bands across all of the fractions indicating that a lot of GlgE was bound to the column. A lot of the nonspecific proteins have been removed during this step of purification showing that this technique does increase protein purity.

At the fraction containing 250 mM NaCl (figure 7) there is a large band directly under the band at 80 kDa and rather than being an impurity, it could be due to the fact that proteases are beginning to cleave GlgE as the band was not present after Ni-NTA purification.

Size-exclusion chromatography was carried out using the same protocol as has been carried out by Kalscheuer *et al*. (2010) using the fractions pooled from ion exchange chromatography. The gel filtration column separates proteins on the basis of protein size. The column contains gel beads containing pores which cause smaller proteins to take longer to pass through the column whereas larger proteins cannot enter the pores therefore eluting from the column sooner. The protein that eluted from the column was pure and eluted as a monomer in solution at 180 mL and was concentrated to 6.28 mg/mL which was an increase in comparison to Kalscheuer *et al.* (2010) which concentrated *M. tuberculosis* GlgE to 0.5 mg/mL.

These experiments to express and purify *M. tuberculosis* GlgE were based on a study by Kalscheur *et al*. (2010) and demonstrated that by using larger cultures of transformed bacteria, larger concentrations of GlgE can be produced for use in experiments. This study differed from the previous one in that ion exchange chromatography was used. It could be argued that ion exchange chromatography is an inefficient step in the purification process as protein is visibly lost in the flow through. The extra step of ion-exchange does however remove a lot

of nonspecific proteins that were present after Ni-NTA chromatography. This allows pure GlgE after size exclusion chromatography to be eluted from the column but possibly at the cost of a lower concentration of GlgE. Ion exchange chromatography if used, is a time limiting step as it would need to be repeated a number of times in order to collect the protein lost in the flow through.

4.2 *Mycobacterium tuberculosis* GlgE characterisation

Using the purified GlgE, crystal trays were set up in order to characterise *M. tuberculosis* GlgE. Only one crystal was grown after the structure screen using the commercial solutions. This was at the conditions 0.5 M NaCl, 0.1 M Na citrate, 4% (v/v) polyethyleneimine, pH 5.6 for JCSG plus matrix screen. The crystals were very small and integrated. To try and improve crystallisation, an optimisation screen was carried out using the conditions the crystal grew at in the structure screen using a range of pH and precipitant concentrations. The optimisation attempt was unfortunately unsuccessful and did not produce any crystals.

4.3 *Mycobacterium smegmatis* GlgE purification

For characterisation and for use in experiments to investigate to what extent the four enzymes in the GlgE pathway are involved in complex formation, purified *M. smegmatis* GlgE is required. *M. smegmatis* TreS and Pep2 have already been used to demonstrate complex formation and by purifying GlgE, investigations can continue into its proposed involvement in a multi-catalytic complex with TreS and Pep2 (Roy *et al*., 2013).

The expression of *M. smegmatis* GlgE was carried using the pET28a plasmid

vector for the same reasons as for *M. tuberculosis* GlgE. Protein expression was first tested in 5 L cultures of two types of *E. coli* competent cells: Bl21(DE3) and BL21(DE3)PlysS. BL21(DE3) competent cells had been used previously (Kalscheuer *et al*., 2010) due to the fact that they can be used with pET28a plasmids containing the T7 promoter sequence. BL21(DE3) cells contain the lambda DE3 lysogen prophage which contains the sequence for T7 RNA polymerase. T7 RNA polymerase is required to transcribe the *glgE* gene inserted in the pET28a plasmid once the T7 promoter has been induced using IPTG allowing control of expression.

In a paper that expressed *S. coelicolor* GlgE, BL21(DE3)pLysS competent cells were used (Syson *et al.,* 2011). These cells express T7 lysozyme which suppresses basal level expression before induction with IPTG. Proteins expressed in vectors that are toxic to *E. coli* will give a basal level expression from the T7 RNA polymerase prior to induction which can be harmful to the cells. To overcome this and have a tighter level of expression, T7 lysozyme is encoded for by pLysS which forms a complex at the C- terminal of T7 RNA polymerase causing it to fall off of the T7 promoter to inhibit the gene being transcribed. When IPTG is added, the T7 lysozyme is over ridden by the T7 promoter allowing protein expression to be controlled.

In the case of this project, there was no obvious difference in expression of GlgE using the two types of cells, this could indicate that *M. smegmatis* GlgE is not toxic to BL21(DE3) *E. coli* and the addition of T7 lysozyme is unnecessary. Further experiments in this project used BL21(DE3) cells.

Although Ni-NTA purification of cell extract from a 5 L culture of transformed cells

did produce bands of *M. smegmatis* GlgE at 79 kDa, bands were much smaller in comparison to a 5L culture of *M. tuberculosis* GlgE indicating a lower yield of protein. This prompted the increase of cultures to 10 L in order to purify a higher concentration of *M. smegmatis* GlgE. After Ni-NTA chromatography of the cleared lysate from a 10 L culture of transformed cells, bands on the SDS-PAGE (figure 12) were stronger and indicated an increase in the amount of protein that eluted from the column but there was also an increase in the amount of nonspecific proteins binding to the column. The buffer used in this experiment (20 mM Na_2HPO_4 pH 7.2, 500 mM NaCl, 10% (v/v) glycerol) contained no imidazole, whereas the previous study using the same protocol (Kalscheuer *et al*., 2010) used a buffer containing 20 mM imidazole. This low concentration of imidazole in the sample buffer may help to avoid weak interactions with other proteins and the His-tag and could be added to the buffer used in this experiment in future. Also to overcome the high amount of non-specific binding to the column, an affinity column containing cobalt could be used for better selectivity. A recent paper used a HiTrap TALON crude (GE lifesciences) to purify GlgE (Veleti *et al*., 2014b) which contains a talon ligand that is a tetra-dentate chelator charged with cobalt and is designed to bind with his-tags with a higher selectivity. This column could be considered for future purification to try and decrease the amount of nonspecific proteins binding to the column at this purification step.

Anion exchange chromatography was carried out using the same method as for *M. tuberculosis* GlgE following Ni-NTA chromatography. This further purified the GlgE by removing some nonspecific proteins but quite strong bands were still present at 35 kDa and below (figure 13). There were also weak bands on the SDS-PAGE indicating some proteins present were larger than 79 kDa, this was

unexpected as they were not present on the SDS-PAGE after Ni-NTA chromatography and could have been due to contamination. As was observed previously, there was still GlgE present in the flow through. This time, anion exchange was repeated using the flow through in an attempt to reduce the amount of protein lost at this step of purification (figure 14). The SDS-PAGE for the anion exchange repeat showed that more GlgE and nonspecific proteins were eluted from the column yet not as much as the previous run as all bands were slightly weaker. It is still clear that at this point there was an equal amount of GlgE in the applied and the flow through as both bands were the same strength.

In order to experiment whether the GlgE proteins isoelectric point was accurate, the flow through from the preceding run was applied to a cation exchange column to test whether the protein bound to a column with opposite charge better (figure 15). Results showed that GlgE did bind to the positive column also and eluted across all NaCl fractions but GlgE was still present within the flow though. How the isoelectric point for proteins is worked out is purely theoretical and does not take into account a number of factors such as how the protein is folded and what amino acids will be buried within a folded protein and therefore will have no effect on net charge. It appears that for *M. smegmatis* GlgE, the theoretical isoelectric point is not very accurate.

Fractions from all the ion exchange runs which contained GlgE were pooled and applied to a gel filtration column and the protein was mostly separated from nonspecific proteins that eluted after GlgE (figure 16). The appearance of two bands at around 79 kDa shows that the protein is beginning to be cleaved by proteases present in solution due and this was put down to the time taken to pass all solution repeatedly through the different chromatography columns. This

demonstrates that it is important to only use fresh protein in experiments and complete the experiments one after another.

4.4 *Mycobacterium smegmatis* GlgE characterisation

To assess whether the unexpected bands which were present on the SDS-PAGE gel after Ni-NTA chromatography were either nonspecific proteins or products from GlgE being cleaved by proteases present in solution a western blot was carried out (figure 17). An anti-His antibody was used which binds to the 6xHis-tag present on the GlgE protein. The western blot showed that at 79 kDa, the GlgE protein was present but other areas on the western blot where there were bands present on SDS-PAGE produced no bands indicating that the proteins present are in fact nonspecific proteins that are impurities.

Next, to try and search for a plasmid that may contain an insert of the *M. smegmatis glgE* gene and expresses the protein better than the one used, six plasmids from a previously completed ligation reaction were purified and cultured in 1 L cultures of BL21(DE3) and the protein expressed. It was decided to attempt this as it was found one plasmid in particular for *M. smegmatis* GlgB expressed very well in comparison to other tested plasmids and it was hoped that applying this method for GlgE, a plasmid with improved expression could be detected. Three of the 1 L culture extracts were applied to the Ni-NTA column but none of the plasmids had any increase in expression.

Within the laboratory, a thermal stability assay was carried out by another student (Ali Kermani, School of Biosciences) which indicated that using a buffer containing HEPES at pH 7.0 in 250 mM NaCl may be better suited to have in solution

with GlgE in comparison to the Na_2HPO_4 pH 7.2 buffer that was currently being. This was due to the fact that protein in HEPES pH 7.0 had a higher melting temperature (T_m) so could have higher protein specificity. No noticeable increase in expression was observed however and a lot of nonspecific proteins were still binding to the column and it was concluded the buffer did not have an obvious effect on proteins binding to the column (figure 20).

With expression and purification of *M. smegmatis* GlgE to a concentration of 9.67 $mgmL^{-1}$, experiments could be carried out to characterise the protein. An enzyme activity assay determined GlgE to be an active enzyme for both *M. smegmatis* (figure 21) and *M. tuberculosis* (data not shown). There was however an issue with the control containing no GlgE which showed small absorbance readings over the course of the assay. It was concluded that the buffer used contained phosphate which was causing small background absorbance. In order to get completely accurate results the protein should be purified in a phosphate free buffer and the activity assay repeated to try and avoid this reoccurring.

From the research that demonstrated how TreS and Pep2 form a complex which increases the activity of Pep2 (Roy *et al*., 2013), it was of interest to the laboratory to investigate whether GlgE, when passed through a gel filtration column in solution with TreS or Pep2 was also involved in complex formation. A previous experiment carried out by another student within the laboratory (Ali Kermani, School of Biosciences) carried out SEC for GlgE and Pep2 in solution at pH 6.0 (figure 22). Pep2 eluted as expected but the peak for GlgE was not as expected in that its peak was very short and eluted at around 140 mL instead of around 170 mL indicating the possibility that GlgE may form higher oligomers at pH 6.0. SEC chromatography was carried out for GlgE in buffers of pH 7.0 and pH 6.0. Due

to mechanical failure, during the SEC run at pH 7.0 results were not accurate (see results for this in appendices). The remaining purified GlgE was aliquoted for repeat SEC runs on a different AKTA prime.

The following two runs proved more successful however there was no absorbance read out available using the AKTA machine and the only way in which GlgE elution could be monitored was by using SDS-PAGE analysis. At pH 7.0 (figure 23) faint bands can be seen across fractions 164 to 192 mL indicating that GlgE eluted where it was expected to at this pH. The bands on the gel are faint due to the fact the remaining GlgE protein after the failed experiment had to be aliquoted into two in order to carry out SEC with the functioning machine. Over some of the fractions it is also clear that at around the size of GlgE two bands are present, this was due to the period of time before a second machine could be found to run the experiments and proteases had begun to cleave the protein.

When GlgE was dialysed into buffer pH 6.0 it was noticed that some of the protein had precipitated from the dialysis tubing and the protein concentration was re-measured at 1.72 mg/mL which was very low. As measuring absorbance was unavailable to monitor elution from the gel filtration column, SDS-PAGE analysis was carried out for fractions from 116 to 156 mL where in the previous GlgE:Pep2 size exclusion chromatography run there was a small peak. In this experiment (figure 24, A) there were no bands between these fractions. There were weak bands visible between fractions 204 and 244 mL (figure 24, B) meaning that the protein eluted from the column later than expected. Some bands at around 79 kDa had double bands present, further indicating cleavage of the protein due to the fact it was not fresh. The bands were so weak due to the fact the concentration of protein was so low.

Although preliminary results from this study indicated that GlgE does not form higher oligomers at pH 6.0, there were a number of problems throughout the experiments that may have affected the accuracy of results such as faulty machinery and protein loss. The peak in the Pep2:GlgE SEC run (figure 22) which could indicate that GlgE forms higher oligomers was so small using a high concentration of protein that in this study using such low concentrations of GlgE would most likely not show any peak at all. Due to this, the experiments should be repeated using a fully functional AKTA prime which is able to record the absorbance of elution from the column. Higher concentrations of GlgE used in experiments will also make results clearer by showing more distinctive bands on SDS-PAGE analyses and peaks on absorbance graphs.

This study expressed and purified *M. smegmatis* GlgE successfully which will allow further experimentation to be carried out to characterise the protein in the future.

4.5 *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* GlgB puirfication

PCR was successful for both *M. tuberculosis* and *M. smegmatis glgB* genes. Ligation was also successful and allowed for pilot scale expression to be carried out.

Of the three plasmids containing the *M. smegmatis glgB* insert that were transfected into BL21(DE3) and cultured and expressed in 1 L medium, one expressed the protein very well as very strong bands at 84 kDa were observed after Ni-NTA chromatography. For the two plasmids containing the *M. tuberculosis*

glgB insert and cultured and expressed in 1 L medium, neither gave good expression in comparison to the *M.smegmatis* GlgB purification. In the case of *M. tuberculosis* GlgB this may not be surprising as a previous paper that expressed and purified the protein used the chaperones GroEL and GroES which function to fold the protein correctly (Pal *et al.,* 2010). In future attempts to purify *M. tuberculosis* GlgB it would be sensible to add in the chaperones in order to increase the chance of successful expression.

For the recombinant plasmid which expressed *M. smegmatis* GlgB well, ion exchange chromatography was carried out following Ni-NTA chromatography. It is clear that all the GlgB protein bound to the ion exchange column as there is no band present in the flow through on the SDS-PAGE gel. Some of the nonspecific proteins that were present after Ni-NTA chromatography were also removed. Ion-exchange chromatography in the instance of GlgB purification would be a sensible technique to use as minimal protein is lost and some non-specific proteins were removed. In a previous paper ion exchange chromatography was used to purify *M. tuberculosis* GlgB (Garg *et al*., 2007) yet another paper does not use ion exchange but uses a TALON resin in affinity chromatography which uses cobalt for higher specificity followed by size exclusion chromatography (Pal *et al*., 2010).

M. smegmatis GlgB expressed very well and was purified to a concentration of 54 $mgmL^{-1}$. For use in further experimentation, the protein will need to be characterised and it will need to be shown that the GlgB protein which has been expressed and purified is an active enzyme. This demonstrates a protocol in which to express and purify *M. smegmatis* GlgB for use in future inhibition studies.

4.6 Conclusions

This project has successfully expressed and purified both *M. tuberculosis* and *M. smegmatis* GlgE which will in the future allow further investigations to be carried out to further characterise the protein and search for potential inhibitors of the enzyme. Along with GlgE, this project also successfully expressed and purified *M. smegmatis* GlgB which will allow the enzyme to be further characterised and also allow inhibitor screens to be carried out. *M. tuberculosis* GlgB was found to be expressed but expression was very low, it is hoped further optimisation of the protocol in the future will improve expression.

5.0 References

Alexander, K.A., Laver, P.N., Michel, A.L., Williams, M., van Helden, P.D., Warren, R.M., Gey van Pittius, N.C., 2010. Novel Mycobacterium tuberculosis Complex Pathogen, M. mungi. Emerg. Infect. Dis. 16, 1296–1299.

Böddinghaus, B., Rogall, T., Flohr, T., Blöcker, H., Böttger, E.C., 1990. Detection and identification of mycobacteria by amplification of rRNA. J. Clin. Microbiol. 28, 1751–1759.

Boehme, C.C., Nabeta, P., Henostroza, G., Raqib, R., Rahim, Z., Gerhardt, M., Sanga, E., Hoelscher, M., Notomi, T., Hase, T., Perkins, M.D., 2007.Operational Feasibility of Using Loop-Mediated Isothermal Amplification for Diagnosis of Pulmonary Tuberculosis in Microscopy Centers of Developing Countries. J. Clin. Microbiol. 45, 1936–1940.

Brennan, P., Crick, D., 2007. The Cell-Wall Core of Mycobacterium tuberculosis in the Context of Drug Discovery. Curr. Top. Med. Chem. 7, 475–488.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L.M., Pym, A.S., Samper, S., van Soolingen, D., Cole, S.T., 2002. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc. Natl. Acad. Sci. U. S. A. 99, 3684–3689.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B., 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 37, D233–D238.

Chandra, G., Chater, K.F., Bornemann, S., 2011. Unexpected and widespread connections between bacterial glycogen and trehalose metabolism.Microbiology 157, 1565–1572.

Chapman, G.B., Hanks, J.H., Wallace, J.H., 1959. An electron microscope study of the disposition and fine structure of *Mycobacterium lepraemurium* in mouse spleen. J. Bacteriol. 77, 205–211.

Chopra, I., Brennan, P., 1998. Molecular action of anti-mycobacterial agents. Tuber.Lung Dis. 78, 89–98.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell,

T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.-A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., Barrell, B.G., 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393, 537–544.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., Yeboah-Manu, D., Bothamley, G., Mei, J., Wei, L., Bentley, S., Harris, S.R., Niemann, S., Diel, R., Aseffa, A., Gao, Q., Young, D., Gagneux, S., 2013. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat. Genet. 45, 1176–1182.

Daffé, M., Etienne, G., 1999. The capsule of Mycobacterium tuberculosis and its implications for pathogenicity. Tuber. Lung Dis. 79, 153–169.

De Jong, B.C., Adetifa, I., Walther, B., Hill, P.C., Antonio, M., Ota, M., Adegbola, R.A., 2010. Differences between TB cases infected with M. africanum, West-African type 2, relative to Euro-American M. tuberculosis- an update. FEMS Immunol. Med. Microbiol. 58, 102–105.

Dheda, K., Ruhwald, M., Theron, G., Peter, J., Yam, W.C., 2013. Point-of-care diagnosis of tuberculosis: Past, present and future. Respirology 18, 217–232.

Fine, P.E.M., 1995. Variation in protection by BCG: implications of and for heterologous immunity. The Lancet 346, 1339–1345.

Fischbach, M.A., Walsh, C.T., 2009. Antibiotics For Emerging Pathogens. Science 325, 1089–1093.

Flynn, J.L., Chan, J., 2003. Immune evasion by Mycobacterium tuberculosis: living with the enemy. Curr. Opin. Immunol. 15, 450–455.

Gandhi, N.R., Nunn, P., Dheda, K., Schaaf, H.S., Zignol, M., van Soolingen, D., Jensen, P., Bayona, J., 2010. Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. The Lancet 375, 1830–1843.

Garg, S.K., Alam, M.S., Kishan, K.V.R., Agrawal, P., 2007. Expression and characterization of α-(1,4)-glucan branching enzyme Rv1326c of Mycobacterium tuberculosis H37Rv. Protein Expr. Purif. 51, 198–208.

Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempe, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, P.R., Parkhill, J., Barrell, B.G., Cole, S.T., Gordon, S.V., Hewinson, R.G., 2003. The complete genome sequence of Mycobacterium bovis. Proc. Natl. Acad. Sci. 100, 7877–7882. Gillespie, S.H., 2007. Tuberculosis: evolution in millennia and minutes. Biochem.Soc. Trans. 35, 1317.

Gler, M.T., Skripconoka, V., Sanchez-Garavito, E., Xiao, H., Cabrera-Rivero, J.L., Vargas-Vasquez, D.E., Gao, M., Awad, M., Park, S.-K., Shim, T.S., Suh, G.Y., Danilovits, M., Ogata, H., Kurve, A., Chang, J., Suzuki, K., Tupasi, T., Koh, W.-J., Seaworth, B., Geiter, L.J., Wells, C.D., 2012. Delamanid for Multidrug- Resistant Pulmonary Tuberculosis. N. Engl. J. Med. 366, 2151–2160.

Imaeda, T., 1985. Deoxyribonucleic Acid Relatedness Among Selected Strains of Mycobacterium tuberculosis, Mycobacterium bovis, Mycobacterium bovis BCG, Mycobacterium microti, and Mycobacterium africanum. Int. J. Syst. Bacteriol. 35, 147–150.

Kalscheuer, R., Syson, K., Veeraraghavan, U., Weinrick, B., Biermann, K.E., Liu, Z., Sacchettini, J.C., Besra, G., Bornemann, S., Jacobs, W.R., 2010. Self- poisoning of Mycobacterium tuberculosis by targeting GlgE in an α-glucan pathway. Nat. Chem. Biol. 6, 376–384.

Kapur, V., Whittam, T.S., Musser, J.M., 1994. Is Mycobacterium tuberculosis 15,000 Years Old? J. Infect. Dis. 170, 1348–1349.

Kaufmann, S.H.E., 1993. Immunity to Intracellular Bacteria. Annu. Rev. Immunol. 11,129–163.

Kaur, D., Guerin, M.E., Skovierova, H., Brennan, P.J., Jackson, M., 2009. Biogenesis of the cell wall and other glycoconjugates of Mycobacterium tuberculosis. Adv. Appl. Microbiol. 69, 23–78.

Koeck, J.-L., Fabre, M., Simon, F., Daffé, M., Garnotel, É. Matan, A.B., Gérôme, P., Bernatas, J.-J., Buisson, Y., Pourcel, C., 2011. Clinical characteristics of the smooth tubercle bacilli “Mycobacterium canettii” infection suggest the existence of an environmental reservoir. Clin. Microbiol. Infect. 17, 1013–1019.

Koul, A., Arnoult, E., Lounis, N., Guillemont, J., Andries, K., 2011. The challenge of new drug discovery for tuberculosis. Nature 469, 483–490.

Koul, A., Dendouga, N., Vergauwen, K., Molenberghs, B., Vranckx, L., Willebrords, R., Ristic, Z., Lill, H., Dorange, I., Guillemont, J., Bald, D., Andries, K., 2007. Diarylquinolines target subunit c of mycobacterial ATP synthase. Nat. Chem. Biol. 3, 323–324.

Lambrecht, B.N., 2006. Alveolar Macrophage in the Driver's Seat. Immunity 24, 366– 368.

Lillebaek, T., Dirksen, A., Baess, I., Strunge, B., Thomsen, V.Ø., Andersen, Å.B., 2002. Molecular Evidence of Endogenous Reactivation of Mycobacterium tuberculosis after 33 Years of Latent Infection. J. Infect. Dis. 185, 401–404.

Ling, D.I., Flores, L.L., Riley, L.W., Pai, M., 2008. Commercial Nucleic-Acid Amplification Tests for Diagnosis of Pulmonary Tuberculosis in Respiratory Specimens: Meta-Analysis and Meta-Regression. PLoS ONE 3.

Mahajan, R., 2013. Bedaquiline: First FDA-approved tuberculosis drug in 40 years. Int. J. Appl. Basic Med. Res. 3, 1–2.

Makarov, V., Manina, G., Mikusova, K., Mollmann, U., Ryabova, O., Saint-Joanis, B., Dhar, N., Pasca, M.R., Buroni, S., Lucarelli, A.P., Milano, A., De Rossi, E., Belanova, M., Bobovska, A., Dianiskova, P., Kordulakova, J., Sala, C., Fullam, E., Schneider, P., McKinney, J.D., Brodin, P., Christophe, T., Waddell, S., Butcher, P., Albrethsen, J., Rosenkrands, I., Brosch, R., Nandi, V., Bharath, S., Gaonkar, S., Shandil, R.K., Balasubramanian, V., Balganesh, T., Tyagi, S., Grosset, J., Riccardi, G., Cole, S.T., 2009. Benzothiazinones Kill Mycobacterium tuberculosis by Blocking Arabinan Synthesis. Science 324, 801–804.

Migliori, G.B., Ortmann, J., Girardi, E., Besozzi, G., Lange, C., Cirillo, D.M., Ferrarese, M., De Iaco, G., Gori, A., Raviglione, M.C., 2007. Extensively Drug- resistant Tuberculosis, Italy and Germany. Emerg. Infect. Dis. 13, 780–782.

Murray, J.F., 1990. Cursed Duet: HIV Infection and Tuberculosis. Respiration 57, 210–220.

Pal, K., Kumar, S., Sharma, S., Garg, S.K., Alam, M.S., Xu, H.E., Agrawal, P., Swaminathan, K., 2010. Crystal Structure of Full-length Mycobacterium tuberculosis H37Rv Glycogen Branching Enzyme. J. Biol. Chem. 285,

20897– 20903.

Palomino, J., Martin, A., 2013. Tuberculosis Clinical Trial Update and the Current Anti-Tuberculosis Drug Portfolio. Curr. Med. Chem. 20, 3785–3796.

Payne, D.J., Gwynn, M.N., Holmes, D.J., Pompliano, D.L., 2007. Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nat. Rev. Drug Discov. 6, 29–40.

Phillips, M., Basa-Dalay, V., Blais, J., Bothamley, G., Chaturvedi, A., Modi, K.D., Pandya, M., Natividad, M.P.R., Patel, U., Ramraje, N.N., Schmitt, P., Udwadia, Z.F., 2012. Point-of-care breath test for biomarkers of active pulmonary tuberculosis. Tuberculosis 92, 314–320.

Pieters, J., 2008. Mycobacterium tuberculosis and the Macrophage: Maintaining a Balance. Cell Host Microbe 3, 399–407.

Rangaka, M.X., Wilkinson, K.A., Glynn, J.R., Ling, D., Menzies, D., Mwansa-Kambafwile, J., Fielding, K., Wilkinson, R.J., Pai, M., 2012. Predictive value of interferon-Ɣ release assays for incident active tuberculosis: a systematic review and meta-analysis. Lancet Infect. Dis. 12, 45–55.

Rattan, A., Kalia, A., Ahmad, N., 1998. Multidrug-resistant Mycobacterium tuberculosis: molecular perspectives. Emerg. Infect. Dis. 4, 195–209.

Riley, R.L., Mills, C.C., Nyka, W., Weinstock, N., Storey, P.B., Sultan, L.U., Riley, M.C., Wells, W.F., 1995. Aerial Dissemination Op Pulmonary Tuberculosis: A Two-Year Study of Contagion in a Tuberculosis Ward. Am. J. Epidemiol. 142, 3–14.

Rodrigues, C., Shenai, S., Sadani, M., Sukhadia, N., Jani, M., Ajbani, K., Sodha, A., Mehta, A., 2009. Evaluation of the bactec MGIT 960 TB 0 system for recovery and identification of *Mycobacterium tuberculosis* complex in a high through put tertiary care centre. Indian J. Med. Microbiol. 27, 217.

Rohde, K., Yates, R.M., Purdy, G.E., Russell, D.G., 2007. Mycobacterium tuberculosis and the environment within the phagosome. Immunol. Rev. 219, 37–54.

Roy, R., Usha, V., Kermani, A., Scott, D.J., Hyde, E.I., Besra, G.S., Alderwick, L.J., Futterer, K., 2013. Synthesis of α-Glucan in Mycobacteria Involves a Hetero-octameric Complex of Trehalose Synthase TreS and Maltokinase Pep2. ACS Chem. Biol. 8, 2245–2255.

Russell, D.G., 2007. Who puts the tubercle in tuberculosis? Nat. Rev. Microbiol. 5,

39–47.

Russell, D.G., Cardona, P.-J., Kim, M.-J., Allain, S., Altare, F., 2009. Foamy macrophages and the progression of the human tuberculosis granuloma. Nat. Immunol. 10, 943–948.

Salo, W.L., Aufderheide, A.C., Buikstra, J., Holcomb, T.A., 1994. Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy. Proc. Natl. Acad. Sci. 91, 2091–2094.

Sambou, T., Dinadayala, P., Stadthagen, G., Barilone, N., Bordat, Y., Constant, P., Levillain, F., Neyrolles, O., Gicquel, B., Lemassu, A., Daffe, M., Jackson, M., 2008. Capsular glucan and intracellular glycogen of Mycobacterium tuberculosis: Biosynthesis and Impact on the Persistence in mice. Mol. Microbiol. 70, 762–774.

Sani, M., Houben, E.N.G., Geurtsen, J., Pierson, J., de Punder, K., van Zon, M., Wever, B., Piersma, S.R., Jimenez, C.R., Daffe, M., Appelmelk, B.J., Bitter, W., van der Wel, N., Peters, P.J., 2010. Direct Visualization by Cryo-EM of the Mycobacterial Capsular Layer: A Labile Structure Containing ESX-1-Secreted Proteins. PLoS Pathog. 6.

Sassetti, C.M., Boyd, D.H., Rubin, E.J., 2003. Genes required for mycobacterial growth defined by high density mutagenesis. Mol. Microbiol. 48, 77–84.

Saunders, B.M., Britton, W.J., 2007. Life and death in the granuloma: immunopathology of tuberculosis. Immunol. Cell Biol. 85, 103–111.

Sengupta, S., Roy, D., Bandyopadhyay, S., 2014. Modeling of a new tubercular maltosyl transferase, GlgE, study of its binding sites and virtual screening. Mol. Biol. Rep. 41, 3549–3560.

Sharma, S.K., Mohan, A., 2013. Tuberculosis: From an incurable scourge to a curable disease - journey over a millennium. Indian J. Med. Res. 137, 455–493.

Silva, P.E.A.D., Palomino, J.C., 2011. Molecular basis and mechanisms of drug resistance in Mycobacterium tuberculosis: classical and new drugs. J. Antimicrob. Chemother. 66, 1417–1430.

Sreevatsan, S., Pan, X., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S., Musser, J.M., 1997. Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global

dissemination. Proc. Natl. Acad. Sci. U. S. A. 94, 9869–9874.
Stead, W.W., Eisenach, K.D., Cave, M.D., Beggs, M.L., Templeton, G.L., Thoen, C.O., Bates, J.H., 1995. When did Mycobacterium tuberculosis infection first occur in the New World? An important question with public health implications. Am. J. Respir. Crit. Care Med. 151, 1267–1268.
Syson, K., Stevenson, C.E.M., Rashid, A.M., Saalbach, G., Tang, M., Tuukkanen, A., Svergun, D.I., Withers, S.G., Lawson, D.M., Bornemann, S., 2014. Structural Insight into How Streptomyces coelicolor Maltosyl Transferase GlgE Binds Maltose 1-Phosphate and Forms a Maltosyl-enzyme Intermediate.Biochemistry (Mosc.) 53, 2494–2504.
Syson, K., Stevenson, C.E.M., Rejzek, M., Fairhurst, S.A., Nair, A., Bruton, C.J., Field, R.A., Chater, K.F., Lawson, D.M., Bornemann, S., 2011. Structure of Streptomyces Maltosyltransferase GlgE, a Homologue of a Genetically Validated Anti-tuberculosis Target. J. Biol. Chem. 286, 38298–38310.
Takayama, K., Armstrong, E.L., Kunugi, K.A., Kilburn, J.O., 1979. Inhibition by Ethambutol of Mycolic Acid Transfer into the Cell Wall of Mycobacterium smegmatis. Antimicrob. Agents Chemother. 16, 240–242.
Tameris, M.D., Hatherill, M., Landry, B.S., Scriba, T.J., Snowden, M.A., Lockhart, S., Shea, J.E., McClain, J.B., Hussey, G.D., Hanekom, W.A., Mahomed, H., McShane, H., 2013. Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. The Lancet 381, 1021–1028.
Treatment of Pulmonary Tuberculosis with Streptomycin and Para-Amino-Salicylic Acid, 1950. . Br. Med. J. 2, 1073–1085.
Velayati, A.A., Farnia, P., Masjedi, M.R., 2013. The totally drug resistant tuberculosis (TDR-TB). Int. J. Clin. Exp. Med. 6, 307–309.
Velayati, A.A., Masjedi, M.R., Farnia, P., Tabarsi, P., Ghanavi, J., ZiaZarifi, A.H., Hoffner, S.E., 2009. Emergence of new forms of totally drug-resistant tuberculosis bacilli: Super extensively drug-resistant tuberculosis or totally drug resistant strains in Iran. CHEST J. 136, 420–425.
Veleti, S.K., Lindenberger, J.J., Ronning, D.R., Sucheck, S.J., 2014a. Synthesis of a C-phosphonate mimic of maltose-1-phosphate and inhibition studies on Mycobacterium tuberculosis GlgE. Bioorg. Med. Chem. 22, 1404–1411.
Veleti, S.K., Lindenberger, J.J., Ronning, D.R., Sucheck, S.J., 2014b. Synthesis of

a C-phosphonate mimic of maltose-1-phosphate and inhibition studies on Mycobacterium tuberculosis GlgE. Bioorg. Med. Chem. 22, 1404–1411.

Veleti, S.K., Lindenberger, J.J., Thanna, S., Ronning, D.R., Sucheck, S.J., 2014c. Synthesis of a poly-hydroxypyrolidine-based inhibitor of Mycobacterium tuberculosis GlgE. J. Org. Chem.

Weinstein, E.A., Yano, T., Li, L.-S., Avarbock, D., Avarbock, A., Helm, D., McColm, A.A., Duncan, K., Lonsdale, J.T., Rubin, H., 2005. Inhibitors of type II NADH:menaquinone oxidoreductase represent a class of antitubercular drugs. Proc. Natl. Acad. Sci. U. S. A. 102, 4548–4553.

WHO, (2013) Global Tuberculosis Report 2013

Williams, K.N., Stover, C.K., Zhu, T., Tasneen, R., Tyagi, S., Grosset, J.H., Nuermberger, E., 2009. Promising Antituberculosis Activity of the Oxazolidinone PNU-100480 Relative to That of Linezolid in a Murine Model. Antimicrob. Agents Chemother. 53, 1314–1319.

6.0 Appendices

Terrific Broth Media (5L)

Tryptone 60 g
Yeast Extract 120 g
ddH_2O Glycerol 18 g
} Make up to 4500 mL with

Potassium Phosphate monobasic 11.5 g
ddH_2O Potassium phosphate dibasic 62.5 g
} Make up to 500 mL with

Autoclave the media and salts separately and combine just before use.

Western Buffer

10% 10X running buffer
20% methanol
70% H_2O

Running Buffer (10X)

30.28 g Tris base
144.13 g Glycine
10 g SDS

Size exclusion chromatography *M. smegmatis* GlgE pH 7.0

Size exclusion chromatography of *M. smegmatis* GlgE in solution at pH 7.0 (figure 30) was not successful as can be seen in the image below where protein can be seen to elute at a range of volumes giving a strange absorbance reading also. It was deemed that the AKTA prime machine used for the experiment had mechanical failure.

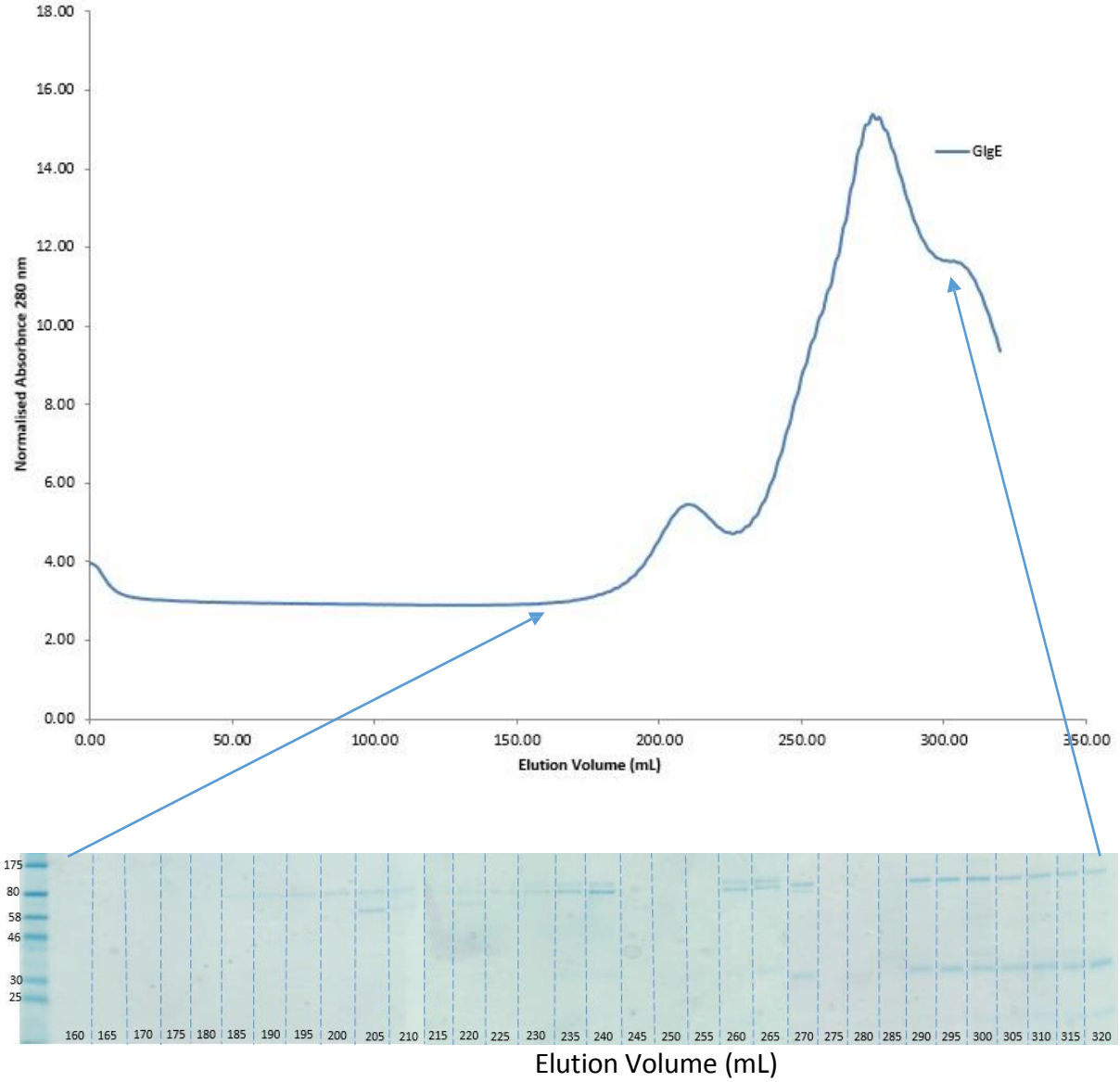


Figure 30. Elution volume (mL) vs. normalised absorbance A_{280} of *M. smegmatis* GlgE elution from a gel filtration column at pH 7.0. An unexpected result was further confirmed by SDS-PAGE analysis of fractions (5 mL per fraction). Molecular weight standards are given in units of kDa and the first lane is the protein marker.

Project Two

Large subunit ribosomal protein 41 is preferentially found in *Drosophila melanogaster* larval midgut stem cells

A thesis submitted in partial fulfillment for the award of MRes in Molecular and Cellular Biology

Supervisor: Dr Saverio Brogna

School of Biosciences
University of Birmingham

September 2014

Acknowledgements

I would first like to thank my supervisor, Dr Saverio Brogna, for offering me the opportunity to carry out my research project within his lab and for his support and advice throughout the project. I would also like to thank everybody else within the Brogna lab for the kind help they have given me throughout this project.

Abstract

Ribosomal proteins have an essential role in ribosome biogenesis and function but research is beginning to show they can also have extra ribosomal functions while not part of the ribosome. It has also been reported that ribosomes are heterogeneous so that they for example can differ in the subset of ribosomal proteins they incorporate. This project characterised a number of transgenic *Drosophila melanogaster* strains expressing a number of ribosomal proteins tagged with YFP at their endogenous loci. Expression of these proteins was visualised in salivary gland, fat body and midgut cells. While most of these proteins were detected throughout the cytoplasm and nucleus in all of the cell types, within the midgut RpL41-YFP showed strong expression in adult midgut precursor (AMP) stem cells. An immunostain using an antibody against the endogenous RpL41 confirmed higher expression in AMP stem cells and also in a small area within the nucleus of differentiated cells, but the specificity of the antibody used has yet to be confirmed. Visualisation of these proteins was also carried out during larval development and the expression within the enterocyte nuclei was detected only during early second larval instar. In summary, the observations made indicate that RpL41 is differentially expressed in different cells. Remarkably within the gut this protein is more abundant in the nucleus of stem cells.

Contents

1.0 Introduction 1
1.1 Regulation of eukaryotic gene expression 1
1.2 The eukaryotic ribosome 2
1.3 Specialised ribosomes 3
1.3.1 rRNA Heterogeneity 3
1.3.2 Ribosomal protein gene paralogues 4
1.3.3 Differences in ribosomal protein expression 5
1.3.4 Ribosomal protein post translational modifications 6
1.3.5 Other levels of regulatory control of the ribosome 6
1.4 Extra-ribosomal protein functions 7
1.5 Nuclear translation 10
1.5.1 Past evidence 11
1.5.2 Ribopuromycylation 11
1.5.3 Bimolecular fluorescence complementation 12
1.6 Research project aims 15
2.0 Materials and methods 16
2.1 *Drosophila melanogaster* fly lines and transgenes 16
2.2 Fly stock handling and food 17
2.3 Midgut dissection 17
2.4 RpL41 Immunostain 18
2.5 Microscopy 19
2.6 PCR 19
2.7 DNA Sequencing 21
2.8 Western blot 22
3.0 Results 24
3.1 *Drosophila melanogaster* salivary gland cell transgene expression 24
3.2 *Drosophila melanogaster* fat body cell transgene expression 26
3.3 *Drosophila melanogaster* midgut transgene expression 28
3.4 Confirming YFP inserts using PCR 32
3.5 Sequencing PCR products 34
3.6 Western blot 35
3.7 Immunostain 40
3.8 RpL41 expression during larval development 42

4.0 Discussion..... 44
4.1 Conclusion…..…….....50
5.0 References.....….51

1.0 Introduction

1.1 Regulation of eukaryotic gene expression

The discovery of the structure of DNA in 1953 by Francis Crick and James Watson (Watson and Crick, 1953) using Rosalind Franklin's data (Maddox, 2003) led to a flourish of research in the field of molecular biology. In the years after this initial discovery the process of gene expression; transcribing DNA into messenger RNA (mRNA) and translating it into proteins, was deduced. Gene expression has proven to be a highly regulated process with regulatory control found at almost every step which in turn allows a diverse range of proteins to be synthesised in different cells types and under a range of conditions and stimuli (Struhl, 1999).

Transcription for example is regulated by a range of transcription factors (TFs) which are proteins that are able to bind to specific DNA sequences and these are able to control expression by either activating or blocking a sequence of DNA (Latchman, 1997).

Chromatin, formed using histone proteins which supercoil DNA into nucleosomes within the eukaryotic nucleus is also able to regulate transcription by changing its state under certain stimuli or for specific cell types. Loosely packed chromatin, known as euchromatin, allows TFs to bind to the DNA and transcribe it whereas heterochromatin is tightly packaged and blocks access of TF's therefore blocking transcription (Dame, 2005). Regulation of this manner is very diverse in that clusters of genes which are simultaneous in function or are structurally related can be expressed in a co-ordinated fashion (Sproul *et al*., 2005).

Histone proteins themselves can also be subject to modifications in that their amino acids can be for example acetylated, methylated, phosphorylated or

ubiquitinated which can alter the structure of the proteins and the histones ability to bind DNA (Bannister and Kouzarides, 2011). In addition, the DNA itself can also be methylated which can lock genes from being transcribed (Suzuki and Bird, 2008). Gene expression can be regulated also post-transcriptionally, for instance by microRNAs (miRNAs); miRNAs are small, non-coding RNAs which have the ability to bind to complementary mRNA sequences and effectively silence the translation of proteins (Bartel, 2009).

A final example of post-transcriptional gene expression control is pre-mRNA processing. Following transcription, single stranded immature pre-mRNA are produced and are made up of introns and exons. Splicing removes the introns and covalently bonds the exons together to form mature mRNA which can then be exported from the nucleus into the cytoplasm to be translated. Pre-mRNA processing takes place in the spliceosome within the nucleus which is made up of small nuclear ribonucleic particles (snRNPs) and complexes of proteins (Watson and Baker, 2007).

1.2 The eukaryotic ribosome

Taking into account the many levels of regulation that gene expression goes through when synthesising proteins, the ribosome has long been somewhat overlooked in its ability to also play a role in regulation. The eukaryotic ribosome is the 'machine' within the cell which translates mRNA into proteins and it consists of four species of ribosomal RNAs (rRNAs): 18S, 5.8S, 28S, 5S and about eighty ribosomal proteins (RPs).

The ribosome is made of two subunits; the smaller 40S subunit (made up of 18S rRNA and 32 RPs), and the larger 60S subunit (made up of 28S, 5.8S, 5S rRNAs

and 47 RPs). When mRNA binds, the two subunits come together to form an 80S ribosome and translation begins (Watson and Baker, 2007). Ribosome biogenesis is a complex and highly co-ordinated process that uses three RNA polymerases. 45S pre-rRNA (which later matures into 18S, 5.8S and 28S rRNA) is transcribed in the nucleolus by RNA Pol I and 5S pre-rRNA is transcribed by Pol III in the nucleus. RPs are transcribed by Pol II and are translated in the cytoplasm before getting imported into the nucleus where they assemble on the pre-rRNA. The pre- rRNAs are cleaved to form mature rRNA as the RPs assemble and once the 40S and the 60S subunits are assembled they are exported into the cytoplasm (Kressler *et al*., 2010; Xue and Barna, 2012).

The ribosome has been stereotyped by researchers over the years into the view that it is a 'machine' that is homogenous across the different cell types of multicellular organisms and unaffected by changes in the cell environment (Frank, 2000; Xue and Barna, 2012). However, as detailed below, research now is beginning to indicate that the ribosome is much more plastic and variable than was once thought.

1.3 Specialised ribosomes

1.3.1 rRNA heterogeneity

There is evidence that heterogeneity between ribosomes in different cell types and at different stages of development exists at a number of levels. A good example of heterogeneity at the rRNA level is found in the parasitic protozoan *Plasmodium berghei*. It was discovered that two structurally distinct rRNAs are expressed at two different points in the organism's life cycle. One rRNA gene is expressed during sporozoite development in a mosquito which then switches to the second

gene when inside a mammalian host (Gunderson *et al*., 1987). Another study indicated that the two ribosomes are not functionally equivalent as when *Saccharmoyces cerevisiae* 25S rRNA was substituted for the two types of *Plasmodium falciparum* rRNA, the sporozoite form was shown to be lethal and the second type produced no phenotype (Velichutina *et al*., 1998).

1.3.2 Ribosomal protein gene paralogues

There is also clear heterogeneity when it comes to ribosomal protein genes. In a number of organisms such as flies, yeast and plants, many RPs are encoded for by more than one gene. The plant species *Arabidopsis thaliana* has 2 -7 gene variants for RPs (Barakat *et al*., 2001). This diversity in paralogues allows the ribosome to become specialised in a number of ways. For example, RpS5 in *A. thaliana* has two paralogues; RpS5A and RpS5B. It was discovered that RpS5A was the most abundantly expressed of the two paralogues and was correlated to cells which were dividing. RpS5B was associated with cells which were differentiating (Weijers *et al*., 2001) showing a requirement for alternative paralogues at different stages of development.

Brassica napus has 996 putative genes for 76 RPs. It was found that a large number of paralogues for *B. napus* RPs are regulated in different patterns among varying tissue types and many RPs were expressed in an opposite manner during early development (Whittle and Krochko, 2009). The yeast *S. cerevisiae* has two RP gene copies for 59 of 78 RPs (Kellis *et al*., 2004; Planta and Mager, 1998). Deletions of one of the paralogues led to differing phenotypes such as random bud patterning and bud site selection in comparison to wild type cells (Ni and Snyder, 2001).

The eukaryotic organism *Drosophila melanogaster* has 79 RP genes, 9 of which are found to be duplicated. The paralogue that has the greatest homology to its human gene ortholog is found to be more highly expressed whereas the duplicate gene with less homology to the human ortholog has limited expression. The lesser expressedgenes duplicates RpS5b, RpS19b, RpL10Aa and RpL37b are found to be enriched within the testes (Marygold *et al*., 2007). RpS5b, RpS10a, RpS19b have also been found to be enriched in germ line cells of embryonic gonads (Shigenobu *et al*., 2006) and in the stem cells of adult ovaries (Kai *et al*., 2005) indicating that the genes have a germ line specific role.

In mammals most ribosomal proteins are encoded for by only one gene and it has been found that in humans, chimps, mice and rats 1400–2800 RP pseudogenes exist (Balasubramanian *et al*., 2009). Despite this three paralogues were found in rodents: RpL10-like, RpL22-like and RpL39-like. Expression was evaluated in liver cells, mammary glands and testis and it was found RpL39-like paralogue was expressed as a small ribosomal component and RpL10-like and RpL22-like are expressed only in ribosomes in the testis (Sugihara *et al*., 2010). The human Y-linked paralogue RpS4Y2 is expressed in only testis and prostate while its X-linked homolog is expressed ubiquitously indicating the paralogue could have a function in spermatogenesis (Lopes *et al*., 2010).

1.3.3 Differences in ribosomal protein expression

Ribosomal protein expression is another level at which ribosomes can be specialised. Previously it was thought that RPs were produced in every cell type in an equimolar manner and were transcribed at the same rate, but research was not carried out in a variety of cell types for fair comparison (Hariharan *et al.,* 1989). When a study was carried out looking at adult human brain, liver, skeletal

muscle, ovary, retina and uterus cells large RP expression variation was identified (Bortoluzzi *et al*., 2001).

1.3.4 Ribosomal protein post translational modifications

As for most proteins, ribosomal proteins undergo post translational modifications (PTMs). Large ribosomal subunit 60S was extracted from human placenta and studies showed RPs to be methylated, acetylated and hydroxylated and losses of methionine (Odintsova *et al.*, 2003). Further analysis on the smaller 40S subunit revealed the same types of PTMs plus disulphide bond formation (Yu *et al*., 2005). A number of RPs from *Arabidopsis thaliana* (Carroll *et al*., 2008) and *Saccharomyces cerevisiae* (Lee *et al*., 2002) also have been demonstrated to have a number of PTMs.

A clear example of the requirement for RP PTMs can be found in RpL28 in *S. cerevisiae* which can ligate ubiquitin to itself. In the cell cycle, during S phase the RP is heavily ubiquitinated and then during G_1 phase ubiquitination is reduced (Spence *et al*., 2000) indicating a regulatory role in translation.

1.3.5 Other levels of regulatory control of the ribosome

Additional levels of ribosomal regulation include taking into account ribosome associated proteins and how they act. In *Drosophila melanogaster* Reaper is an apoptotic regulator which can inhibit protein synthesis by binding to the 40S subunit during periods of stress or during apoptosis (Colon-Ramos *et al*., 2006). Another example has been found in HeLa cells; it was shown when phosphorylated glycogen synthase (GSY1) was depleted, translation was affected as subsets of mRNAs were not transcribed to produce proteins required for the

modulation of protein synthesis (Fuchs *et al*., 2011).

Other proposed but lesser researched areas of ribosomal regulatory control includes looking at how *cis*-regulatory elements on mRNA interact with ribosomes during translation (Mauro and Edelman, 2002). It has also been suggested that ten to fifteen percent of proteins are synthesised using a cap-independent mechanism using an internal ribosome entry site (IRES) (Spriggs *et al*., 2008). Studies have also shown that modified psuedouridylated rRNAs do not translate some mRNAs correctly (Jack *et al*., 2011). A final level of regulatory control can be found in upstream open reading frames (uORFs) which have been found to be present in around half of human and mouse transcripts. Most uORFs repress translation of the downstream main uORF (Calvo *et al*., 2009). RpL24 has been identified as an important factor in reinitiating translation of mRNAs which have uORFs in *Arabidopsis thaliana* (Nishimura *et al*., 2005; Park *et al*., 2001).

In summary it is clear that the ribosome can be regulated in a variety of ways to allow translation to be carried out for a cells specific requirements or during different stages of development and there are many factors which are involved to allow the ribosome to become specialised in its function.

1.4 Extra-ribosomal protein functions

Taking into consideration how complex most biological mechanisms are it would not be unrealistic to consider that the many RPs that make up the ribosome could also have roles while not part of the ribosome. Since the 1970's there has been research backing this idea with some research being more conclusive than others due to the difficulties in providing abundant evidence (Lindström, 2009; Warner and McIntosh, 2009; Wool, 1996).

Studies indicate that some RPs are able to regulate their own expression by their ability to bind their own promoter, mRNAs or pre-mRNAs which gives them the ability to regulate translation, splicing or transcription (Warner and McIntosh, 2009). An example of this can be found in the mammalian RpS13 which has been found to repress its own gene expression by binding its own pre-mRNA near splice sites of the first intron which inhibits splicing (Malygin *et al*., 2007). Other examples include human RpS26 which also inhibits splicing of its own pre-mRNA (Ivanov *et al*., 2004) and *S. cerevisiae* RpS14B which regulates its expression by binding to RpS14B pre-mRNA and to 18S rRNA (Fewell and Woolford, 1999).

It was also discovered that RpL11 could bind to specific transcription factors (TFs) and regulate genes in that manner. The mammalian RpL11 can associate with the oncoprotein c-Myc which in turn inactivates c-Myc target genes (Dai *et al*., 2010, 2007). In addition to this, RpS3 has been shown to associate with the nuclear factor κB (NF- κB) DNA binding protein complex and allows the complex to stably associate with its target regulatory site (Wan *et al*., 2007).

For a while it was understood that RPs performing extra ribosomal functions within the nucleus were free and not bound to any subunit. In opposition to this view, research has also shown that in biological preparations of complexes which are involved in transcription and pre-mRNA processing, RPs are often found together but never as 40S or 60S subunits (Gavin *et al*., 2002; Jurica and Moore, 2003; Shi *et al*., 2009). The RPs were at the time identified as contamination which forms when proteins associate after cell lysis (Gavin *et al*., 2002).

In view of these findings, it was considered that what was assumed were free RPs with extra ribosomal functions could actually have once been a RP complex which

during cell lysate extraction lost all of the RPs which were not bound as tightly (De and Brogna, 2010). For example, RpL11 which was shown to be the sole RP present interacting with c-Myc could have been present as a complex of RPs which after extraction left only RpL11 to be detected. When the protocol for this study was studied it was found EDTA was used in the lysis solution which is an acid known to cause RPs to dissociate (Steitz *et al*., 1988).

This evidence that ribosomal subunits could carry out extra ribosomal functions at chromosomal sites was further backed up by another study which used genome wide ChIP-on-chip analysis to show that along with RpL11, two other 60S RPs associate with the same sub set of specific chromosomal loci (De *et al*., 2011). In further support of ribosomal subunits having extra ribosomal functions, a study of *D. melanogaster* found that 20 RPs and rRNAs were present at a number of transcription sites on polytene chromosomes which were interacting with nascent mRNA (Brogna *et al*., 2002). This study was however later criticised in that the antibodies used in the study may not have been specific enough and may have cross reacted with common components of transcription sites (Bohnsack *et al*., 2002). Following this, a study on *S. cerevisiae* carried out ChIP experiments and found that 4 RPs were associated with transcription sites, yet it was still not distinguishable whether the RPs interacted as free RPs or as ribosomal subunits (Schroder and Moore, 2005).

A more recent investigation claims to have overcome the issue of non-specific antibodies detecting extra ribosomal proteins at transcription sites on polytene chromosomes. Rugjee *et al*. (2013) tagged several RPs with either GFP or RFP and found accumulation within the cytoplasm and nucleolus in the salivary gland cells of *D. melanogaster*. Polytene chromosome immunostaining confirmed the

RPs associate with transcription loci which supports findings from the previous study by Brogna *et al*. (2002).

In summary it is clear that some RPs do have extra ribosomal functions and appear able to interact with sites of transcription, but it has yet to be concluded whether RPs are present as subunits, on their own, or both.

1.5 Nuclear translation

Bearing in mind the evidence for ribosomal subunits being present in at transcription sites in the nucleus, emerging research should also be taken into consideration which indicates that translation can take place within a cells nuclear compartment. There are however a number of observations arguing against nuclear translation. For a start it is thought that the 40S subunit is only able to become active in translation once it has been transported out of the nucleus and into the cytoplasm where its 20S rRNA is trimmed to 18S and assembly factors are able to bind to the pre-40S to prevent translation (Rouquette *et al*., 2005; Strunk *et al*., 2011; Udem and Warner, 1973). There is however evidence which opposes this and indicates that pre-40S subunits are able to associate with ribosomal complexes either independently or dependently on translation (Lebaron *et al*., 2012; Soudet *et al*., 2010).

Secondly, the 60S subunit is thought to only be able to become active for translation once in the cytoplasm (Udem and Warner, 1973). The 60S subunit is exported from the nucleus along with non-ribosomal factors which are removed once in the cytoplasm where the 60S can also gain RPs to become fully functional (Panse and Johnson, 2010). Despite this, evidence for nuclear translation has been growing since the 1950s and two recent studies in particular

cannot go unnoticed.

1.5.1 Past evidence

Older reports of nuclear translation were controversial, heavily debated and mostly dismissed at the time (Dahlberg and Lund, 2004). One study found that radioactively labelled cysteine C^{14} could incorporate into nascent proteins within nuclei which were isolated from calf thymus (Allfrey, 1954). Another reported the presence of polyribosomes in HeLa cells which could be released from the nucleus in the presence of Poly(U) (Goidl *et al*., 1975). Many of these results were dismissed as contamination and argued that evidence was not conclusive enough to warrant it being approved by the research community.

Later, a study which focussed on HeLa cells found 9-15% of permeabilised nascent peptides were from cell nuclei (Iborra *et al*., 2001). Yet two years later another study cast doubt on these results and estimated that if protein synthesis did take place within the nuclear compartment it would be limited to around 1% of total protein synthesis within the whole cell (Nathanson *et al*., 2003). The later study put the previous studies results down to over-permeabilisation which would have caused damage to cells and led to cytoplasmic contamination.

1.5.2 Ribopuromycylation

A more recent study used a technique known as ribopuromycylation (RPM) to detect nascent peptides still attached to translating ribosomes in the nuclear compartment of HeLa cells. David *et al.* (2002) developed the RPM technique which uses puromycin, an antibiotic which is known to inhibit translation (Nathans, 1964; Zamir *et al*., 1966), along with translation elongation inhibitors emetine and

cycloheximide which stop the nascent peptide being released from the ribosome. RPM then allows direct visualisation of translating ribosomes using puromycin specific monoclonal antibodies in fixed and permeabilised cells (Eggers *et al*., 1997). Using this technique, translating ribosomes were detected mainly in the cytoplasm, but also in the nucleolus (David *et al.*, 2012).

Older studies pass doubt on the effectiveness of cyclohexamide and emetine and their ability of preventing the release of nascent peptides from the ribosome (Baliga *et al.*, 1970; Grollman, 1968). Even if this is the case and the two translation elongation inhibitors are merely slowing down the nascent peptides release, the puromycylated peptides would still remain close to sites of translation due to the brief periods of incubation with the translation elongation inhibitors (McLeod *et al.,* 2014).

1.5.3 Bimolecular fluorescence complementation

A second study used a different technique which has also detected nuclear translation in *Drosophila melanogaster* cells (Al-Jubran *et al.,* 2013). Bimolecular fluorescence complementation (BiFC) uses two complementary halves of YFP which are attached to two proteins and when the proteins interact the complementary halves come together and produce a YFP fluorescent signal (Hu *et al*., 2002).

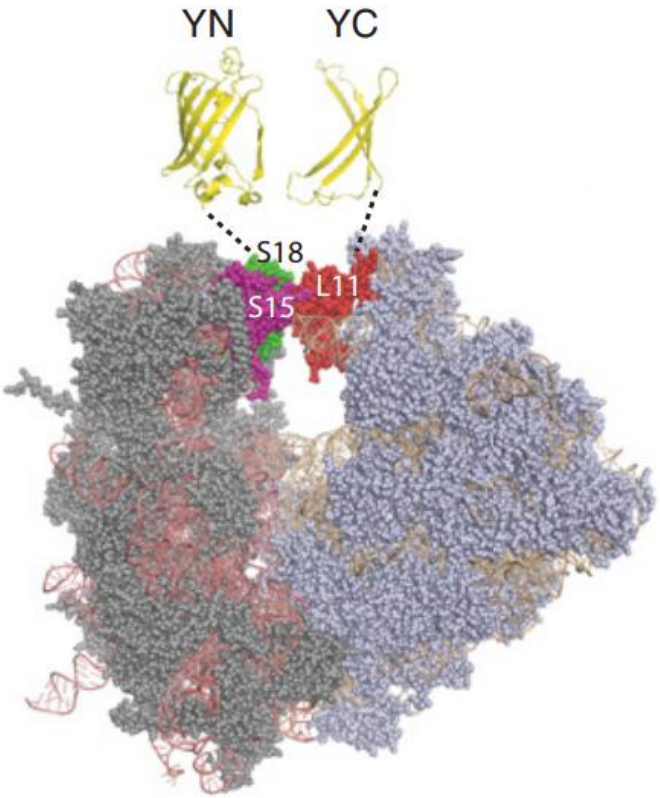


Figure 1. BiFC: Two adjacent tagged RPs S11 and L11 on the 40S (left) and 60S (right) express YFP signal when they come together and join YN and YC, the two complementary halves of YFP, during 80S formation [Image retrieved and edited from Al- Jubran *et al*. (2013)].

In the study by Al-Jubran *et al*. (2013), the YFP complementary halves were tethered to two adjacent proteins on the 40S and 60S ribosomal subunits. When the two subunits came together to form the 80S ribosome, it was found that the two tagged proteins which were adjacent to each other on the 80S ribosome would produce YFP signal, but proteins which were further apart would not. For example RpS18 and RpL11 are adjacent and produce a signal whereas RpS11 and RpL32, which are further apart from each other, produced no signal. Using this method of tagging RPs which are adjacent on the two ribosomal subunits, YFP signal was found mostly within in the cytoplasm but was also present in the nucleus and nucleolus (Al-Jubran *et al*., 2013). This result indicated

that 80S ribosomes, which are the hallmark of translation, initiation, elongation and termination (Jackson *et al*., 2010) are present within the nucleus in both cell culture and live fly tissue.

Translation elongation inhibitors such as emetine were found to increase YFP signal whereas translation initiation inhibitors such a pactamycin plus other inhibitors which cause the 80S to dissociate decreased signal. Signal within the nucleus was most strong within the nucleolus and high concentrations of actinomycin D, a transcription inhibitor, produced little or no signal. Lower concentrations of actinomycin D, caused signal loss within the inner core of the nucleolus, signifying that there may be an abundance of mRNAs within the nucleolus and the signal produced indicates translating 80S ribosomes (Al-Jubran *et al*., 2013).

In the study by Al-Jubran *et al.* (2013), puromycin incorporation was also tested and results found puromycin incorporation within the nucleolus which was consistent with results from David *et al*. Puromycin incorporation was identified at specific transcription sites on the *Drosophila melanogaster* polytene chromosomes signifying that ribosomal subunits may associate with nascent transcripts. This data also supports previous findings where RPs, rRNA and some transcription factors were found at transcription sites on polytene chromosomes (Brogna *et al*., 2002; Rugjee *et al*., 2013).

1.6 Research project aims

The aim of this research project was to assess whether RpL41 and other ribosomal proteins are differentially expressed in *Drosophila melanogaster*. This project focussed on characterising RpL41-YFP expression within the *Drosophila melanogaster* midgut.

The aims of this research project were:

- To characterise *Drosophila melanogaster* transgenes expressing RpL41-YFP along with three other YFP fusion tagged ribosomal proteins: $RpL10_{AB}$-YFP, RpS9 034-YFP, RpS9 504-YFP and a eukaryotic initiation factor: eIF-1A-YFP.
- To compare RpL41-YFP expression in salivary gland cells, fat body cells and midgut cells along with the other tagged proteins.
- To study RpL41 distribution throughout the midgut using anti-RpL41 antibodies.
- To study RpL41-YFP expression in the midgut during larval development.

2.0 Materials and methods

2.1 *Drosophila melanogaster* fly lines and transgenes

Five YFP fusion protein expression fly lines were purchased from the DGRC, Kyoto Stock Centre. Flyprot (http://www.flyprot.org/) developed the fly lines using Cambridge Protein Trap Insertions which used a PiggyBac transposable element to insert the YFP gene into introns of the endogenous *Drosophila melanogaster* protein genes. Details of each fly line and their corresponding constructs can be seen in table 1 and diagram representations of the genes containing the YFP gene insert can be viewed in Figure 2.

Affected Gene	Inserted Element	Cambridge Protein Trap Line	DGRC Number	Annotation ID	Chromosome Insert Location	Genomic Location of YFP Insert
RpL41	PBac{754.P.SVS-0}	CPTI002881	115344	CG30425	2R (60E5)	20,791,354 [+]
RpL10$_{AB}$	PBac{802.P.SVS-2}	CPTI003957	115463	CG7283	3L (68E1)	11816014 [+]
RpS9 034	PBac{602.P.SVS-1}	CPTI000493	115034	CG3395	3L (67B11)	9503969 [+]
RpS9 504	PBac{602.P.SVS-1}	CPTI000785	115504	CG3395	3L (67B11)	9504066 [+]
eIF-1A	PBac{754.P.SVS-0}	CPTI002562	115303	CG8053	3R (91A1)	14103733 [-]

Table 1. This table displays the *Drosophila melanogaster* genes which were used in the project and contain an endogenous YFP insert. Details of insert location, the transposable elements used for each gene and the fly line numbers for their respective source are also shown.

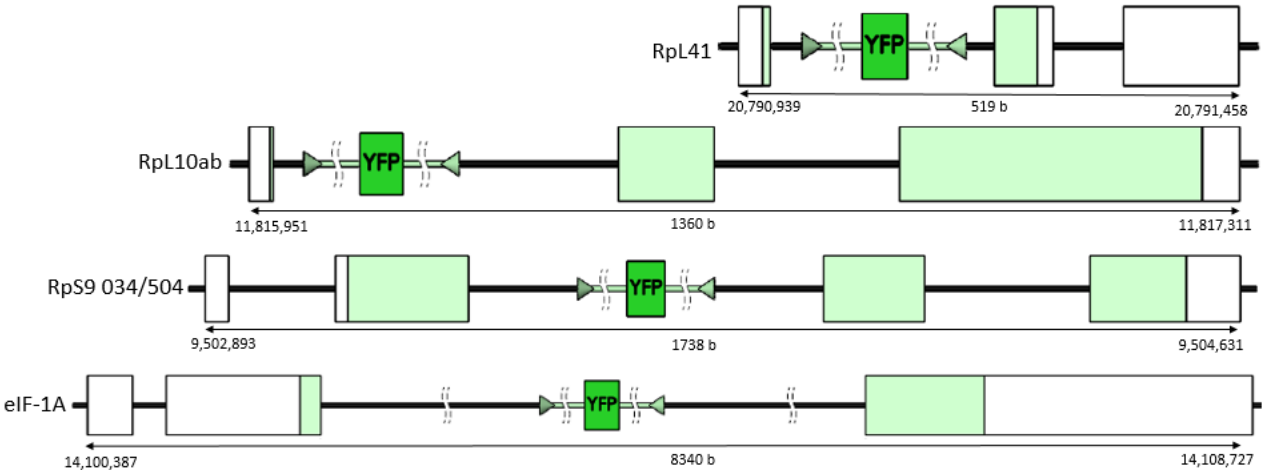


Figure 2. A diagram representation of the five *Drosophila melanogaster* protein transgenes used in this project; RpL41, $RpL10_{AB}$, RpS9 034, RpS9 504 and eIF-1A which are endogenously tagged in their introns with a YFP gene. Boxes represent exons with coloured areas representing the CDS. Black lines represent introns. The YFP insert contains a splice acceptor and splice donor site and wavy lines represent genes of the transposable element which were not spliced into the gene. Gene locations and sizes are also displayed.

2.2 Fly stock handling and food

Fly stocks containing the endogenous YFP gene were grown in glass vials at 18°C in order to grow larger third instar wandering larvae for dissection.

All flies were grown in vials containing standard fly food mix made up of yeast, agar, sugar, nipagin and cornmeal.

2.3 Midgut dissection

Third instar wandering larvae were collected, washed briefly in water and transferred to a sucrose solution (0.5M) and left at room temperature for two hours in order to remove yeast present inside the guts so as to reduce auto-fluorescence. The larvae were then transferred to water and kept on ice for ten minutes to inhibit movement to allow for easier dissection. Midguts were dissected individually in glass wells containing 20 µL cold 1X PBS and

immediately transferred to a 1 mL Eppendorf tube containing 200 µL fixation solution (4 % HCHO in 1X PBS) and kept on ice. Once five midguts had been dissected the fixation solution was removed from the tube by pipetting and replaced with 200 µL fresh fixation solution and incubated on ice for 30 mins. Following fixation the midguts were rinsed with 200 µL washing solution (1X PBS/ 0.3% Triton v/v) twice before washing twice for ten minutes using the same solution at room temperature. To stain the guts were then incubated in 200 µL 1X PBS/ 0.3% Triton v/v and DAPI (1:1000) (100 µg/mL) for 30 mins at room temperature. The guts were then rinsed and washed (10 mins) in 1X PBS/ 0.3% Triton v/v at room temperature. The midguts were then mounted on a slide under the microscope in 1X PBS/ 0.3% Triton v/v and arranged. Excess 1X PBS/ 0.3% Triton v/v was then soaked up using tissue paper (Whatman, Sigma Aldrich) and the midguts immediately overlaid with 80% glycerol in 1x PBS. On either side of the guts a cover glass was placed and another was placed on top of the guts directly so as not to squash them. Slides were sealed and stored at 4 °C.

2.4 RpL41 immunostain

RpL41-YFP third instar wandering larvae were prepared for dissection in the same way as before in sucrose solution and dissected, fixed and washed also in the same manner. After washing the guts were blocked in 200 µL blocking solution (10% fetal bovine serum and 1X PBS/ 0.3% Triton v/v) for 1 hour at room temperature. The guts were then incubated with the anti-RpL41 antibody (primary antibody) (1:500 dilution) in 200 µL blocking solution overnight at 4 °C in a humid chamber (sealed plastic lunchbox containing wet tissue paper). The guts were then rinsed and washed (10 mins) twice in 200 µL washing solution (1X PBS/

0.3% Triton v/v) at room temperature and then incubated with the secondary antibody (Cy3 Conjugated Affinipure goat anti-rabbit IgG H+L) (1:250) in 200 µL blocking solution for 3 hours at room temperature. The rinse and wash step was repeated before the guts were stained with DAPI (1:1000) in 200µL washing solution for 20 mins at room temperature. The guts were rinsed and then washed in washing solution before mounting on to a slide as done in the previous section.

2.5 Microscopy

Slides were analysed using a Nikon Eclipse Ti epifluorescence microscope which was equipped with an ORCA-R2 camera (Hamamatsu Photonics) to take images of the midguts. A Leica SP2-A0BS confocal microscope was used to take higher quality images of slides which required further analysis.

2.6 PCR

Primers were designed in order to carry out PCR and confirm the location of the YFP gene inserts in the five endogenously tagged proteins. Forward primers were designed to start at the 5’ Terminal repeat of the P-element constructs and the reverse primers were designed to go from the 3’ UTR on the gene itself. Table 2 lists the primers used in this project.

Gene	Reverse Primer	Forward Primer
RpL41	CTGCGCTTACTTGGACCTTG	CTAAATGCACAGCGACGGAT
RpL10Ab	CTAGTAAAGACGCTGGGGC	ATGCACAGCGACGGATTC
RpS9 034	AGTCTGCGCCATGTATTTGTC	CGGATTCGCGCTATTTAGAAAG
RpS9 504	AGTCTGCGCCATGTATTTGTC	CGGATTCGCGCTATTTAGAAAG
eIF-1A	ACATCGACTACTACGTGCGA	CTAAATGCACAGCGACGGAT

Table 2. A list of the forward and reverse primers used for each transgene to carry out PCR.

Genomic DNA was extracted from the five YFP tagged fly lines using 'squish' buffer (1M Tris pH 8.0, 0.5M EDTA, 5M NaCl). A single male fly from each line was frozen at -20 °C for 30 mins in a 100 µL PCR tube. 50 µL 'squish' buffer was pipetted up, the tip was used to mash the fly before the buffer was ejected and grinded further against the tube wall. Samples were then placed in a thermo cycler and ran at the following parameters: 37 °C for 30 mins (digestion), 95 °C for 2 mins (heat kill enzyme) and then 4 °C for ∞ (storage). Samples were centrifuged at 16,000 g for 7 mins at 4 °C and the supernatant (genomic DNA) was transferred to a fresh tube and stored at -20 °C until required.

PCR amplification was carried out in 50 µL reactions using Phusion® high fidelity DNA polymerase and reaction specifics can be viewed in table 3. The PCR reaction was carried out in a thermal cycler under the following parameters: 98 °C for 2 mins for the initial denaturation followed by 32 cycles of 98 °C for 10 mins, 58 °C for 30 s and 72 °C for 1 min before a final extension of 10 mins at 72 °C.

PCR products were analysed by electrophoresis using a 1% agarose gel stained

with 2 µL ethidium bromide. PCR products were ran next to a 1 kb DNA ladder at 100 V for approximately 30 mins. The gel was analysed under UV light (Genesnap, Syngene) and images taken.

Component	**Amount added for 50 µL reaction:**
5x Phusion HF Buffer	10 µL
10 mM dNTPd	0.4 µL
10 µM Forward primer	2.5 µL
10 µM Reverse primer	2.5 µL
Genomic DNA template	2.5 µL
Phusion DNA Polymerase	2.5 µL
Nuclease-free water	31.6 µL

Table 3. A list of all of the components and the amounts used for PCR reactions.

2.7 DNA sequencing

100 µL PCR reactions were carried out following the same PCR protocol as before for each fly line genomic DNA and analysed on a 1% agarose gel. Bands of DNA were excised from the gel under UV light and transferred into a 1 mL Eppendorf tube. DNA was extracted and purified using a Quiagen Gel extraction Kit following manufacturer's instructions. The concentration of the DNA was measured using a Nanodrop and PCR products confirmed on a 1% agarose gel. 5 µL of each

purified PCR product was mixed with 5 µL of the corresponding forward primer were then sent to GATC Biotech (London) for sequencing.

2.8 Western blot

20 fly heads were collected on ice for each fly line and homogenised in 30 µL 2X SDS sample buffer (1 M Tris, 50% glycerol, 10% SDS, 0.5% bromophenol blue and 0.5% β-mercaptoethanol) using a pestle and mortar on ice to extract protein. The sample was spun down briefly in a centrifuge and squashed again. This was repeated twice more before the samples were boiled for 5 mins at 100 °C and then transferred directly onto ice for 1 min before being centrifuged at 16,000 g for 2 mins to collect the condensation.

20 µL samples from each protein extraction were loaded onto a 12% SDS-PAGE gel next to a 1 kDa protein ladder (Bio-rad). The gel was ran at 80 V for the first 10 mins and was then increased to 100 V for 60 mins. The protein was transferred onto a 0.2 µm nitrocellulose membrane (Life technologies) at 350 mA for 1.5 hours at 4 °C in 1X transfer buffer (25 mM Tris, 192 mM Glycine, 10% methanol). Transfer was verified by staining the membrane with Ponceau S (Sigma Aldrich) for 5 mins with rocking before being rinsed off with ddH_2O.

The membrane was incubated with blocking solution (0.5g dry milk power in 10 mL TBST) for 1 hour at room temperature with rocking before incubation with an anti-GFP primary antibody (AbD serotec) (1:1000) in 5 mL TBST overnight at 4 °C. The next day the membrane was washed (2x quick and 3x 10 mins) in 10 mL TBST before incubation with the secondary antibody (Look up bottle) (1:10,000) in 10 mL TBST for 1 hour at room temperature with rocking before being washed again as

before (2x quick and 3x10 mins). The membrane was incubated in 1 mL ECL solution (luminal and pico stable peroxide, Thermo Scientific) for 5 mins at room temperature and then exposed for 2 hours under UV light (GeneSnap, SynGene).

3.0 Results

3.1 *Drosophila melanogaster* salivary gland cell transgene expression

The salivary glands in *Drosophila melanogaster* are made of two cell types. Columnar secretory cells are large and make up the epithelium of the salivary glands and synthesise and secrete large amounts of 'glue' protein at the end of third instar in order to stick to a surface and aid pupal development (Andrew *et al*., 2000). These cells also have large polytenic chromosomes and have been used to visualise proteins within the nucleus (Rugjee *et al*., 2013). The second type of cell which makes up the salivary glands are the duct cells which make up the tubes that connect the glands to the larval mouth.

Expression of the transgenes RpL41-YFP, $RpL10_{AB}$-YFP, RpS9 034-YFP, RpS9 504- YFP and eIF-1A-YFP were compared in *Drosophila melanogaster* salivary gland cells. Microscopy images of RpL41-YFP expression displayed signal in the cytoplasm of cells and in the nucleus strong YFP signal was observed (Figure 3, A). For $RpL10_{AB}$- YFP signal was present throughout the cytoplasm and there was no concentrated signal present in the nucleus although signal did not decrease around the area of the nucleus (Figure 3, B). Signal for RpS9 034-YFP was present throughout the cells cytoplasm but signal was observed to decrease in intensity in the nucleus (Figure 3, C). For RpS9 504-YFP no YFP signal was observed in any of the cells (Figure 3, D). eIF-1A-YFP expression produced YFP signal throughout the cytoplasm and was observed to be stronger in the nucleus (Figure 3, E).

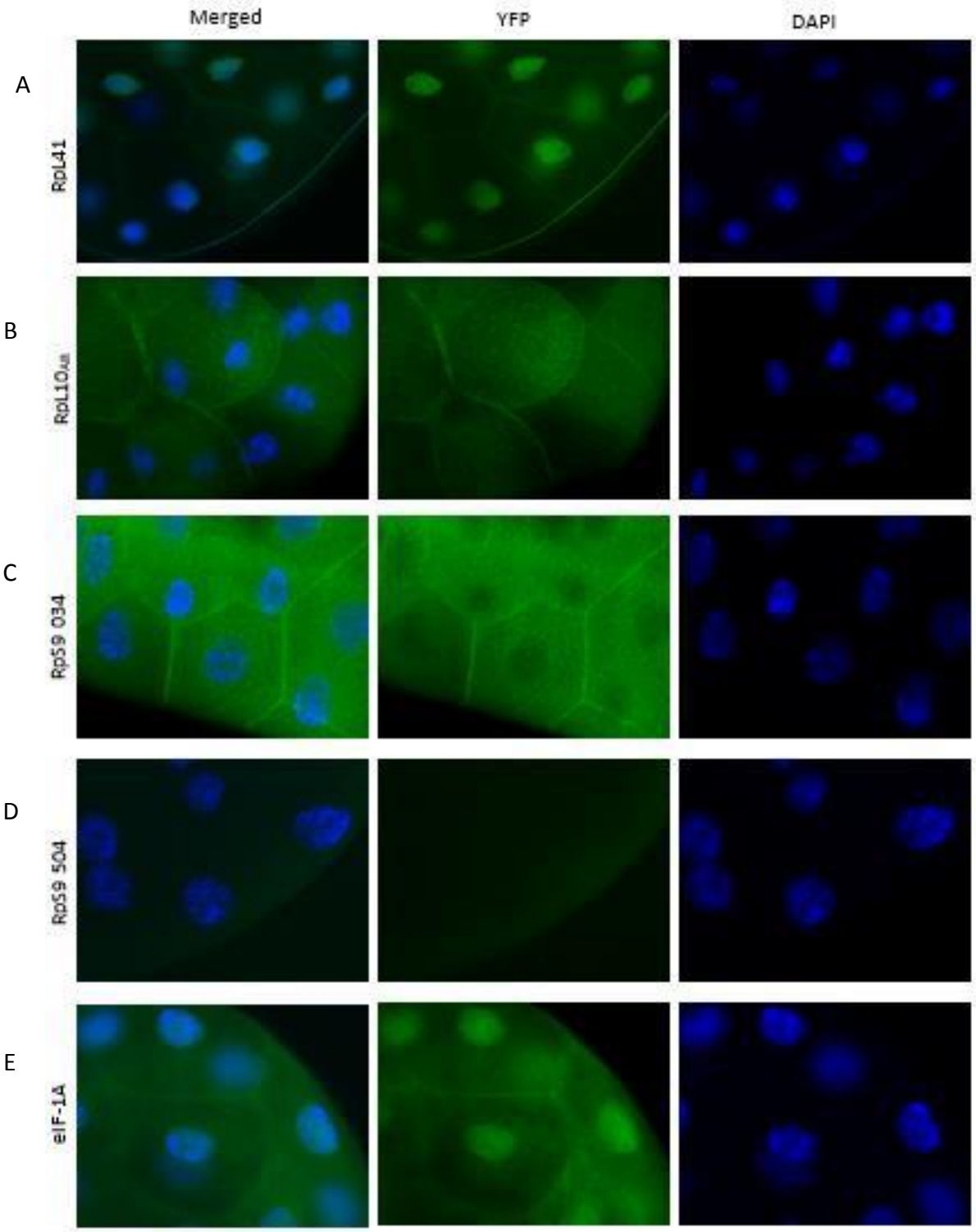


Figure 3. Endogenous expression of the YFP tagged RpL41, $RpL10_{AB}$, RpS9 304, RpS9 504 and eIF-1A genes under native promoter control in *D. melanogaster* salivary gland cells. Images on the right display a magnified view of salivary gland cells with YFP fluorescence and DAPI staining. Images in the middle show YFP signal in salivary gland cells, and images on the right display DAPI staining. Images were taken on an epi-fluorescence microscope using a 40x objective.

3.2 *Drosophila melanogaster* fat body cell transgene expression

Fat body cells make up a single cell thick tissue and are composed of mostly lipids. The cells accumulate nutrients in the last three days of larval development in order to have enough energy stored for pupal development, a period in which no feeding takes place (Aguila *et al*., 2007).

Expression of the transgenes was compared in fat body cells. Microscopy images were used to analyse RpL41-YFP expression and found signal was present throughout the cytoplasm of cells. There was a stronger YFP signal observed within the nucleus of some but not all of the cells (Figure 4, A). Expression of $RpL10_{AB}$-YFP could be seen in the area of the cytoplasm surrounding the nucleus and there was little to no signal at the outer areas of the cells cytoplasm. Signal was also observed within the nucleus of about equal intensity in comparison to the cytoplasm (Figure 4, B). Signal for the transgene RpS9 034-YFP was similar to that of $RpL10_{AB}$-YFP in that most signal was present in the area surrounding the nucleus and was patchy. In the nucleus however there was a clear decrease in signal strength in comparison to the immediate area of cytoplasm surrounding the nucleus (Figure 4, C). Signal for RpS9 504-YFP expression was very weak in comparison to the other transgenes which was most likely auto fluorescence due to the large amount of lipids present in the cells (Figure4, D). Expression of eIF-1A-YFP in fat body cells was observed throughout the cytoplasm and the signal was more concentrated in the nucleus than the cytoplasm (Figure 4, E).

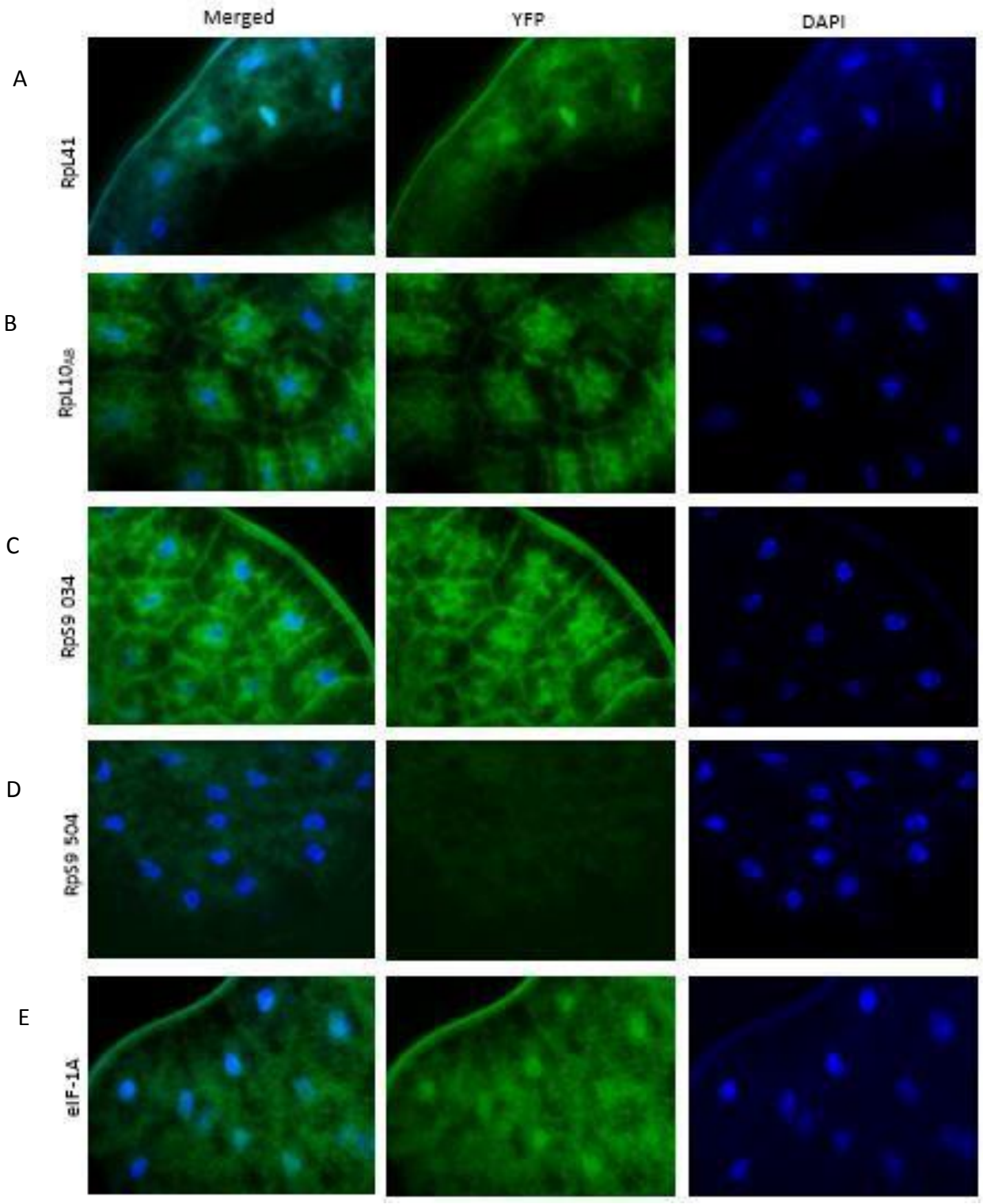


Figure 4. Endogenous expression of the YFP tagged RpL41, $RpL10_{AB}$, RpS9 304, RpS9 504 and eIF-1A genes under native promoter control in *D. melanogaster* fat body cells. Images on the right display a magnified view of fat body cells with YFP fluorescence and DAPI staining. Images in the middle show YFP signal in fat body cells, and images on the right display DAPI staining. Images were taken on an epi-fluorescence microscope using a 40x objective.

3.3 *Drosophila melanogaster* midgut transgene expression

The *Drosophila melanogaster* midgut has a complex organ structure and has similarities to that of mammalian intestines. The adult fly midgut is derived from the endoderm and is made up one epithelium layer surrounded by two layers of visceral muscle, nerves and trachea with a peritrophic membrane which separates the epithelium from food inside of the gut (Jiang and Edgar, 2009).

The adult midgut epithelium is made up of three cell types; absorptive enterocytes (ECs), secretive enteroendocrine cells (EEs) and intestinal stem cells (ISCs). It is the ISCs which maintain gut homeostasis by differentiating to replace lost or damaged ECs and EEs (Mathur *et al*., 2010). ISCs give rise to two cell types; an immature daughter cell called an enteroblast (EB) and a self-renewed stem cell. The EB exits the cell cycle and differentiates into an EC or an EE which is driven by Notch signalling. Normally 90% EBs become ECs (Ohlstein and Spradling, 2006).

Drosophila melanogaster goes through a number of stages of development before becoming an adult: from an embryo it follows through three larval stages (first, second and third instar) before entering a pupal stage where it metamorphoses into an adult. The midgut at the larval stage is made up of two cell types; larval enterocytes and adult midgut progenitor (AMP) stem cells, also known as midgut imaginal discs (Hartenstein *et al*., 1992; Jiang and Edgar, 2009). Notch and also EGRF signalling drives AMP cell development and during later stages of larval development AMP cells which are scattered move from the apical to the basal layer of the epithelium where they form islet clusters (Jiang and Edgar, 2009; Mathur *et al*., 2010). During larval development the ECs grow in size and ploidy (Lamb, 1982).

During the pupal stage the ECs degenerate and the AMP cell islets merge to form a new epithelial tube. ECs are born in the second half of pupal development from differentiating AMP cells (Takashima *et al*., 2011).

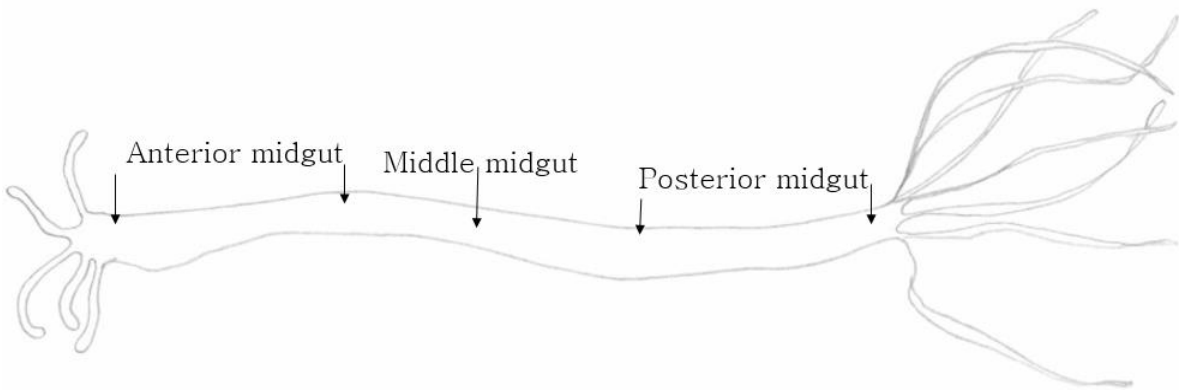


Figure 5. The *Drosophila melanogaster* larval midgut.

In the midgut, RpL41-YFP was observed by microscopy imaging and found to only be expressed within a specific region in the smaller cells nuclei as there is no YFP signal present in the larger cells nuclei or in the cytoplasm of either cell types (Figure 6, A). RpL10$_{AB}$-YFP was expressed throughout the cytoplasm of what appears to be the larger cells yet it is indistinguishable if it is expressed in the smaller cells cytoplasm also. In a region of the larger cells nucleus it is clear that in a small region where DAPI staining appears weaker to the rest of the nucleus RpL10$_{AB}$-YFP is expressed. There appears to be no expression within the smaller cells nuclei (Figure 6, B). YFP signal for RpS9 034-YFP expression was strong for what appears to be the cytoplasm of the smaller cells which are present in clusters. There is no apparent signal within the nucleus of the smaller cells. In the larger cells, there is no apparent localisation within the nucleus and localisation in the cytoplasm is weaker than for the smaller cells (Figure 6, C). RpS9 504-YFP expression gave very weak YFP signal (Figure 6, D). eIF-1A- YFP expression

was unexpected in that none of the cells had YFP signal, there are small spots of YFP signal within the gut, but this was more than likely due yeast not being successfully removed from the midgut prior to dissection (Figure 6, E).

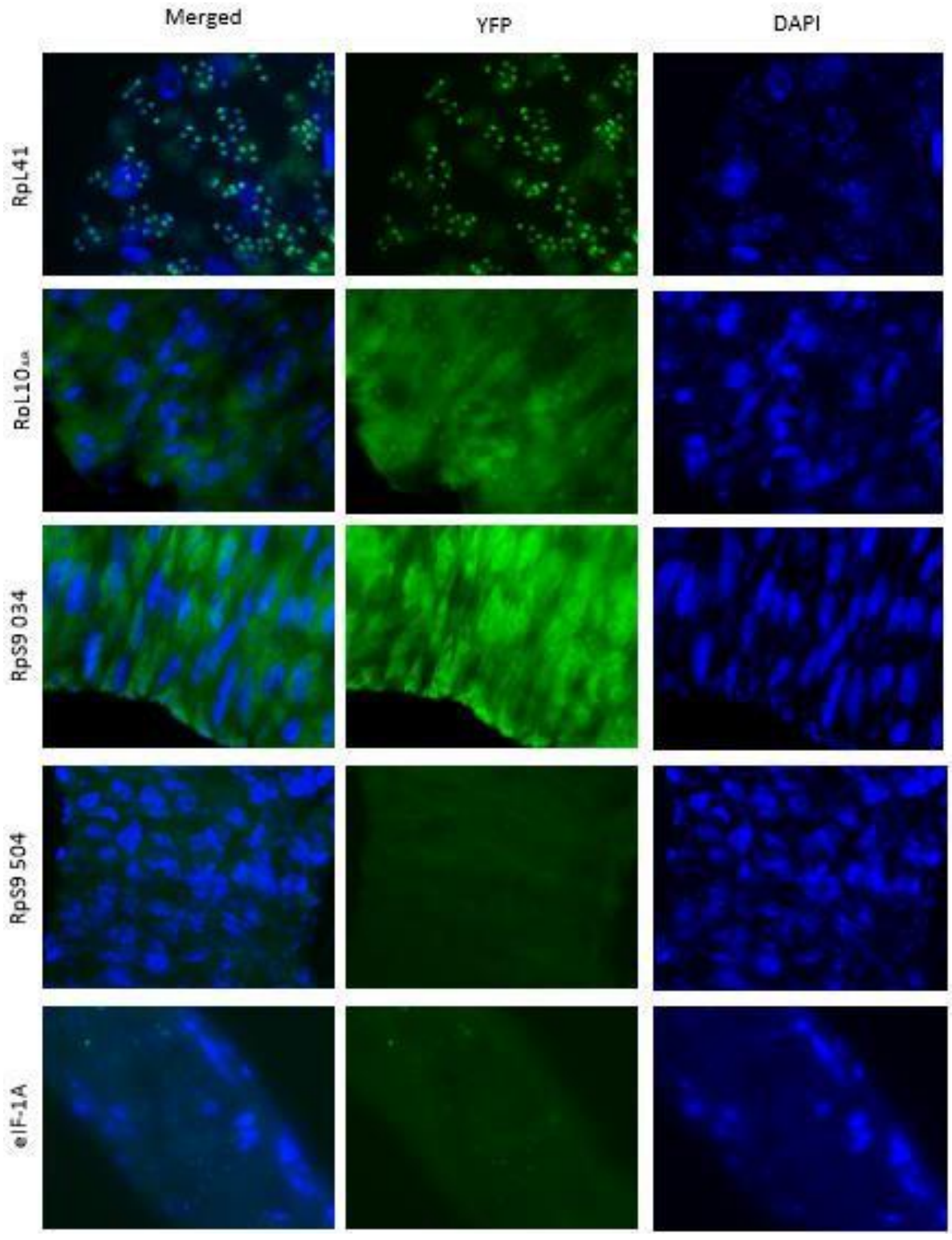


Figure 6. Endogenous expression of the YFP tagged RpL41, $RpL10_{AB}$, RpS9 304, RpS9 504 and eIF-1A genes under native promoter control in *D. melanogaster* midgut cells. Images on the right display a magnified view of a section of the midgut with YFP fluorescence and DAPI staining. Images in the middle show YFP signal in midgut cells and images on the right display DAPI staining. Images were taken on an epi-fluorescence microscope using a 40x objective.

3.4 Confirming YFP inserts using PCR

To confirm the endogenous YFP inserts were present and in the correct position for each of the transgenes, primers were designed in order to carry out PCR and compare actual fragment sizes to estimated fragment sizes. All of the forward primers were designed to start from the 5' terminal repeat of the YFP gene on each construct (Figure 7). The reverse primer for RpL41-YFP was designed at the 3' UTR on the 2nd intron (Figure 7, A) and the resulting PCR fragment size was estimated to be around 189 bases (Table 4). The reverse primer for the transgene $RpL10_{AB}$-YFP was designed at the 3' UTR on the third exon (Figure 7, B) and the fragment produced from PCR was estimated to be 1393 bases (Table 4). The reverse primer for both RpS9 034-YFP and 504-YFP was designed from the 3' UTR on exon 4 (Figure 7, C) and the estimated PCR fragment size was calculated to be around 740 bases for RpS9 034-YFP and around 643 bases for RpS9 504-YFP (Table 4). The reverse primer for eIF-1A-YFP had to be designed in the second intron near to the insert due to the fact that the intron where the YFP gene is inserted is 6787 bases long and for optimal PCR a fragment of only a few hundred can be produced (Figure 7, D). The estimated PCR fragment size for eIF-1A-YFP was 413 bases (Table 4).

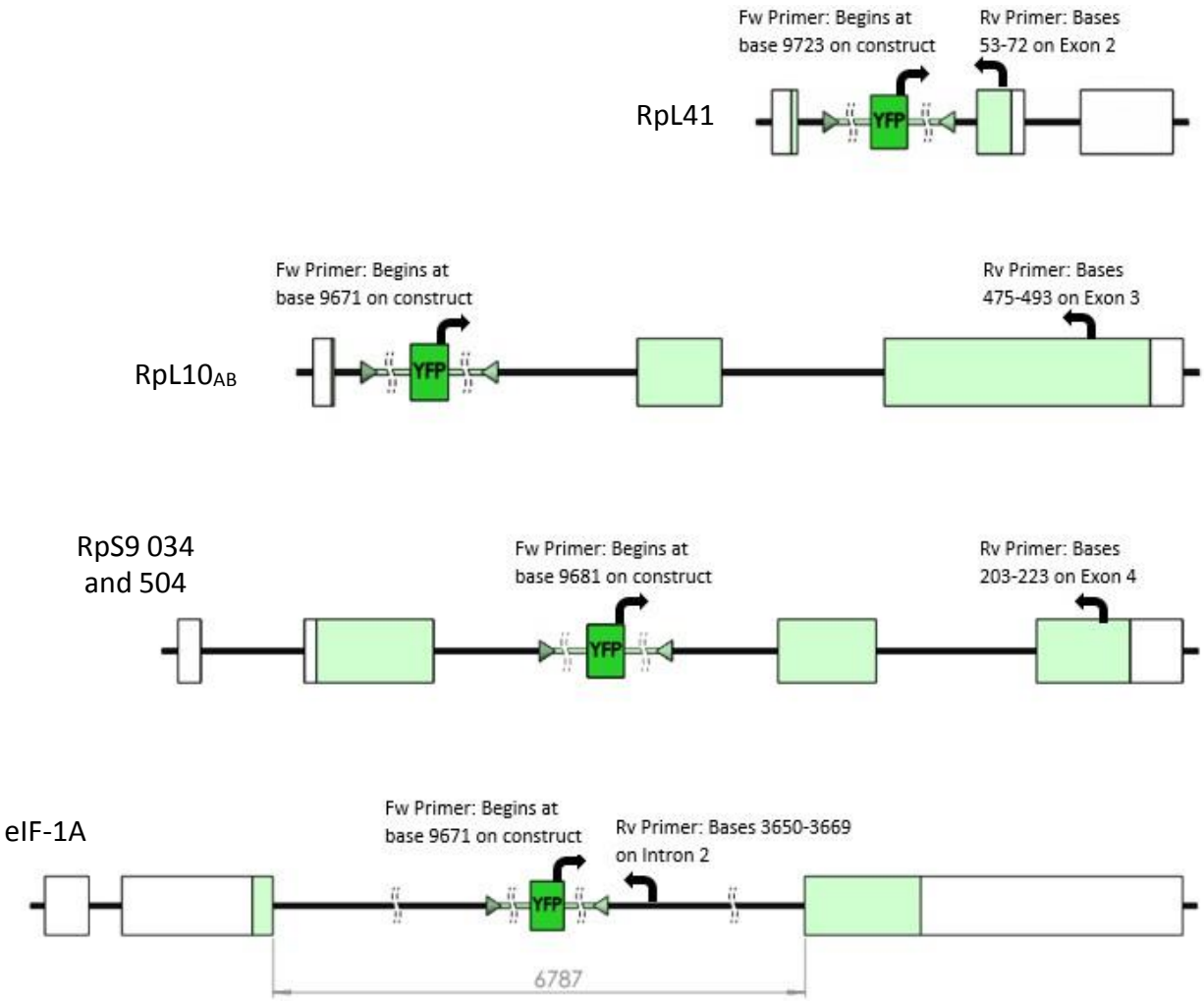


Figure 7. The four sets of forward and reverse primer locations for each of the five transgenes for PCR.

Transgene	Estimated PCR fragment size (bp)
RpL41-YFP	~189
$RpL10_{AB}$-YFP	~1393
RpS9 034-YFP	~740
RpS9 504-YFP	~643
eIF-1A-YFP	~413

Table 4. The estimated PCR product fragment sizes for each of the five transgenes.

PCR was carried out for each of the transgenic fly strains using the corresponding forward and reverse primers (figure 7). It can be seen in figure 8 (Lane 2) that the resultant RpL41-YFP fragment is the estimated size of around 189 bases confirming the YFP insert for this transgene is in the correct position. The $RpL10_{AB}$-YFP fragment produced was also of expected size of around 1393 bases confirming the YFP insert position (Figure 8, lane 3). The fragment produced for RpS9 034-YFP however was much larger than the estimated 740 bases and was more around the size of 1000 bases (Figure 8, lane 4). For RpS9 504-YFP, no PCR product was present (Figure 8, lane 6) showing that PCR was unsuccessful for this transgene. For this reason the fly strain expressing RpS9-504-YFP was removed from further experimentation in this project. The PCR fragment for eIF-1A-YFP was much larger than the estimated fragment size of 413 base pairs, the fragment was actually around 900 bases and not only was it not the expected size, the abundance of the fragments was less than for the other transgene fragments(Figure 8, lane 5).

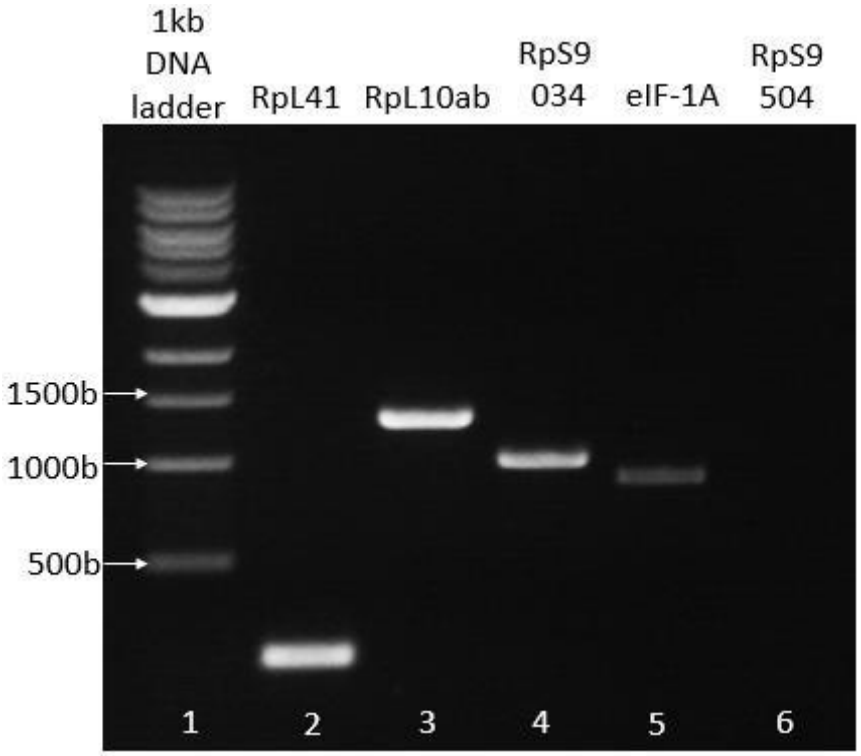


Figure 8. Gel electrophoresis using a 1% agarose gel of PCR amplification of RpL41-YFP (lane 2), $RpL10_{AB}$-YFP (lane 3), RpS9 034-YFP (lane 4), eIF-1A-YFP (lane 5) and RpS9 504-YFP (lane 6). Lane 1 shows a 1 kb DNA ladder.

3.5 Sequencing PCR products

To further attempt to confirm the YFP insert in the transgenes, the PCR fragments were purified and sent along with the corresponding forward primer to GATC, Biotech (London). The resultant sequences were BLAST searched and the insert was confirmed for RpL41 (figure 9), $RpL10_{AB}$ (figure 10) and unexpectedly for RpS9 034 (figure 11). The sequence for eIF-1A-YFP gave no matches on BLAST to eIF-1A.

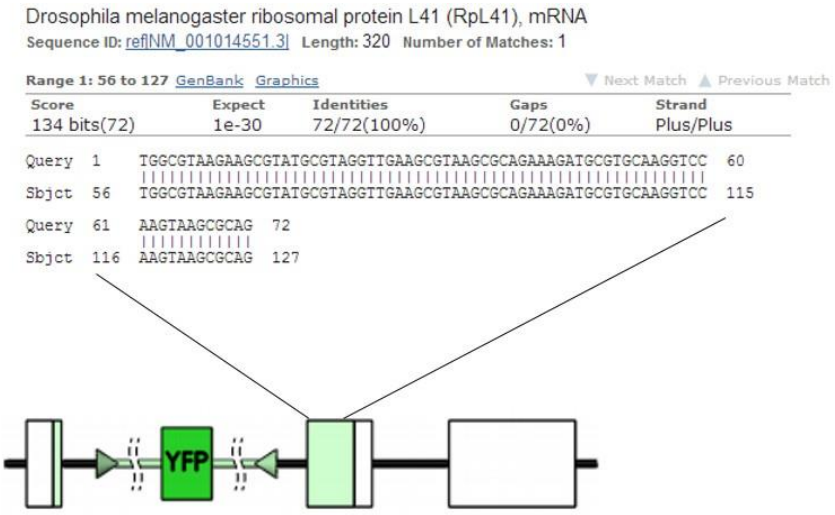


Figure 9. BLAST search results matched the sequenced fragment to an exon of the RpL41 gene. Coding exons of the gene are green in this diagram.



Figure 10. BLAST search results match the sequenced fragment to two exons and intron regions of the $RpL10_{AB}$ gene. Coding exons of the gene are green in this diagram.



Figure 11. BLAST search results match the sequenced fragment to exons of the RpS9 gene. Coding exons of the gene are green in this diagram.

3.6 Western blot

For further confirmation of the YFP insert in each of the transgenes, a western blot was carried out using an anti-GFP antibody. In the first Western blot (figure 12), only 20 fly heads were prepared per transgenic fly strain and a 1:2000 dilution primary antibody was used. No band was present for RpL41-YFP (figure 12, lane 2) whereas RpL10$_{AB}$- YFP had a dark band of expected size of the protein (24.28 kDa) plus the YFP protein (26.9 kDa) of 51.18 kDa (figure 12, lane 3). For RpS9 034-YFP there was also a band present at the expected

size of 49.53 kDa (RpS9 is 22.63 kDa plus YFP which is 26.9 kDa). There was no band present for eIF-1A-YFP which would have had an estimated size of 44 kDa. A second western was carried out to try and detect RpL41-YFP using 30 fly heads and the concentration of the primary antibody was increased to 1:1000 dilution. Indeed, a band was present at 30.3 kDa indicating the presence of RpL41- YFP (Figure 13, lane 2). RpL10$_{AB}$-YFP also had a strong band indicating the protein is present in abundance while the band for RpL41-YFP is much weaker (figure 13, lane 3).

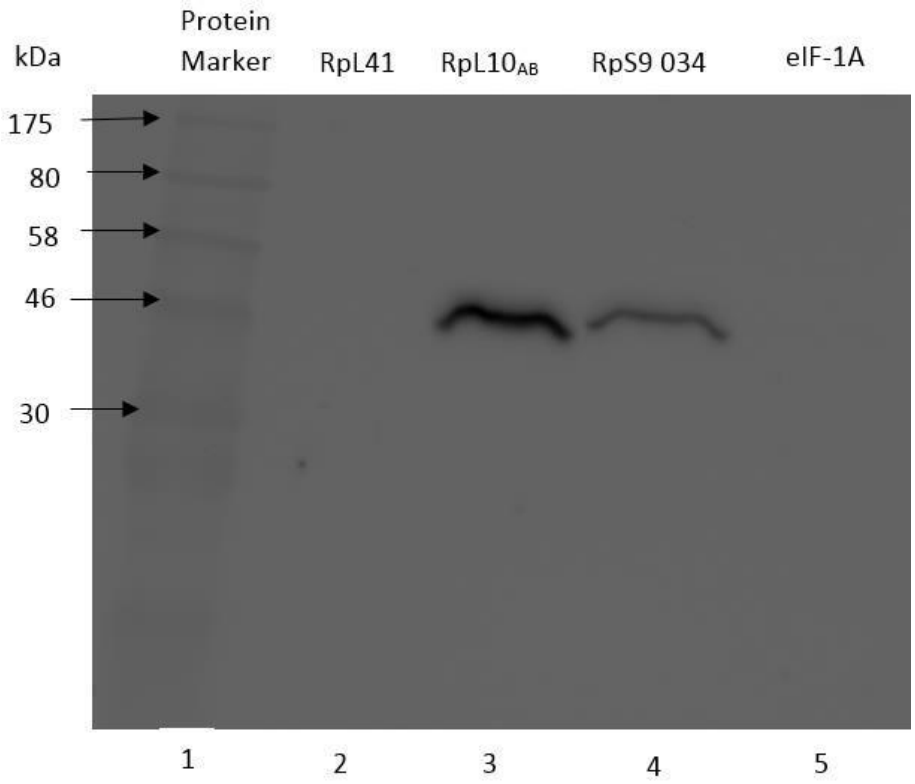


Figure 12. Western blot analysis confirms the YFP insert of RpL10$_{AB}$-YFP (lane 3) and RpS9 034-YFP (lane 4) using anti-GFP antibodies. The YFP insert for RpL41-YFP (lane 2) and eIF-1A-YFP (lane 5) could not be confirmed. Lane 1 shows a protein ladder.

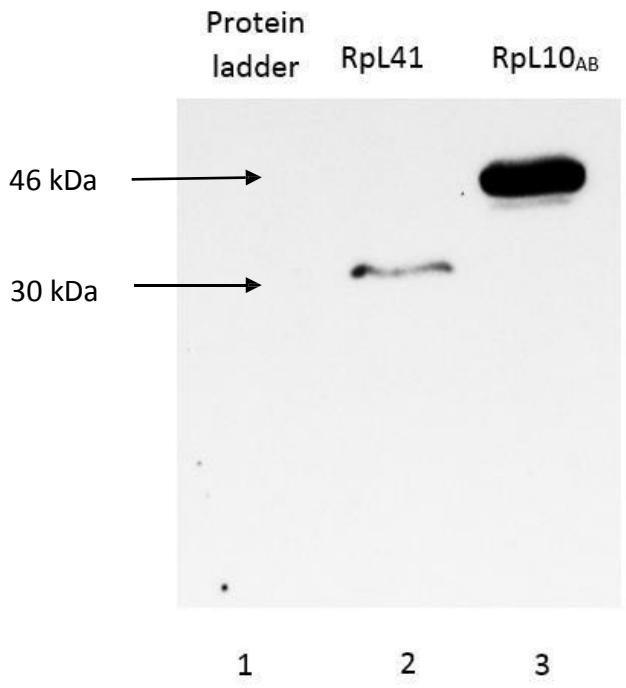


Figure 13. Western blot analysis confirmed the YFP insert of RpL41-YFP (lane 2) and $RpL10_{AB}$-YFP (lane 3). Lane one is a protein ladder.

3.7 Immunostain

Human anti-RpL41 antibodies were used to carry out an immunostain on third instar larval midguts for the fly strain expressing RpL41-YFP and stained using Cy3. Confocal imaging of a section of the midgut (figure 14) shows that in some but not all of the smaller cells where there is YFP signal, there is no Cy3 fluorescence present (blue arrow) but there is Cy3 signal in others (purple arrow) . Unexpectedly, in the larger cells, there is a small dot present of Cy3 fluorescence signal and no YFP signal (white arrow).

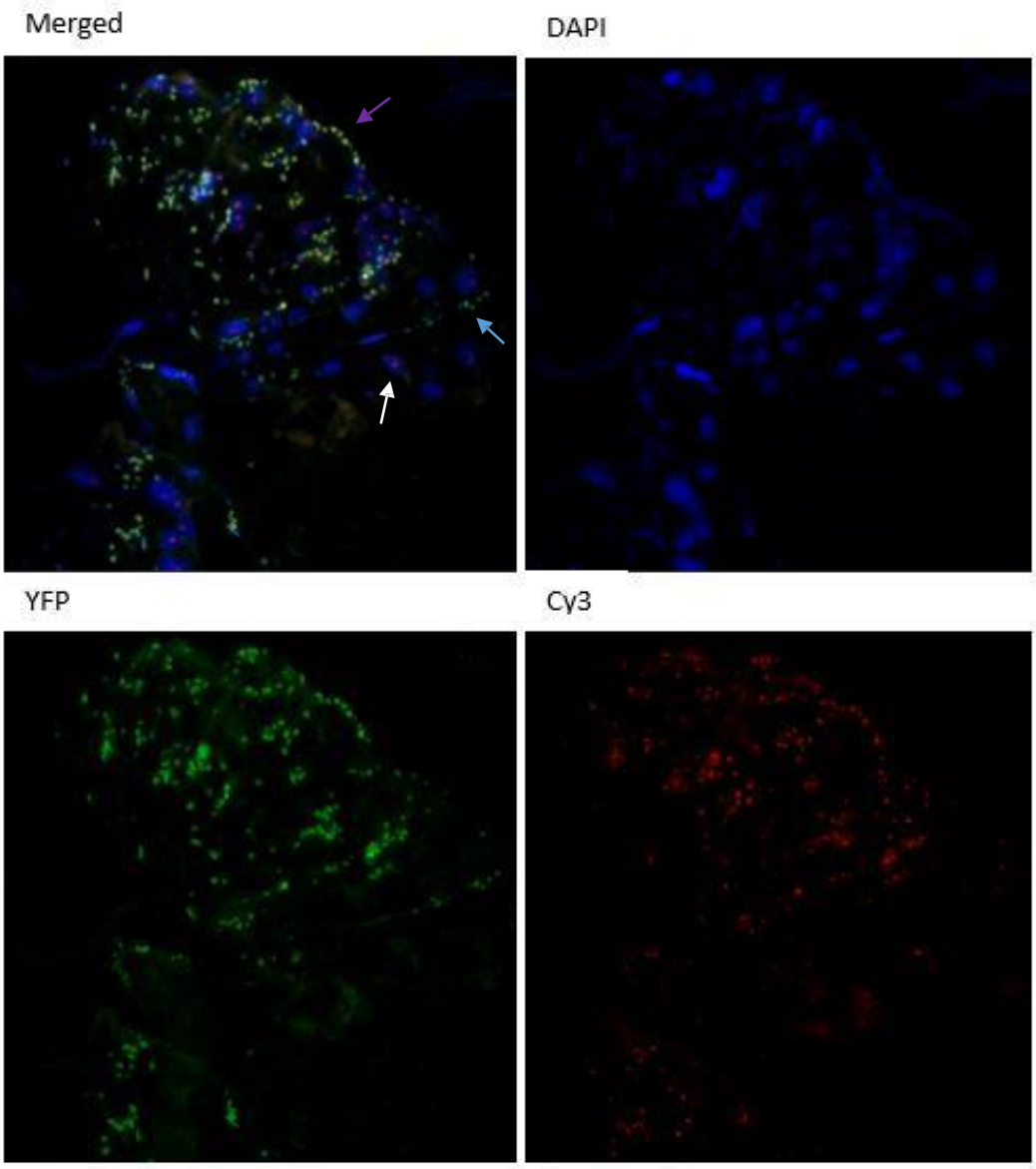


Figure 14. A confocal section of a third instar larval *D. melanogaster* midgut from the fly line expressing the transgene RpL41-YFP and immunostained using human anti-RpL41 antibodies. Top right image shows YFP flourescence, DAPI staining and Cy3 stain. The top left image shows only DAPI staining, bottom right shows YFP signal and bottom left shows Cy3 staining. This confocal section image was taken using a 40x oil immersion objective.

3.8 RpL41 expression during larval development

To assess the expression of RpL41-YFP during larval development, an experiment was set up to dissect the midgut from larvae at the different instars of the *Drosophila melanogaster* life cycle. Dissection of larva at early second instar (figure 15) was analysed by microscopy imaging and showed that all the cells nuclei were much smaller than third instar larva enterocyte nuclei and there appeared to be no smaller nuclei therefore it appears no AMP cells are present in the midgut. RpL41-YFP expression gave strong signal in the majority of the cells present in the midgut.

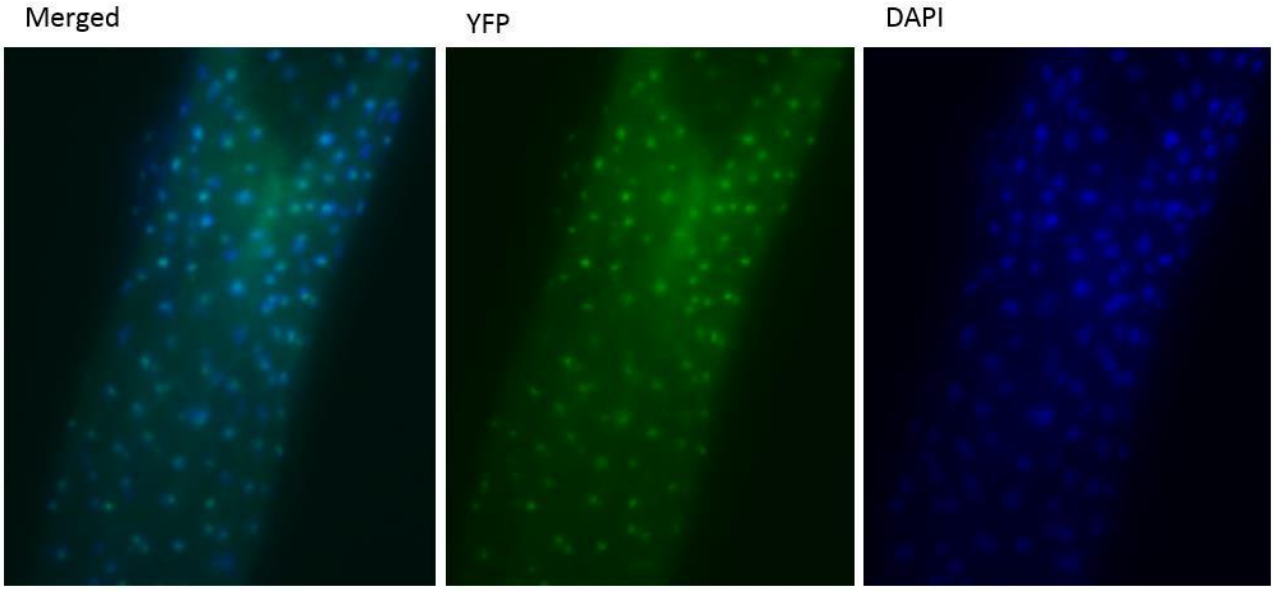


Figure 15. Early second instar larva midgut dissection to assess RpL41-YFP transgene expression. Images were taken using a 40x objective. On the left is an image of a section of the midgut and displays both YFP fluorescence and DAPI staining. In the middle only YFP signal is observed and on the right only DAPI staining.

Dissection was also carried out at early third instar (figure 16) and analysed by microscopy imaging. RpL41-YFP expression is strong in the nuclei of most of the larger cells. Smaller cell nuclei are also at this stage visible (white arrow) and it is clear there is no YFP signal at all in these cells while YFP signal appears stronger in some but not all of the larger cells.

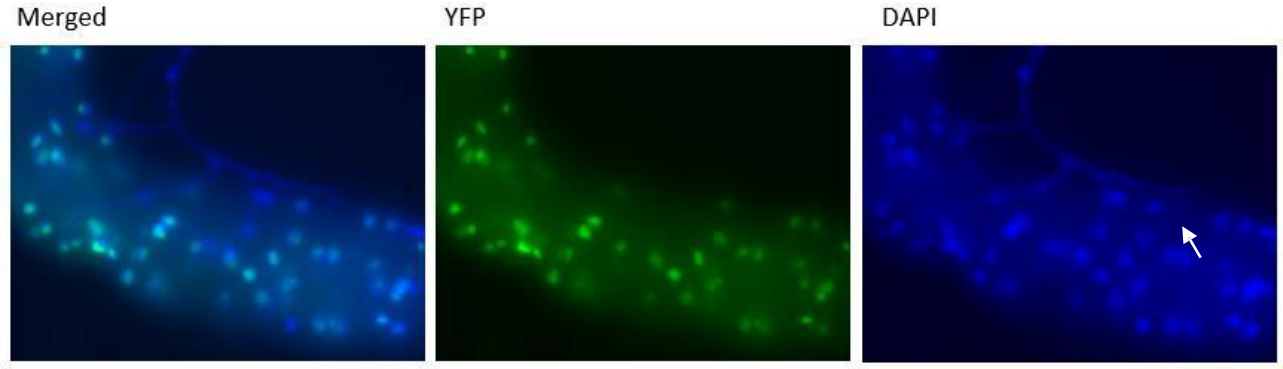


Figure 16. Early third instar larva midgut dissection of the fly strain expressing RpL41-YFP, taking using a 40x obective. On the left is an image of a section of the midgut and displays both YFP fluorescence and DAPI staining. In the middle only YFP signal is observed and on the right only DAPI staining.

4.0 Discussion

This project first aimed to characterise the expression of the *Drosophila melanogaster* transgene RpL41-YFP and compare its expression between salivary gland, fat body and midgut cells along with four other transgenes: $RpL10_{AB}$-YFP, RpS9 034-YFP, RpS9 504-YFP and eukaryotic initiation factor-1A-YFP (eIF-1A-YFP).

A previous paper had previously characterised RpL41-YFP expression within *Drosophila melanogaster* salivary gland cells and found the protein to be abundant in the cytoplasm and the nucleus with nucleolar enrichment (Rugjee *et al*., 2013). Similar expression was found in this study in the microscopy imaging of the salivary gland cells and expression was also the same in the fat body cells. There was however a clear difference in expression in the midgut cells. In the midgut, strong YFP signal was present only in a specific area of the nucleus, possibly the nucleolus, of the smaller cells present in the midgut. The RpL41-YFP protein is thought to be functional due to the fact that the tagged gene is homozygous viable (flyprot.org) and the genome of *Drosophila melanogaster* has just one gene that encodes RpL41 and is thought to be essential for viability (Marygold *et al*., 2007).

Although we cannot be sure that the smaller cells present within the midgut are AMP stem cells, their presence in clusters around the larger cells strongly indicate they are AMP cells which have formed the expected islets in late third instar (Jiang and Edgar, 2009).

RpL41 is a small protein which is made up of only 25 amino acids and crystal structure studies of the ribosome showed the protein to only associate with the 80S

ribosome at the interface where the 40S and 60S rRNAs make contact (Klinge *et al*., 2012).

The other transgenes used in this study were not homozygous viable and were in fact homozygous lethal. RpS9 034-YFP expression in salivary glands in comparison to a previous paper which tagged RpS9 with GFP on a 40S subunit was much less as signal was not enriched within the nucleus (Rugjee *et al*., 2013). This was most likely due to the fact that in the previous RpS9 was overexpressed and in this study expression of the transgene was endogenous. Another reason is that it could be due to the protein RpS9-034-YFP not being fully functional as it is homozygous lethal but it has previously been demonstrated that tagged RpS9 behaves in a fully functional manner (Al-Jubran *et al*., 2013). The RpS9 gene in *D. melanogaster* is present as a single copy (Marygold *et al*., 2007).

RpL10 can be expressed by two genes in *D. melanogaster*: $RpL10_{AA}$ and $RpL10_{AB}$. $RpL10_{AA}$ however lacks introns and it is thought the gene arose from retro transposition (Marygold *et al*., 2007). Expression of the transgene $RpL10_{AB}$-YFP was present throughout the cytoplasm and in the nucleus in all cell types indicating the protein is functional.

eIF-1A is a eukaryotic initiation factor which is known to transfer Met-tRNA to ribosomal 40S subunits in the absence of mRNA (Chaudhuri *et al*., 1997). eIF-1A-YFP appears functional as it is present in the cytoplasm and nucleus of salivary gland and fat body cells. In the midgut however, no YFP was found, this was an odd result as translation must be taking place in the cells throughout the midgut. The little YFP signal that is present is more than likely due to fluorescence signal interacting with yeast that was not successfully removed from the midgut prior to

dissection. The YFP insert for eIF- 1A could not be confirmed in PCR or sequencing which could have been due to the primers used not being specific enough. The Western blot however did not detect the YFP insert either. Although eIF-1A expression does appear in salivary gland cells and fat body cells, more work will be required to confirm the YFP insert.

By comparing all of the transgenes and their expression in the three different cell types it is clear that RpL41-YFP expression within the midgut is greatly different to expression in the other cell types. Also, all of the other transgenes (bar eIF-1A) had the same expression levels present throughout each cell type. This is interesting as it raises the question of why RpL41-YFP expression is required so strongly in the nuclei of AMP cells at this stage of larval development. It was discussed in detail in the introduction how some ribosomes can become specialised at specific periods in development or under certain conditions and it could be speculated that RpL41 is required for a specific period in larval development in AMP cells.

The question is also raised of whether RpL41 is present in the nucleus as a free protein with extra ribosomal functions, or as part of a ribosomal subunit. A few ribosomal proteins are already known to have extra ribosomal functions (Warner and McIntosh, 2009). RpL41 has been found to only associate the 80S ribosome (Klinge *et al*., 2012) which would link the protein to translation. Currently nuclear translation is hotly debated as to whether it takes place. If RpL41 is present in the nucleus it could be as a free protein (or part of a ribosomal subunit) carrying out extra ribosomal functions or it could also be speculated that it is actually integrated into a fully functioning ribosome carrying out translation within the nucleus. Al-Jubran *et al*. (2013) have previously reported nuclear translation as was

discussed in detail in the introduction and found some translation to be taking place within the nucleolus. RpL41-YFP expression in midgut AMP cells is present only within a specific area of the nuclei, which could be the nucleolus and it could be proposed that nuclear translation may be required in the nucleus of AMP cells in order for translation of proteins required at the late stages of larval development. In order to uncover the real role of RpL41 in the midgut AMP stem cells of third instar larvae, further experimentation will be necessary to characterise the protein and its expression.

To verify the transgenes, this project also set out to confirm the YFP inserts to confirm their locations on the corresponding protein gene there-by confirming that the transgene is being expressed at the correct location. PCR confirmed the YFP insert for only RpL41-YFP and $RpL10_{AB}$-YFP. The band for RpS9 034-YFP was around 1000 bases, and with the fragment estimated to be around 740 bases, 260 bases were unaccounted for. eIF-1A-YFP also produced a PCR product that was larger than expected, however this was most likely due to the primers not being specific enough. The reverse primer for this reaction was designed from the middle of an intron and not the 3' UTR of an exon as the others were due to the fact the gene was so long and it proved difficult to locate a primer that was specific enough to that gene and would also create a PCR product. RpS9 504-YFP produced no product in PCR, which backed up why there was a lack of YFP signal in microscopy imaging indicating the YFP insert was not present in the fly line. For this reason RpS9 504-YFP was removed from further analysis.

For further confirmation of the YFP inserts, PCR products of expected size were confirmed by sequencing for RpL41-YFP and also $RpL10_{AB}$-YFP. RpS9 034-YFP was also sequenced and BLAST search indicated that the insert was present

on the gene in the correct position.

A western blot using anti-GFP antibodies was carried out which further confirmed the YFP insert for RpL41-YFP, $RpL10_{AB}$-YFP and RpS9 034-YFP. YFP is a genetic mutant of GFP and only differs by a few bases so anti-GFP antibodies are still specific enough to detect the YFP on the transgenic proteins.

In an attempt to try and characterise RpL41-YFP expression within the larval midgut, anti-RpL41 antibodies were purchased in order to try and detect RpL41 expression within the transgenic fly line. The results of the immunostain were interesting and unexpected in that the confocal image showed that there was Cy3 fluorescence signal present in most, but not all of the small cells nuclei whereas YFP signal was present in all of them, but also Cy3 signal appeared in a small area of all of the larger cells nuclei towards the centre. To be able to interpret this result however it would first be sensible to confirm that the antibody used is specific enough. Only human anti-RpL41 was available at the time of the project which was purchased due to the fact its sequence only differs from the *Drosophila melanogaster* by four amino acids. In order to verify the antibodies specificity a western blot was attempted but was unsuccessful due to the fact RpL41 is such a small protein (3.40 kDa). A Western blot using a tris-tricine gel and buffer for the detection of small proteins was attempted which was also unsuccessful and due to time restraints of the project, optimisation of the protocol used was not possible.

To investigate how RpL41 is expressed during development, dissection was attempted at different time points in the *Drosophila melanogaster* larval life cycle. Microscopy images of early second instar larva indicate that RpL41-YFP is expressed in the larger cells nucleus and not at all in the smaller cells nuclei

although at this stage of larval development no smaller cells were observed in the midgut. In early third instar however, smaller cells nuclei were observed by DAPI staining, although not in abundance, and YFP signal was still only detected within the larger cells nuclei. This was an interesting result as initially due to ease of dissection only late third instar larva were dissected and by this point the YFP signal becomes very strong within the nucleus of only the small AMP cells so it was unexpected to find YFP signal in larger cell nuclei at an earlier stage in development. Due to difficulties in the dissection and fixing of the small guts, it proved difficult to dissect and fix some of the guts successfully and time restraints of the project meant a repeat was not possible to dissect the larvae at more stages of development. A repeat in the future will be able to further define how expression changes. These results do however indicate a change in expression during development in the midgut which may highlight a specific role for RpL41 during specific stages of development and this suggestion will need to be researched further.

4.1 Conclusion

Overall, this research project has indicated that in the midgut of *Drosophila melanogaster* the protein RpL41 appears to be differentially expressed in the nucleus of midgut stem cells. To investigate the role of RpL41 and if it has extra ribosomal functions or is involved in nuclear translation it will be important to further characterise RpL41-YFP expression. By defining the expression of RpL41-YFP during larval development to understand at which point the transgenic protein becomes concentrated in the nucleus and by also confirming the anti-RpL41 antibody is specific for *Drosophila melanogaster* RpL41 and RpL41-YFP results can then be interpreted with more confidence and will allow experimentation for investigating the role of RpL41 to continue with confidence.

It will also will be important to define specifically which cell type is expressing RpL41-YFP within the midgut in order to state specifically whether the YFP signal is present within small AMP cells or larger enterocytes and the use of Gal4/UAS genetic crosses will aid this. The study of RpL41 RNAi fly lines which are available from the Vienna Drosophila Stock Centre, will also allow further characterisation of the protein within the midgut and by silencing RpL41 the resultant phenotype can be assessed and a possible role for RpL41 hypothesised.

5.0 References

Aguila, J.R., Suszko, J., Gibbs, A.G., Hoshizaki, D.K., 2007. The role of larval fat cells in adult Drosophila melanogaster. J. Exp. Biol. 210, 956–963.

Al-Jubran, K., Wen, J., Abdullahi, A., Chaudhury, S.R., Li, M., Ramanathan, P., Matina, A., De, S., Piechocki, K., Rugjee, K.N., Brogna, S., 2013. Visualization of the joining of ribosomal subunits reveals the presence of 80S ribosomes in the nucleus. RNA 19, 1669–1683.

Allfrey, V.G., 1954. Amino acid incorporation by isolated thymus nuclei. I. The role of deoxyribonucleic acid in protein synthesis. Proc. Natl. Acad. Sci. U. S. A. 40, 881–885.

Andrew, D.J., Henderson, K.D., Seshaiah, P., 2000. Salivary gland development in Drosophila melanogaster. Mech. Dev. 92, 5–17.

Balasubramanian, S., Zheng, D., Liu, Y.-J., Fang, G., Frankish, A., Carriero, N., Robilotto, R., Cayting, P., Gerstein, M., 2009. Comparative analysis of processed ribosomal protein pseudogenes in four mammalian genomes. Genome Biol. 10, R2.

Baliga, B.S., Cohen, S.A., Munro, H.N., 1970. Effect of cycloheximide on the reaction of puromycin with polysome-bound peptidyl-tRNA. FEBS Lett. 8, 249–252.

Bannister, A.J., Kouzarides, T., 2011. Regulation of chromatin by histone modifications. Cell Res. 21, 381–395.

Barakat, A., Szick-Miranda, K., Chang, I.-F., Guyot, R., Blanc, G., Cooke, R., Delseny, M., Bailey-Serres, J., 2001. The Organization of Cytoplasmic Ribosomal Protein Genes in the Arabidopsis Genome. Plant Physiol. 127, 398–415.

Bartel, D.P., 2009. MicroRNA Target Recognition and Regulatory Functions. Cell 136, 215–233.

Bohnsack, M.T., Regener, K., Schwappach, B., Saffrich, R., Paraskeva, E., Hartmann, E., Gorlich, D., 2002. Exp5 exports eEF1A via tRNA from nuclei and synergizes with other transport pathways to confine translation to the cytoplasm. EMBO J. 21, 6205–6215.

Bortoluzzi, S., d' Alessi, F., Romualdi, C., Danieli, G.A., 2001. Differential expression of genes coding for ribosomal proteins in different human

tissues.
Bioinformatics 17, 1152–1157.
Brogna, S., Sato, T.-A., Rosbash, M., 2002. Ribosome Components Are Associated with Sites of Transcription. Mol. Cell 10, 93–104.
Calvo, S.E., Pagliarini, D.J., Mootha, V.K., 2009. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc. Natl. Acad. Sci. U. S. A. 106, 7507–7512.
Carroll, A.J., Heazlewood, J.L., Ito, J., Millar, A.H., 2008. Analysis of the Arabidopsis Cytosolic Ribosome Proteome Provides Detailed Insights into Its Components and Their Post-translational Modification. Mol. Cell. Proteomics 7, 347–369.
Chaudhuri, J., Si, K., Maitra, U., 1997. Function of Eukaryotic Translation Initiation Factor 1A (eIF1A) (Formerly Called eIF-4C) in Initiation of Protein Synthesis. J. Biol. Chem. 272, 7883–7891.
Colon-Ramos, D.A., Shenvi, C.L., Weitzel, D.H., Gan, E.C., Matts, R., Cate, J., Kornbluth, S., 2006. Direct ribosomal binding by a cellular inhibitor of translation. Nat. Struct. Mol. Biol. 13, 103–111.
Dahlberg, J.E., Lund, E., 2004. Does protein synthesis occur in the nucleus? Curr. Opin. Cell Biol. 16, 335–338.
Dai, M.-S., Arnold, H., Sun, X.-X., Sears, R., Lu, H., 2007. Inhibition of c-Myc activity by ribosomal protein L11. EMBO J. 26, 3332–3345.
Dai, M.-S., Sun, X.-X., Lu, H., 2010. Ribosomal Protein L11 Associates with c-Myc at 5 S rRNA and tRNA Genes and Regulates Their Expression. J. Biol. Chem.
285, 12587–12594.
Dame, R.T., 2005. The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. Mol. Microbiol. 56, 858–870.
David, A., Dolan, B.P., Hickman, H.D., Knowlton, J.J., Clavarino, G., Pierre, P., Bennink, J.R., Yewdell, J.W., 2012. Nuclear translation visualized by ribosome-bound nascent chain puromycylation. J. Cell Biol. 197, 45–57.
De, S., Brogna, S., 2010. Are ribosomal proteins present at transcription sites on or off ribosomal subunits? Biochem. Soc. Trans. 38, 1543.
De, S., Varsally, W., Falciani, F., Brogna, S., 2011. Ribosomal proteins' association with transcription sites peaks at tRNA genes in Schizosaccharomyces

pombe. RNA 17, 1713–1726.

Eggers, D.K., Welch, W.J., Hansen, W.J., 1997. Complexes between nascent polypeptides and their molecular chaperones in the cytosol of mammalian cells. Mol. Biol. Cell 8, 1559–1573.

Frank, J., 2000. The ribosome – a macromolecular machine par excellence. Chem. Biol. 7, R133–R141.

Fuchs, G., Diges, C., Kohlstaedt, L.A., Wehner, K.A., Sarnow, P., 2011. Proteomic Analysis of Ribosomes: Translational Control of mRNA populations by Glycogen Synthase GYS1. J. Mol. Biol. 410, 118–130.

Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J.,
Rick, J.M., Michon, A.-M., Cruciat, C.-M., Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.-A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., Superti- Furga, G., 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141–147.

Goidl, J.A., Canaani, D., Boublik, M., Weissbach, H., Dickerman, H., 1975. ..Polyanion-induced release of polyribosomes from HeLa cell nuclei. J. Biol. Chem. 250, 9198–9205.

Grollman, A.P., 1968. Inhibitors of Protein Biosynthesis V. Effects of emetine on protein and nucleic acid biosynthesis in HeLa cells. J. Biol. Chem. 243, 4089–4094.

Gunderson, J.H., Sogin, M.L., Wollett, G., Hollingdale, M., Cruz, V. de la, Waters, A.P., McCutchan, T.F., 1987. Structurally distinct, stage-specific ribosomes occur in Plasmodium. Science 238, 933–937.

Hariharan, N., Kelley, D.E., Perry, R.P., 1989. Equipotent mouse ribosomal protein promoters have a similar architecture that includes internal sequence elements. Genes Dev. 3, 1789–1800.

Hartenstein, A.Y., Rugendorff, A., Tepass, U., Hartenstein, V., 1992. The function of the neurogenic genes during epithelial development in the Drosophila embryo. Development 116, 1203–1220.

Hu, C.-D., Chinenov, Y., Kerppola, T.K., 2002. Visualization of Interactions among bZIP and Rel Family Proteins in Living Cells Using Bimolecular Fluorescence Complementation. Mol. Cell 9, 789–798.

Iborra, F.J., Jackson, D.A., Cook, P.R., 2001. Coupled Transcription and Translation Within Nuclei of Mammalian Cells. Science 293, 1139–1142.

Ivanov, A.V., Malygin, A.A., Karpova, G.G., 2004. Human Ribosomal Protein S26 Inhibits Splicing of Its Own Pre-mRNA. Mol. Biol. 38, 568–574.

Jack, K., Bellodi, C., Landry, D.M., Niederer, R., Meskauskas, A., Musalgaonkar, S., Kopmar, N., Krasnykh, O., Dean, A.M., Thompson, S.R., Ruggero, D., Dinman, J.D., 2011. rRNA Pseudouridylation Defects Affect Ribosomal Ligand Binding and Translational Fidelity from Yeast to Human Cells. Mol. Cell 44, 660–666.

Jackson, R.J., Hellen, C.U.T., Pestova, T.V., 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. 11, 113–127.

Jiang, H., Edgar, B.A., 2009. EGFR signaling regulates the proliferation of Drosophila adult midgut progenitors. Dev. Camb. Engl. 136, 483–493.

Jurica, M.S., Moore, M.J., 2003. Pre-mRNA Splicing: Awash in a Sea of Proteins. Mol. Cell 12, 5–14.

Kai, T., Williams, D., Spradling, A.C., 2005. The expression profile of purified Drosophila germline stem cells. Dev. Biol. 283, 486–502.

Kellis, M., Birren, B.W., Lander, E.S., 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature 428, 617–624.

Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Ban, N., 2012. Atomic structures of the eukaryotic ribosome. Trends Biochem. Sci. 37, 189–198.

Kressler, D., Hurt, E., Baβler, J., 2010. Driving ribosome assembly. Biochim. Biophys. Acta BBA - Mol. Cell Res., Molecular Chaperones and Intracellular Protein Transport 1803, 673–683.

Lamb, M.J., 1982. The DNA content of polytene nuclei in midgut and Malpighian tubule cells of adultDrosophila melanogaster. Wilhelm Rouxs Arch. Dev. Biol. 191, 381–384.

Latchman, D.S., 1997. Transcription factors: An overview. Int. J. Biochem. Cell Biol. 29, 1305–1312.

Lebaron, S., Schneider, C., van Nues, R.W., Swiatkowska, A., Walsh, D., Bottcher, B., Granneman, S., Watkins, N.J., Tollervey, D., 2012. Proof reading of pre-40S ribosome maturation by a translation initiation factor and 60S subunits. Nat. Struct. Mol. Biol. 19, 744–753.

Lee, S.-W., Berger, S.J., Martinovic, S., Pasa-Tolic, L., Anderson, G.A., Shen, Y., Zhao, R., Smith, R.D., 2002. Direct mass spectrometric analysis of intact proteins of the yeast large ribosomal subunit using capillary LC/FTICR. Proc. Natl. Acad. Sci. U. S. A. 99, 5942–5947.

Lindström, M.S., 2009. Emerging functions of ribosomal proteins in gene-specific transcription and translation. Biochem. Biophys. Res. Commun. 379, 167–170.

Lopes, A.M., Miguel, R.N., Sargent, C.A., Ellis, P.J., Amorim, A., Affara, N.A., 2010. The human RPS4 paralogue on Yq11.223 encodes a structurally conserved ribosomal protein and is preferentially expressed during spermatogenesis. BMC Mol. Biol. 11, 33.

Maddox, B., 2003. The double helix and the “wronged heroine”. Nature 421, 407-408.

Malygin, A.A., Parakhnevitch, N.M., Ivanov, A.V., Eperon, I.C., Karpova, G.G., 2007. Human ribosomal protein S13 regulates expression of its own gene at the splicing step by a feedback mechanism. Nucleic Acids Res. 35, 6414–6423.

Marygold, S.J., Roote, J., Reuter, G., Lambertsson, A., Ashburner, M., Millburn, G.H., Harrison, P.M., Yu, Z., Kenmochi, N., Kaufman, T.C., Leevers, S.J., Cook, K.R., 2007. The ribosomal protein genes and Minute loci of Drosophila melanogaster. Genome Biol. 8, R216.

Mathur, D., Bost, A., Driver, I., Ohlstein, B., 2010. A Transient Niche Regulates the Specification of Drosophila Intestinal Stem Cells. Science 327, 210–213.

Mauro, V.P., Edelman, G.M., 2002. The ribosome filter hypothesis. Proc. Natl. Acad. Sci. 99, 12031–12036.

McLeod, T., Abdullahi, A., Li, M., Brogna, S., 2014. Recent studies implicate the nucleolus as the major site of nuclear translation. Biochem. Soc. Trans. 42, 1224–1228.

Nathans, D., 1964. Puromycin inhibition of protein synthesis: Incorporation of puromycin into peptide chains. Proc. Natl. Acad. Sci. U. S. A. 51, 585–592.

Nathanson, L., Xia, T., Deutscher, M.P., 2003. Nuclear protein synthesis: A re-

evaluation. RNA 9, 9–13.
Ni, L., Snyder, M., 2001. A Genomic Study of the Bipolar Bud Site Selection Pattern in Saccharomyces cerevisiae. Mol. Biol. Cell 12, 2147–2170.
Nishimura, T., Wada, T., Yamamoto, K.T., Okada, K., 2005. The Arabidopsis STV1 Protein, Responsible for Translation Reinitiation, Is Required for Auxin-Mediated Gynoecium Patterning. Plant Cell 17, 2940–2953.
Odintsova, T.I., Müller, E.-C., Ivanov, A.V., Egorov, T.A., Bienert, R., Vladimirov, S.N., Kostka, S., Otto, A., Wittmann-Liebold, B., Karpova, G.G., 2003. Characterization and analysis of posttranslational modifications of the human large cytoplasmic ribosomal subunit proteins by mass spectrometry and Edman sequencing. J. Protein Chem. 22, 249–258.
Ohlstein, B., Spradling, A., 2006. The adult Drosophila posterior midgut is maintained by pluripotent stem cells. Nature 439, 470–474.
Panse, V.G., Johnson, A.W., 2010. Maturation of Eukaryotic Ribosomes: Acquisition of Functionality. Trends Biochem. Sci. 35, 260–266.
Park, H.-S., Himmelbach, A., Browning, K.S., Hohn, T., Ryabova, L.A., 2001. A Plant Viral "Reinitiation" Factor Interacts with the Host Translational Machinery. Cell 106, 723–733.
Planta, R.J., Mager, W.H., 1998. The list of cytoplasmic ribosomal proteins of Saccharomyces cerevisiae. Yeast 14, 471–477.
Rouquette, J., Choesmel, V., Gleizes, P.-E., 2005. Nuclear export and cytoplasmic processing of precursors to the 40S ribosomal subunits in mammalian cells. EMBO J. 24, 2862–2872.
Rugjee, K.N., Roy Chaudhury, S., Al-Jubran, K., Ramanathan, P., Matina, T., Wen, J., Brogna, S., 2013. Fluorescent protein tagging confirms the presence of ribosomal proteins at Drosophila polytene chromosomes. PeerJ 1.
Schroder, P.A., Moore, M.J., 2005. Association of ribosomal proteins with nascent transcripts in S. cerevisiae. RNA 11, 1521–1529.
Shanbhag, S., Tripathi, S., n.d. Epithelial ultrastructure and cellular mechanisms of acid and base transport in the Drosophila midgut [WWW Document]. URL http://jeb.biologists.org (accessed 8.29.14).
Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates III, J.R., Frank, J., Manley, J.L., 2009. Molecular Architecture of the Human Pre-

mRNA 3′ Processing Complex. Mol. Cell 33, 365–376.
Shigenobu, S., Kitadate, Y., Noda, C., Kobayashi, S., 2006. Molecular characterization of embryonic gonads by gene expression profiling in Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. 103, 13728–13733.
Soudet, J., Gelugne, J.-P., Belhabich-Baumas, K., Caizergues-Ferrer, M., Mougin, A., 2010. Immature small ribosomal subunits can engage in translation initiation in Saccharomyces cerevisiae. EMBO J. 29, 80–92.
Spence, J., Gali, R.R., Dittmar, G., Sherman, F., Karin, M., Finley, D., 2000. Cell Cycle–Regulated Modification of the Ribosome by a Variant Multi-ubiquitin Chain. Cell 102, 67–76.
Spriggs, K.A., Stoneley, M., Bushell, M., Willis, A.E., 2008. Re-programming of translation following cell stress allows IRES-mediated translation to predominate. Biol. Cell 100, 27–38.
Sproul, D., Gilbert, N., Bickmore, W.A., 2005. The role of chromatin structure in regulating the expression of clustered genes. Nat. Rev. Genet. 6, 775–781.
Steitz, J.A., Berg, C., Hendrick, J.P., Branche-Chabot, H.L., Metspalu, A., Rinke, J., Yario, T., 1988. A 5S rRNA/L5 complex is a precursor to ribosome assembly in mammalian cells. J. Cell Biol. 106, 545–556.
Struhl, K., 1999. Fundamentally Different Logic of Gene Regulation in Eukaryotes and Prokaryotes. Cell 98, 1–4.
Strunk, B.S., Loucks, C.R., Su, M., Vashisth, H., Cheng, S., Schilling, J., Brooks, C.L., Karbstein, K., Skiniotis, G., 2011. Ribosome Assembly Factors Prevent Premature Translation Initiation by 40S Assembly Intermediates. Science 333, 1449–1453.
Sugihara, Y., Honda, H., Iida, T., Morinaga, T., Hino, S., Okajima, T., Matsuda, T., Nadano, D., 2010. Proteomic Analysis of Rodent Ribosomes Revealed Heterogeneity Including Ribosomal Proteins L10-like, L22-like 1, and L39-like. J. Proteome Res. 9, 1351–1366.
Suzuki, M.M., Bird, A., 2008. DNA methylation landscapes: provocative insights from epigenomics. Nat. Rev. Genet. 9, 465–476.
Takashima, S., Adams, K.L., Ortiz, P.A., Ying, C.T., Moridzadeh, R., Younossi-Hartenstein, A., Hartenstein, V., 2011. Development of the Drosophila entero- endocrine lineage and its specification by the Notch signaling

pathway. Dev. Biol. 353.
Udem, S.A., Warner, J.R., 1973. The Cytoplasmic Maturation of a Ribosomal Precursor Ribonucleic Acid in Yeast. J. Biol. Chem. 248, 1412–1416.
Velichutina, I.V., Rogers, M.J., McCutchan, T.F., Liebman, S.W., 1998. Chimeric rRNAs containing the GTPase centers of the developmentally regulated ribosomal rRNAs of Plasmodium falciparum are functionally distinct. RNA 4, 594–602.
Wan, F., Anderson, D.E., Barnitz, R.A., Snow, A., Bidere, N., Zheng, L., Hegde, V., Lam, L.T., Staudt, L.M., Levens, D., Deutsch, W.A., Lenardo, M.J., 2007. Ribosomal Protein S3: A KH Domain Subunit in NF-κB Complexes that Mediates Selective Gene Regulation. Cell 131, 927–939.
Warner, J.R., McIntosh, K.B., 2009. How Common Are Extraribosomal Functions of Ribosomal Proteins? Mol. Cell 34, 3–11.
Watson, J.D., Baker, T.A., 2007. Molecular Biology of the Gene, 6 edition. ed. Pearson, San Francisco; Cold Spring Harbor, N.Y.
Watson, J.D., Crick, F.H.C., 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature 171, 737–738.
Weijers, D., Dijk, M.F., Vencken, R.-J., Quint, A., Hooykaas, P., Offringa, R., 2001. An Arabidopsis Minute-like phenotype caused by a semi-dominant mutation in a ribosomal.protein S5 gene. Development 128, 4289–4299.
Whittle, C.A., Krochko, J.E., 2009. Transcript Profiling Provides Evidence of Functional Divergence and Expression Networks among Ribosomal Protein Gene Paralogs in Brassica napus. Plant Cell Online 21, 2203–2219.
Wool, I.G., 1996. Extraribosomal functions of ribosomal proteins. Trends Biochem.
Sci. 21, 164–165.
Xue, S., Barna, M., 2012. Specialized ribosomes: a new frontier in gene regulation and organismal biology. Nat. Rev. Mol. Cell Biol. 13, 355–369.
Yu, Y., Ji, H., Doudna, J.A., Leary, J.A., 2005. Mass spectrometric analysis of the human 40S ribosomal subunit: Native and HCV IRES-bound complexes. Protein Sci. Publ. Protein Soc. 14, 1438–1446.
Zamir, A., Leder, P., Elson, D., 1966. A ribosome-catalyzed reaction between N-Formylmethionyl-tRNA and puromycin. Proc. Natl. Acad. Sci. U. S. A. 56, 1794–1801.