Caldalkalibacillus thermarum PhoH2: Solving the solubility puzzle

A thesis

submitted in partial fulfilment

of the requirements for the degree

of

Master of Science

at

The University of Waikato

by

Brooke Jane Dillon

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2017

Abstract

Previous attempts to determine the structure of the protein PhoH2 from *Mycobacterium tuberculosis*, *Mycobacterium smegmatis* and *Thermobispora bispora* have been unsuccessful producing diffraction data to low resolution. The thermophilic organism *Caldalkalibacillus thermarum* was thought to be an advantageous option from which to clone, express and purify PhoH2 from for the purpose of structure determination. PhoH2 consists of two domains, an N terminal PIN domain, known for its toxic properties as part of many toxin-antitoxin systems present in *M. tuberculosis*, and a C-terminal PhoH domain, an RNA helicase suspected to be involved in phosphate starvation responses. The cloning and expression of three variations of recombinant *C. thermarum* PhoH2 was successful in *Escherichia coli*. However the purification of PhoH2 continues to yield insoluble protein, despite the range of purification buffer conditions screened, thus preventing downstream biochemical characterisation and structural investigations. Multiple options exist to overcome this insolubility problem; including alternate plasmids for protein expression and purification that alter the tag location and type such as C-terminal His-tags and alternative fusion tags.

Acknowledgements

I would first like to thank Professor Vic Arcus for presenting this opportunity, and the support and abundance of knowledge and experience you have provided towards completing this project.

Secondly to Dr Emma Andrews, thank you for the support whenever I ran into trouble or had a question about this research, often with simple answers. Your patience and expertise is much appreciated.

I will treasure this opportunity to work with you both, and am grateful of the experience and biochemistry techniques I have gained during this time.

To the remainder of the Protein and Microbes Lab, thanks for the collective brainstorming, and the reminder that nothing is ever as simple as it should seem.

To Tom, and Mum and Dad; thank you for convincing me to complete this qualification and then listening to the endless and repetitive frustration over the duration of the project. Without you all, this thesis would not be what it is today.

"*You don't have time to be timid. You must be bold and daring!*"

- Lumiere, Beauty and the Beast.

Table of Contents

Abstract ii

Acknowledgements iii

Table of Contents iv

List of Figures vii

List of Tables x

List of Abbreviations xi

1 Introduction 1

1.1 Tuberculosis 1

1.2 Toxin-Antitoxin Systems 3

1.2.1 PIN Domains 8

1.3 Helicase Proteins 11

1.3.1 PhoH Domains 12

1.4 PhoH2 fusion protein 13

1.5 *Caldalkalibacillusthermarum* 14

1.6 Objectives 15

2 Materials and Methods 16

DNA Manipulations 16

2.1 Genomic DNA extraction 16

2.1.1.1 Media and Growth Conditions 16

2.1.2 Vector DNA extraction from E coli 16

2.1.3 Agarose Gel Electrophoresis 16

2.1.4 DNA Quantification 16

2.1.5 PCR 17

2.1.5.1 Primers 17

2.1.5.2 PCR for amplification of PhoH2 from Genomic DNA for Vector inserts 17

Table 2.1and Table 2.2describe the PCR amplification methods and cycling parameters used in this study. 17

2.1.6 Purification of PCR product in solution 19

2.1.7 Restriction Enzyme Digest 19

2.1.8 DNA Ligation 19

2.1.9 DNA Transformation 20

2.1.9.1 Preparation and transformation of chemically competent *E. coli*. 20

2.1.9.2 Glycerol Stocks 20

2.1.10 DNA Sequencing 20

2.2 Protein Expression and Purification 21

2.2.1 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) Protein Analysis 21

2.2.3 Coomassie Blue Staining for Protein Gel electrophoresis 21

2.2.4 Measurement of Protein concentration 21

2.2.5 Concentration of Protein Samples 22

2.2.6 Small Scale Protein expression trial 22

2.2.7 Purification of His - Tagged proteins using IMAC/Nickel chromatography 23

2.2.8 Size Exclusion chromatography 23

2.3 Protein ATPase Assays 24

3 Results 25

3.1 Cloning, Expression and Purification of *C. thermarum* PhoH2 25

3.1.1 Cloning and Expression 25

Cloning Construct A 28

Cloning Construct B 31
Cloning Construct C 36
The identification of alternative start sites 36
Discussion 43
References 47

List of Figures

Figure 1.1: Global estimated TB incident rates according to population. Figure taken from WHO (2015) [2]. 1
Figure 1.2: Mode of action of the five classified toxin antitoxin systems. Figure taken from Wen, Behiels & Devreese (2014)[22]. 4
Figure 1.3: The toxin-antitoxin genes present in the M. tuberculosis genome. Dark blue TA systems have been identified as most commonly induced in drug-tolerant cells. Red dots indicate inhibition of growth, while grey indicates no growth inhibition in E. coli, M. smegmatis and M. Tuberculosis. Figure taken from Sala, Bordes & Genevaux (2014) [19]. 5
Figure 1.4: A schematic diagram of the (a) Type I and (b) Type II toxin antitoxin systems. Of particular importance is the direction of the antitoxin gene [1]. 7
Figure 1.5: Hidden Markov Model (HMM) of PIN domain proteins. Of importance is the four conserved residues D, E, D, D that are conserved at strict positions as indicated by the height and width of each letter. Generated from pFam, [35]. 9
Figure 1.6: The structure of PIN domains found in PyrobaculumaerophilumPAE2754. Figures were made with Pymol (http://www.pymol.org/) using PDB 2FEI [31, 32]. 10
Figure 1.7:Schematic diagram outlining the conserved nature of the nucleotides in the hairpin structure in PIN domain toxin RNA substrates (figure taken from McKenzie, Robson [29]). 11
Figure 1.8: Hidden Markov model (HMM) of PhoH proteins. Conserved motifs are annotated as described in [32]. Generated fromPfam[35]. 12
Figure 1.9: Schematic arrangement of the PIN and PhoH domains in $PhoH_2$ from C. thermarum. 13
Figure 3.1: 1% agarose gels of the PhoH2 gene a) prior to restriction endonuclease digestion, and b) following restriction endonuclease digestion..... 25
Figure 3.2: PhoH2 insert following RE digest dropout to confirm the presence of PhoH2 in BL21 E.coli..... 26
Figure 3.3: 12 % SDS-PAGE gels showing the small scale expression trials of PhoH2 cloned intopET28b-PstIfor expression in E. coli BL21at 37 °C, 30 °C, 25 °C and 18

°C. Controls contained empty pET28b-PstI were tested at 30 °C, 25 °C and 18 °C in parallel. Expected band size 56 kDa not present in control or PhoH2 samples. .. 27
Figure 3.4: Schematic diagram of the 3 cloning constructs 28
Figure 3.5: Small scale expression of PhoH2 in pPROEX-Htb plasmid from DH5α E. coli at 37 °C Cells WC - Whole Cell, SN - Supernatant, Insoluble, W1 - Wash 1, Resin - Ni Beads. White box highlights protein band of expected size present in WC and Resin samples 29
Figure 3.6: IMAC chromatography curve and corresponding 12 % SDS-PAGE of the fractions covering the main peaks as shown by the black line across the chromatography curve corresponding to fractions A1 – A12, grown at 37 °C..... 29
Figure 3.7: His-tag stain 12% SDS-PAGE, followed by a standard Coomassie Stain. pFu protein was used as a control for size (47 kDa) and presence of His-tag. Fractions A1-A4 for both expression temperatures (37 °C and 30 °C) were loaded respectively. 30
Figure 3.8: Size Exclusion Chromatography Curve and Corresponding 12% SDS-PAGE a) 37 °C and b) 30 °C. 31
Figure 3.9: Small scale expression of PhoH2 in pPROEX-Htb plasmid in DH5α E.coli cells over 37 °C, 30 °C and 28 °C post-induction. The black boxes highlight theweak protein band of expected size present in Resin samples over all temperatures. Note there is no enrichment of PhoH2 in the 37 °C control (empty pPROEX-HtB). 32
Figure 3.10: SDS-PAGE gels of small scale expression trials screening increased salt concentrations and the addition of 10 % glycerol to lysis buffers on PhoH2 expressed in DH5α E. coli cells. 34
Figure 3.11: SDS-PAGE gels of small scale expression trials screening increased pH of lysis buffers on PhoH2 expressed in BL21 E. coli cells. 34
Figure 3.12: SDS-PAGE gels of small scale expression trials screening increased salt concentrations and the addition of 10 % glycerol to lysis buffers on PhoH2. 35
Figure 3.13: Schematic diagram of 30-PhoH2 construct. 37
Figure 3.14: Schematic diagram of 18-PhoH2 construct. 37
Figure 3.15: Gradient PCR from C. Thermarum gDNA using cloning primers on 30-PhoH2 and 18-PhoH2 constructs respectively. 38

Figure 3.16: SDS-PAGE gels of small scale expression trials screening variable expression temperatures 37°C, 30°C and 28°C using lysis buffer 50 mMTris pH 8.0, 200 mM NaCl, 20 mM Imidazole in 30-PhoH2..... 39
Figure 3.17: SDS-PAGE gels of small scale expression trials screening increased pH of lysis buffers in 30-PhoH2. 39
Figure 3.18: SDS-PAGE gels of small scale expression trials screening increased salt concentrations and the addition of 10 % glycerol to lysis buffers in 30-PhoH2... 40
Figure 3.19: SDS-PAGE gels of small scale expression trials screening variable expression temperatures 37°C, 30°C and 28°C using lysis buffer 50 mMTris pH 8.0, 200 mMNaCl, 20 mM Imidazole in 18-PhoH2..... 40
Figure 3.20: SDS-PAGE gels of small scale expression trials screening increased pH of lysis buffers in 18-PhoH2. 41
Figure 3.21: SDS-PAGE gels of small scale expression trials screening increased pH of lysis buffers in 18-PhoH2. 41

List of Tables

Table 2.1: PCR reactions used in the following study. Hotfire pol used for cloning. KAPA HiFiHotstart used when cloning has been difficult. Taq Polymerase, a non-proofreading Polymerase used for screening colonies. 17

Table 2.2: Cycling conditions for each of the PCR reactions used. 18

Table 2.3: Primers designed for cloning and sequencing C thermarum PhoH2, ordered from IDT technologies. 19

Table 3.1: Composition of purification buffers varying pH conditions. 33

Table 3.2: Composition of purification buffers varying Salt and glycerol conditions. 35

List of Abbreviations

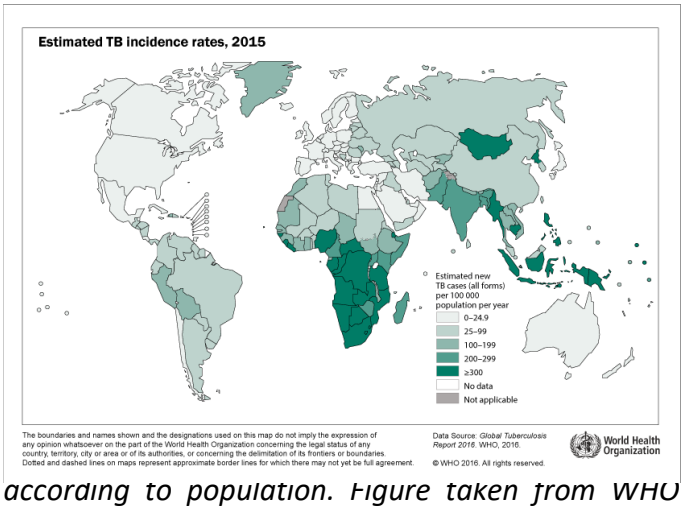
µg	microgram (10^{-6} g)
µl	microliter (10^{-6} L)
µm	micrometer (10^{-6} m)
µM	micromolar (10^{-6} M)
A	adenine
ATP	adenosine triphosphate
APS	ammonium persulfate
bp	base pair(s)
C	cytosine
Da	Daltons
DNA	deoxyribose nucleic acids
EDTA	ethylene diamine tetraacetic acid (disodium salt)
FPLC	Fast Protein Liquid Chromatography
G	guanine
g	centrifugal acceleration force
HIV	Human Immunodeficiency Virus
His-tag	poly-histidine tag
kb	kilobase
kDa	kilo Dalton
LB	Luria Bertani
mg	milligram (10^{-3} g)
MGS	Massey Genome Service
mL	millilitre (10^{-3} L)
mM	(10^{-3} M)
MW	molecular weight
nm	wavelength/ nanometers (10^{-9} M)
OD_{600}	optical density at 600 nm wavelength
PAGE	poly acrylamide gel electrophoresis
PCR	polymerase chain reaction
pI	isoelectric point
PIN	PilT N-terminal domain
RBS	Ribosome Binding Site
RNA	Ribonucleic Acid
RNase	ribonuclease
rpm	revolutions per minute
rRNA	ribosomal RNA
SDS	sodium dodecyl sulphate
SOC	super optimal broth with catabolite repression
T	thymine
TA	toxin-antitoxin
TAE	tris-acetate-EDTA
TB	tuberculosis
TE	tris EDTA buffer
TEMED	N, N, N, N,-tetramethylethylenediamine
Tm	melting temperature

Tris	tris-hydroxymethylaminomethane
tRNA	transfer ribonucleic acid
TSB	Tryptic Soy Broth media
TSP	transcriptional start point
U	uracil
UV	ultraviolet
V	Volts
Vap	virulence associated protein
v/v	volume per volume
WHO	World Health Organisation
w/v	weight per volume
w/w	weight per weight
XDR	Extensively Drug Resistance

1 Introduction

1.1 Tuberculosis

The disease tuberculosis (TB) remains globally the second highest cause of death by a single infectious agent (caused by *Mycobacterium tuberculosis*) ranking behind the human immunodeficiency virus (HIV) [2]. Despite the availability of the preventative Bacille Calmette-Guérin (BCG) vaccine and known therapy options including strict 6 month antibiotic treatments with rifampicin and isoniazid [3], the prevalence of TB worldwide is high particularly in developing countries. According to the Global tuberculosis report 2015, 9.6 million people worldwide are estimated to have fallen ill with the disease, with 1.5 million of those cases being fatal [2]. Increased risk to exposure and infection of TB occurs in Africa and South-East Asia and the Western Pacific, with these regions accounting for 21.5%, 40.95 % and 22.1% respectively, and are now considered to be on the TB burden country list (Figure 1.1) [2].Contributing factors to increased risk of TB burden include poverty, immuno-depressive drugs and infection by HIV [4]. Although it has long been thought human TB disease originated from the Bovine TB disease, genetic research into the two closely related strains suggests this may not be the case, and human TB, as a result of the bacterium *M. tuberculosis*, is likely to be closest to the ancestral bacilli [5].



according to population. Figure taken from WHO (2015) [2].

Pulmonary TB; infection of the lungs, is the most common form of TB, however extra-pulmonary infections have been observed in peripheral areas of the body. TB disease progresses following human inhalation of tuberculosis bacilli in aerosol droplets into the lungs where the Mycobacteria proliferate inside macrophages in the alveolar space, rather than being destroyed by the immune

response[4, 6]. This triggers the attraction of other immune cells which form a granuloma known as the hallmark of TB [7]. The granulomas create a microenvironment with extreme conditions such as nutrient starvation, low pH and hypoxia [8].

The Mycobacteriaceae family of bacteria which are genus of Actinobacteria are responsible for a wide range of often fatal diseases in mammals. In addition to TB, Leprosy (Hansen's disease) is also a mycobacterial disease which is less prevalent. Mycobacterial diseases such as TB are difficult to treat and to do so, require long courses combining multiple antibiotics [3]. Extensively Drug Resistant (XDR) strains are becoming increasingly prevalent, where antibiotic resistance is occurring[9]. Antibiotic resistance is increasing at an alarming rate across many diseases caused by a range of bacteria, including pneumonia, gonorrhoea and tuberculosis [10]. Although resistance occurs naturally, the over-prescription of antibiotics and over-the-counter supply of antibiotic treatments are exacerbating the issue [11]. The World Health Organisation (WHO) issued a report indicating a full review of how and when antibiotics should be prescribed to treat infections and diseases as a way to slow the progression of antibiotic resistance, specifically for TB [10].

A defining feature of Mycobacteria is their unique cell wall structure, thought to be crucial to the strength and resistance of this bacterial family [12].This acts as a barrier to permeation and therefore prevents molecules, including drugs from entering the cell [12]. This unique strength of the cell wall is often a limiting factor to many of the studies which attempt to understand more about this class of bacteria.

Dormant *M. tuberculosis* must be able to survive the toxic conditions in the granuloma for extended periods of time [13]. In addition to the unique cell wall, *M. tuberculosis* has reduced metabolism and replication, the details of which are still poorly understood [14]. There are multiple suggestions as to the changes *M. tuberculosis* undergoes to survive under such conditions over the long term within the granuloma. For example, transcriptional profiling of *M. Tuberculosis* shows that the genes encoding sulphur metabolism are regulated by

environmental cues, such as that of nutrient starvation and decrease in oxygen and carbon dioxide levels [15]. Control of the hosts apoptosis response by adaptation of metabolism has been reviewed in detail [16]. Increased production of Interleukin-10 (also known as human cytokine synthesis inhibitory factor) bound to TNFR2 (Tumor Necrosis Factor Receptor 2) by the Mycobacterium in response to the hosts increasing T cell immune response, results in the Mycobacterium halting the apoptosis cycle and increasing the viability of the infecting organism [17].

A key genomic feature that has been linked to dormancy and long term survival (in addition to the physical features of the Mycobacterial cell wall), is the unusually large number of toxin-antitoxin genes encoded in the *M. tuberculosis* genome, a total of 79 at last count (Figure 1.3) [18, 19].

1.2 Toxin-Antitoxin Systems

A toxin-antitoxin (TA) system is two closely spaced genes, often as part of an operon. The first TA system was discovered on a plasmid and had a role in maintaining the plasmid in bacterial daughter cells [19]. These genes encode for a specific protein toxin which evokes a highly toxic effect on the cell, such as cell death or growth arrest. A second unstable antagonistic antitoxin, which can be either a protein or non-coding RNA, is also expressed which is able to neutralise the toxin under strict optimum conditions [20], [1].

Toxin-antitoxin systems have been classified into 5 distinct groups, according to their antitoxin mechanism of inhibition (Figure 1.2) [21, 22].

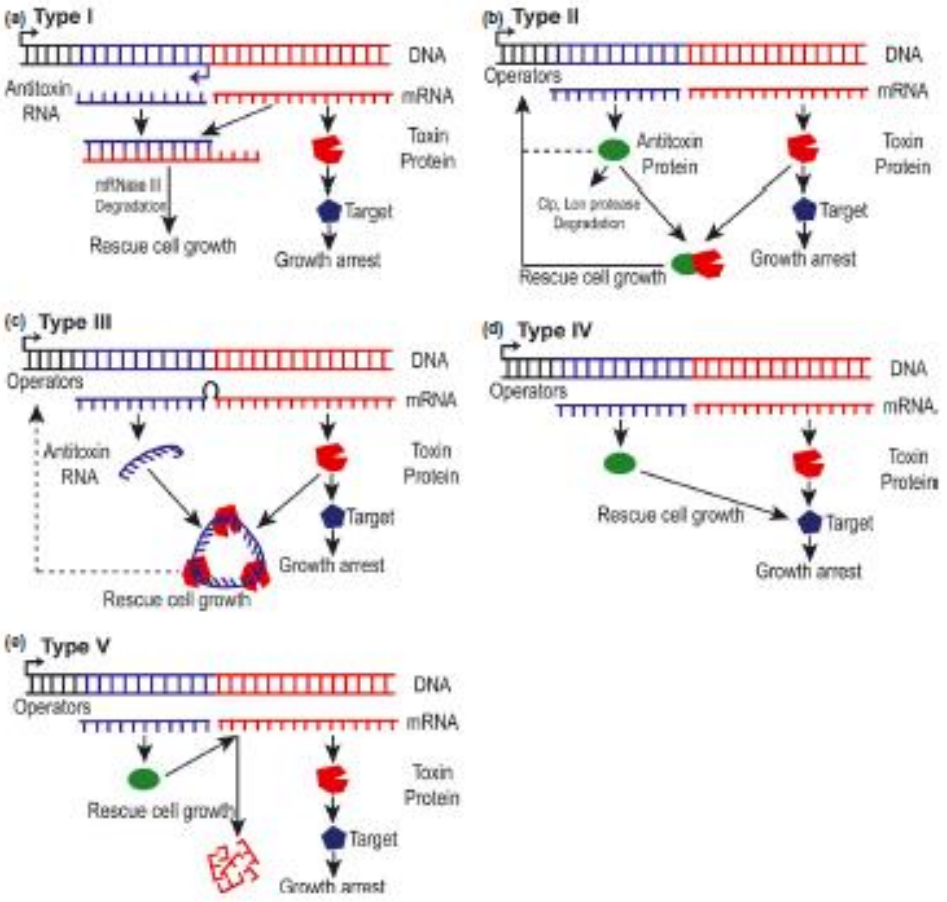


Figure 1.2: Mode of action of the five classified toxin antitoxin systems. Figure taken from Wen, Behiels & Devreese (2014)[22].

- Type I - non-coding sRNA inhibits the translation of the toxin mRNA by degrading or blocking the ribosome binding site.
- Type II - both toxin and antitoxin are small peptides, in which the binding of the antitoxin to the toxin causes neutralisation of the toxin (Figure 1.4).
- Type III - toxins are proteins which are inhibited by RNA antitoxin pseudoknots.
- Type IV - TA systems have protein components for both the toxin and the antitoxin. Inhibition of the toxin occurs when the antitoxin stabilises the target proteins, thereby blocking the toxin.
- The final group of TA systems has only recently been discovered, Type V systems have a protein antitoxin which binds and degrades the mRNA of the toxin [23].

Type I and Type II TA systems are of particular interest to the field of Mycobacteria research, such as *M. tuberculosis* and *M. Leprae* [23, 24]. In the *M. tuberculosis* genome H37Rv, there are 77 TA systems known, 68 are Type II TA systems of which 50 are Virulence Associated Protein (VapBC) TA systems [19].

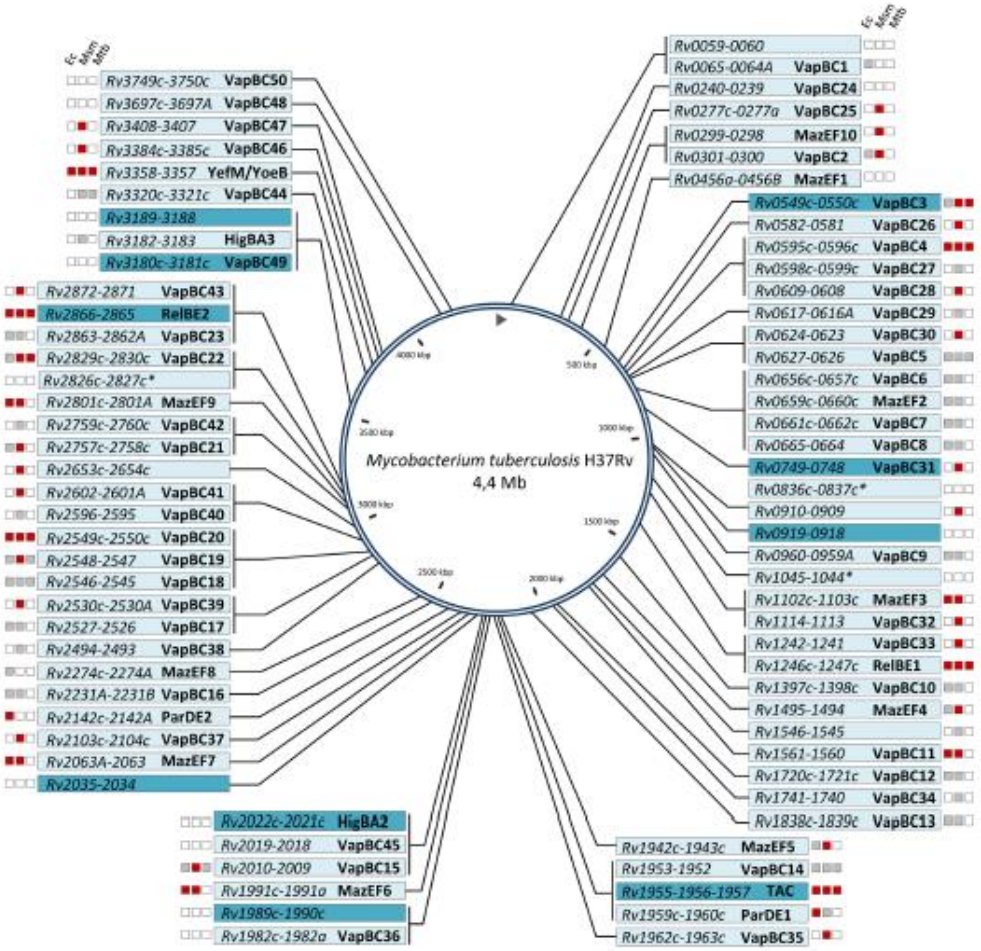


Figure 1.3: The toxin-antitoxin genes present in the M. tuberculosis genome. Dark blue TA systems have been identified as most commonly induced in drug-tolerant cells. Red dots indicate inhibition of growth, while grey indicates no growth inhibition in E. coli, M. smegmatis and M. Tuberculosis. Figure taken from Sala, Bordes & Genevaux (2014) [19].

Type I toxins are small hydrophobic peptides ranging between 20-65 amino acids in length[1, 20]. These act by disrupting the membrane potential of invading cells [25]. Type I antitoxins are small non-coding antisense RNAs which inhibit the toxin either by blocking the ribosome binding site, or degrading the toxin's mRNA [23]. Multiple Type I TA systems have been discovered in both Gram-positive and Gram-negative bacteria and archaea [26]. The transcription of the toxin gene is regulated

by the antisense RNA antitoxin, which covers the toxin gene, but is in the reverse direction (Figure 1.4) [1, 24]. The antisense RNA antitoxin of type I TA systems, such as the Hok-Sok, SymR-SymE and IstR-TisB TA systems, when transcribed, binds to the toxin mRNA which forms a double stranded RNA molecule [1]. It is believed that this formation of the double stranded RNA molecule targets it for degradation by exoribonucleases, thereby inhibiting the expression of the toxic protein [24]. The most understood of Type I TA systems is the Hok-Sok system, the first TA system to be discovered in 1986 in *Escherichia coli*. Studies of the Hok-Sok system along with other type I TA systems predominantly in *E.coli* have identified a conserved secondary structure to the toxin protein with a single α-helical structure thought to be crucial to the mechanism of toxicity; disruption of the cellular membrane potential [1, 19, 26].

Type II TA systems differ from Type I TA systems as both components of the Type II systems are proteins (Figure 1.4) [19, 25]. Like the Type I TA systems, the gene encoding the antitoxin component is upstream of the toxin gene, but in Type II systems is in the same 5' to 3' orientation as the toxin gene [1]. Typically, the toxin proteins are larger than the Type I toxins, with protein sequences ranging between 40-206 amino acids in length [27].The antitoxin protein binds to the toxin protein (Figure 1.4), and either blocks the active site of the toxin, or induces steric changes to the toxin, resulting in a non-functional protein, unable to exert its toxicity on the cell [19].

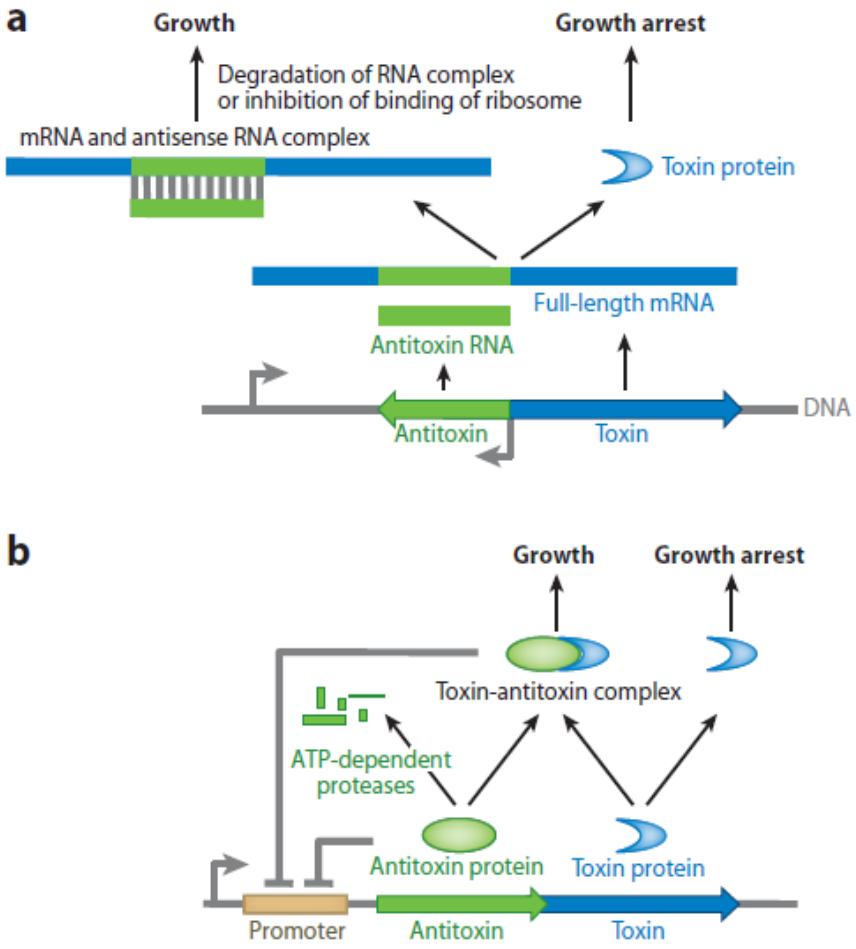


Figure 1.4: A schematic diagram of the (a) Type I and (b) Type II toxin antitoxin systems. Of particular importance is the direction of the antitoxin gene [1].

Type II TA systems which have been identified and characterised include 5 families of toxin Proteins; MazEF, RelBE, ParDE, YefM/YoeB, HigB, TAC and VapBC.

In *M. tuberculosis* these TA systems have been identified as occurring in the genome 10, 2, 2, 1, 2, 1 and 45+ times respectively. In contrast to other bacteria where very few (1-5) TA operons are observed, this uniquely high VapBC TA operon presence has prompted multiple studies into the purpose of TA systems, and the possible implications these have in relation to the unique characteristics of this pathogenic mycobacteria [28].

Although specific roles of the VapBC TA systems in general are still not fully understood, there recently has been some evidence that these are involved in

regulating cellular processes, particularly in Mycobacteria[24]. These include cellular processes such as bacterial growth, stabilisation of genomic regions, defence against invading pathogens, nitrogen fixation, biomass production, induction and control of stress response, and bacterial persistence [29]. In *E. coli*, by inducing a range of Type II TA systems it was observed that bacterial persistence was significantly increased, thus indicating a major role of TA systems [19]. It is speculated that understanding and possibly exploiting these TA systems in *M. tuberculosis* may lead to the development of effective and efficient treatment options.

1.2.1 PIN Domains

The VapBC TA systems are defined by their toxin (VapC) component containing a PilT-N terminus (PIN) domain [30]. PIN domains are small proteins of approximately 130 amino acids long. These PIN domains have since been discovered in bacteria, archaea and eukaryotes which has generated the thought that these domains have roles in cellular maintenance [31]. While the amino acid sequence for the PIN domain family has very low identity, four amino acids are strictly conserved and form the enzymatic active site of the three-dimensional structure (Asp, Glu, Asp, Asp) (Figure 1.5)[32, 33]. Approximately 95% of PIN domains exist as single domain proteins, while the remaining 5% are fused with AAA+ ATPase domains, TRAM and KH domains [34].

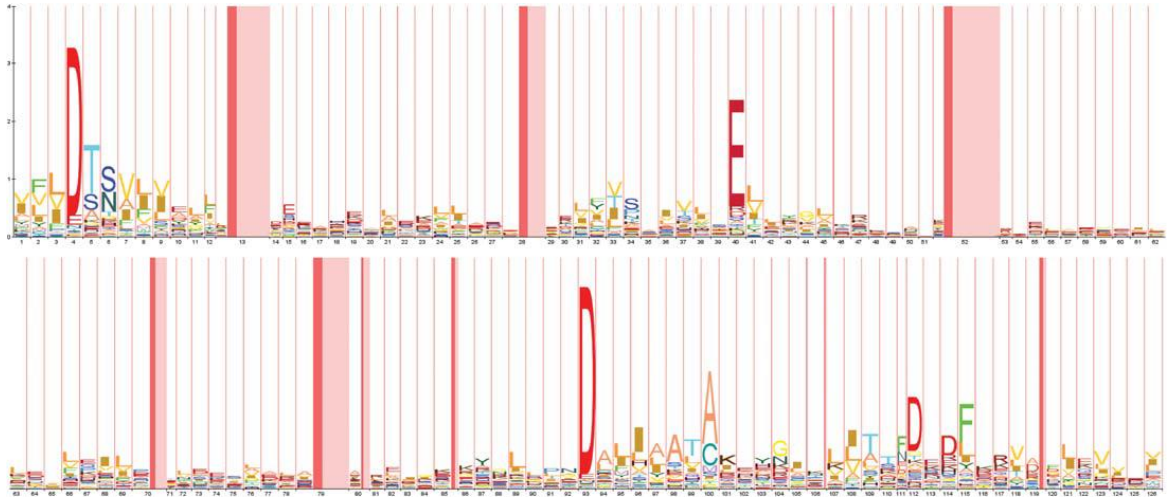


Figure 1.5: Hidden Markov Model (HMM) of PIN domain proteins. Of importance is the four conserved residues D, E, D, D that are conserved at strict positions as indicated by the height and width of each letter. Generated from pFam, [35].

Structurally, PIN domains have been characterised with alpha-beta-alpha (αβα) sandwich topology, with the conserved amino acid residues composing the active site responsible for exerting the toxic effect (Figure 1.6) [36]. While the four conserved amino acids vary in sequence distance, spatially they are positioned closely to enable the binding of Mg^{2+} ions and form the active site of the toxin nuclease (Figure 1.6) [31]. To date, all characterised PIN domain toxins are active as dimers. This dimerisation occurs along the long-axis of the PIN structure, positioning the active sites in a groove, enabling their toxic effect [32].

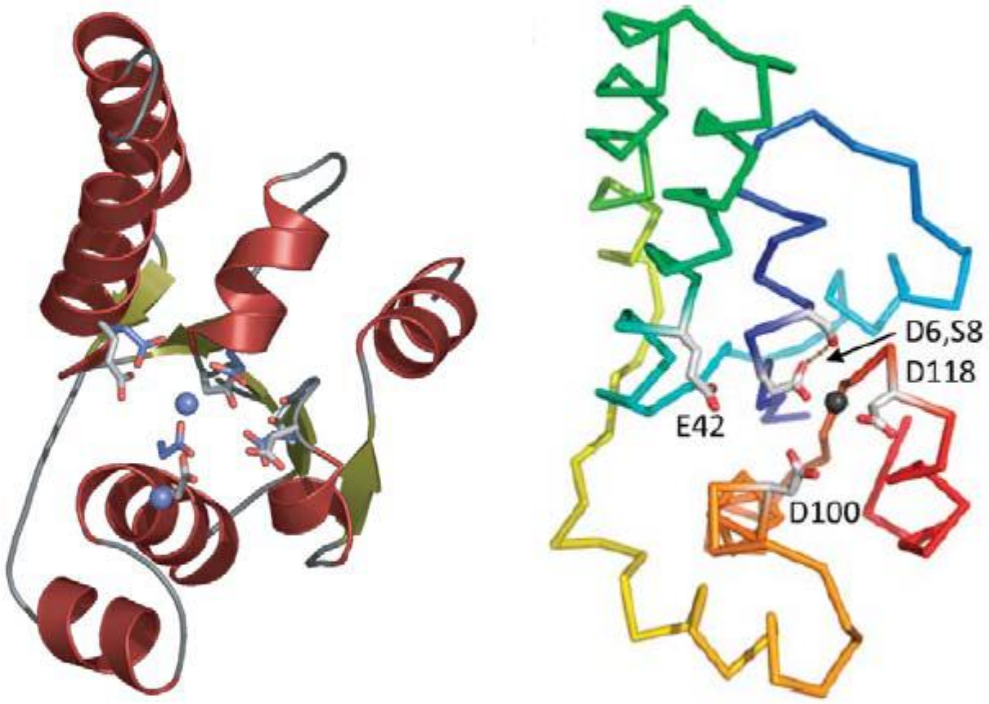


Figure 1.6: The structure of PIN domains found in PyrobaculumaerophilumPAE2754. Figures were made with Pymol (http://www.pymol.org/) using PDB 2FEI [31, 32].

PIN domain toxins are RNAses that target RNA by specific mechanisms, with varying PIN domain proteins targeting tRNA, rRNA or mRNA transcripts [19, 29]. The *M. tuberculosis* genome has a rich GC content of 66%, thus the observation that PIN domain toxins target and cleave GC rich regions in mRNA is no surprise [19]. How the PIN domain toxins specifically recognise their cleavage targets remains enigmatic, however a few studies have identified a stable hairpin motif which creates double stranded RNA (Figure 1.7) [29]. A consensus sequence in *M. smegmatis* has been identified following sequencing where up-stream of the hairpin loop is AUA(A/G) and other conserved nucleotides [37].

An identified rRNA targeting PIN domain toxin, VapBC20, specifically cleaves the Sarcin-Ricin loop. This inhibits translation as the Sarcin-Ricin loop is crucial for the assembly of the functional group of the 50S ribosomal subunit [38].

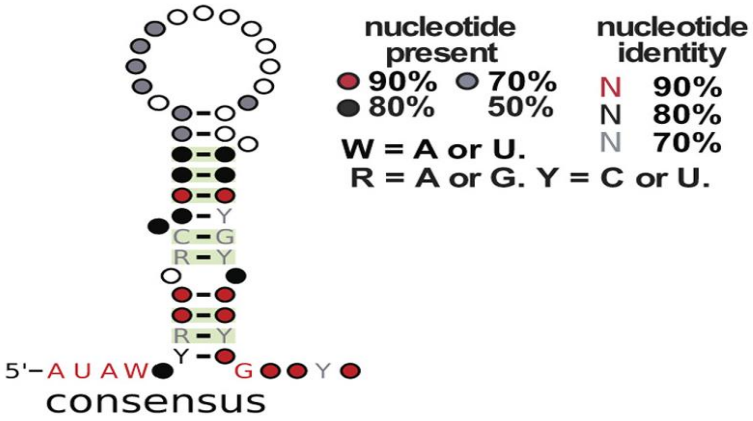


Figure 1.7: Schematic diagram outlining the conserved nature of the nucleotides in the hairpin structure in PIN domain toxin RNA substrates (figure taken from McKenzie, Robson [29]).

While no definitive studies into the characteristics of the type II antitoxins have been conducted to date, there are a few observations. An extensive search and analysis of the TBDB (Tuberculosis Data Base) revealed that the 50 VapBC antitoxins in *M. tuberculosis*, are generally half the length of the toxin protein that it inhibits, with lengths varying between 40% to 60% of the toxin component. Other features of antitoxins appear to be the presence of an arginine (R) amino acid residue in the last approximately 5 residues. In many structures of TA complexes, this positively charged side chain interacts with the negative side chains of the conserved residues forming the active site of PIN domain proteins [28].

1.3 Helicase Proteins

In order for the PIN domains to be able to execute their toxicity on their associated targets, the RNA sequence must be exposed, typically this is conducted by helicase proteins. Helicase proteins target either DNA or RNA and unwind the helices in order to expose the strands. Helicase enzymes require the hydrolysis of ATP for energy to unwind the strands. The helicase superfamily I has long been known to be comprised of 2 independent domains linked together by a non-conserved flexible linker region [39]. The N-terminal domain, and the YjhR C-terminal domain [39].

1.3.1 PhoH Domains

Very little is known about PhoH domain proteins with literature searches showing less than 5 publications. Bioinformatic investigations of PhoH proteins from *E.coli* and *Corynebacterium glutamicum* have identified similarities to helicase superfamily 1/2 proteins [39]. PhoH contains some of the conserved motifs in superfamily 1 helicase proteins of the N-terminal domain, but lacks the motifs found in the C-terminal domain (Figure 1.8). One such conserved motif is the Walker B of DNA and RNA helicases, also known as the DEAD box due to the residues present. This motif is characteristic of all ATPases and helicase proteins and is involved in the turnover of ATP [40]. The helicase motifs in the N-terminal domain of Superfamily I are required for ATP binding and hydrolysis, and intramolecular rearrangements [39]. It was identified that PhoH also appears to have 2 sequence motifs unique to PhoH; an arginine residue and a conserved sequence RGRTL, these have been experimentally verified to be involved with RNA recognition and binding(Figure 1.8)[34].The other conserved motifs found in PhoH are identified across proteins belonging to all 6 of the helicase families [34].

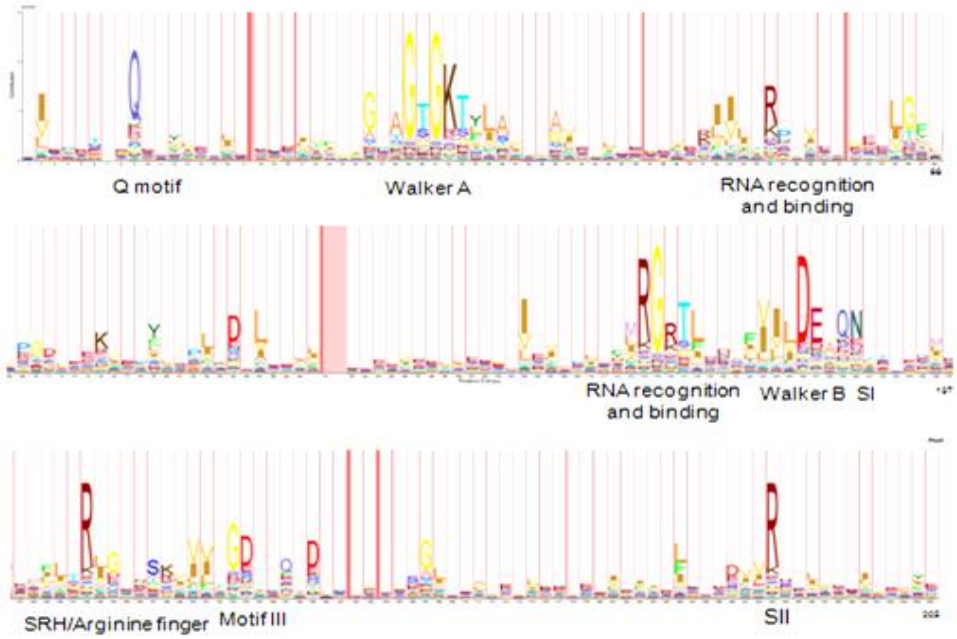


Figure 1.8: Hidden Markov model (HMM) of PhoH proteins. Conserved motifs are annotated as described in [32]. Generated from Pfam [35].

Like superfamily 1 helicase proteins, PhoH from *C. glutamicum*, the only PhoH structure in the PDB, shows a strong beta strand topology[41].

Because of these similarities, PhoH is likely to have helicase activity. It was reported when PhoH is found fused to a PIN domain protein to have unwinding activity (See 1.4 PhoH2 fusion protein) [34]. It has been speculated that PhoH proteins have a role during conditions of phosphate starvation [42]. In *E. coli* phosphate starvation induces a change in metabolism and gene expression to adapt to the conditions. This has been observed as a trigger of bacterial virulence in Mycobacteria, although the details are still not well known[43, 44].

1.4 PhoH2 fusion protein

PhoH2 proteins, the protein family of interest in this study, is a fusion between a PIN domain (VapC) at its N-terminus and a PhoH domain at its C-terminus. These are linked together by a long flexible linker region that does not correspond to any known conserved domains (Figure 1.9).

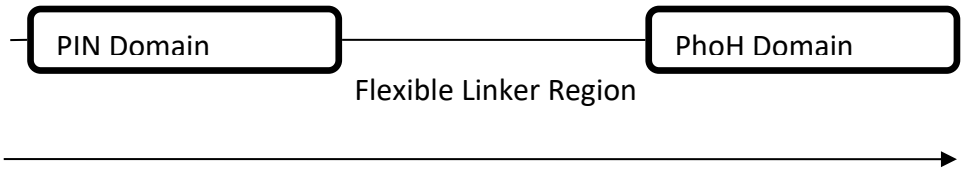


Figure 1.2: Schematic arrangement of the PIN and PhoH domains in $PhoH_2$ from C. thermarum.

PhoH2 from Mycobacteria has been biochemically characterised as an ATP-dependent RNA helicase and degradation protein specifically targeting RNA sequences with the 5' sequence AC[A/U][A/U][G/C] in the presence of ATP. The two unique motifs identified in PhoH not observed in any of the other 6 helicase families were found to be essential for unwinding activity [34]. In Mycobacteria, it was found that PhoH2 is posing as a variation of a type II TA system in which a small protein antitoxin is located upstream and overlapping the start of the PhoH2

gene. This small antitoxin was required to gain soluble protein expression of PhoH2 from *M. tuberculosis* but was not required to gain soluble PhoH2 protein from *M. smegmatis.* The presence of antitoxins upstream of the PhoH2 gene is limited to Mycobacteria as none have been identified in other organisms to date. This short antitoxin when expressed with PhoH2 from *M. tuberculosis* prevents an otherwise toxic effect that PhoH2 expression has on the growth of *M. smegmatis*[34].

Extensive efforts in the Proteins and Microbes lab (University of Waikato) have been unsuccessful in producing structural data to a high resolution using PhoH2 proteins from *M. tuberculosis*, *M. smegmatis* or *Thermobispora bispora*. While modest resolution has been achieved with a mutant of *T. bispora*, the 3-dimensional structure and therefore, insights into the functioning of PhoH2 proteins and how an RNase fused with an RNA helicase is coordinated, remains enigmatic.

1.5 *Caldalkalibacillus thermarum*

The genome of *Caldalkalibacillius thermarum* (*C. thermarum*) TA2.A1 has been sequenced (GenBank: AFCE00000000.1) [45] and a homologue of PhoH2 was identified using BLAST with a 39 % sequence similarity to PhoH2 from *M. tuberculosis* and *T. bispora*.

C. thermarum TA2.A1 is a thermophilic alkaphile, found to be growing in hot springs around the world, and locally in Te Aroha, Waikato, New Zealand [46]. This environment has extreme conditions and as such *C. thermarum* TA2.A1 grows optimally at 65 °C and at pH 9.5 [46, 47]. PhoH2 from *C. thermarum* is particularly attractive as a research target due to its lack of corresponding antitoxin suggesting a lesser toxic effect on cell growth than typically found with other PIN domain operons. The lack of antitoxin protein has been defined by the absence of an arginine residue in the approximately last 5 residues prior to the beginning of the PIN domain and the absence of start sites in frame with a stop codon overlapping the start of the PhoH2 gene. Secondly, other research labs have had success in studying large and difficult proteins to solve structures and other biochemical

attributes from *C. thermarum*. A recent example of this is the study of ATP synthase conducted by a research group at the University of Otago in 2005 [48]. Prior to studying the ATP synthase from *C. thermarum*, the structural investigations into this protein from *E.coli* and *Bacillus sp. PS3* had failed in providing structural features key to the biochemical function of the enzyme [48]. Following solving the structure of ATP synthase from *C. thermarum*, this research group has gone onto using molecular modelling to solve the structure of similar ATP synthase enzymes from other organisms [49].Thirdly, *C. thermarum* poses as an alternate thermophile to *T. bispora* that led to moderate success in crystallographic studies. Therefore, *C. thermarum* poses as an attractive candidate for further structural studies of PhoH2.

1.6 Objectives

The overall aim of the following research is to determine the structure and three-dimensional characteristics of the novel protein PhoH2 from *C. thermarum*. The characterisation of PhoH2 was carried out in terms of the following objectives.

- Clone, express and purify PhoH2 protein from *C. thermarum.*
- Conduct biochemical assays to characterise the activity of PhoH2 from *C. thermarum.*
- Using purified protein, grow diffraction quality crystals of PhoH2.
- Solve the structure of PhoH2 from *C. Thermarum* by X-ray crystallography.

2 Materials and Methods

2.1 DNA Manipulations

2.1.1 Genomic DNA extraction

2.1.1.1 Media and Growth Conditions

Caldalkalibacillus thermarum TA2.A1 was grown routinely in TA2.A1 classic media containing 2.5 mL of 200x Dictyoglomus trace elements (2.8 mg $ZnSO_4$, 0.38 g $MgCl_2.6H_2O$, 0.05 g $CaCl_2$, 0.039 g $Fe(NH_4)_2(SO_4).6H2O$, 1 g Na_2CO3, 2.9 mg $CoCl_2.6H_2O$, 2.4 mg $NaMoO_4.2H_2O$, 0.17 mg Na_2SeO3 and 2.0 mg $MnCl_2.4H_2O$, pH 9.5) at 65 °C with agitation at 200 rpm.

E. coli strains were grown in Luria broth (LB); 1 % peptone, 0.5 % yeast, 0.5 % NaCl) at 37 °C with agitation at 200 rpm, and supplemented with appropriate antibiotic, Kanamycin 50 mg/mL.

2.1.2 Vector DNA extraction from E coli

A Qiagen Plasmid extraction kit was used to extract DNA from *E. coli* cells following growth overnight at 37 °C in LB medium.

2.1.3 Agarose Gel Electrophoresis

One % Agarose Gel electrophoresis was used to separate DNA fragments. A 1 kb PLUS DNA Ladder was included to provide a standard of DNA band sizes (Invitrogen, USA). Samples were mixed with 10x DNA loading dye (50 % glycerol, 0.4 % bromophenol blue) prior to loading. Gels were prepared using 1x SYBRSafe™ DNA gel stain (Invitrogen, USA), and run in 1x TAE buffer (40 mM Tris acetate, 2 mM EDTA) at 100 V for 40 minutes. Gels were visualised using a blue light box (Invitrogen, USA) and images captured.

2.1.4 DNA Quantification

Quantification of DNA was done using a Nanodrop ND-1000 Spectrophotometer (Nanodrop Technologies, USA). The Nanodrop quantifies DNA by measuring absorbance of DNA at 260 nm, and the purity of samples determined by the

260/280 and 260/230 ratios indicating respective protein and carbon contamination.

2.1.5 PCR

PCR reaction temperatures were optimised by a gradient of temperatures between 5 °C above and below the primers calculated T_m.

2.1.5.1 Primers

All primers were a minimum size of 20 bp in length, with primers used for ligase-based cloning containing flanking restriction enzyme sites with an additional 2-4 bases to increase efficient cleavage. All primers were supplied by IDT (USA) and dissolved to a 100 µM concentration in 1x TE buffer, then further into a working concentration of 10 µM.

2.1.5.2 PCR for amplification of PhoH2 from Genomic DNA for Vector inserts

Table 2.1and Table 2.2 describe the PCR amplification methods and cycling parameters used in this study.

Hot FirePol DNA Polymerase	KAPA HiFi™ HotStart	Taq Polymerase
1 x Buffer B1	1x KAPA 5x HIFI GC Buffer	1x PCR buffer
0.05 U HOT FIRE pol	1 U KAPA HiFi Hotstart DNA Polymerase	2.5 U Taq DNA Polymerase
2.5 mM $MgCl_2$	10 mM KAPA dNTP Mix	1.5 mM $MgCl_2$
200 µM dNTP Mix	0.3 µM Each Primer	0.2 mM dNTP Mix
0.3 µM Each Primer	1-10 ng DNA template	0.5 µM Each Primer
1-10 ng DNA template		20-120 ng DNA Template

Table 2.1: PCR reactions used in the following study. Hotfire pol used for cloning. KAPA HiFi Hotstart used when cloning has been difficult. Taq Polymerase, a non-proofreading Polymerase used for screening colonies.

Hot FirePol DNA Polymerase	KAPA HiFi™ HotStart	Taq Polymerase
Cycling conditions (min:sec) x 29 cycles		
95 °C 15:00 (initial denaturation)	95 °C 2:00 (initial denaturation)	95 °C 2:00 (initial denaturation)
95 °C 0:20	98 °C 0:20	95 °C 0:15
T_m * 0:30	T_m * 0:15	T_m * 0:30
72 °C 1:00**	72 °C 0:45**	72 °C 0:45**
72 °C 5:00 (final extension)	72 °C 5:00 (final extension)	72 °C 5:00 (final extension)

Table 2.2: Cycling conditions for each of the PCR reactions used.

*T_m determined by gradient PCR
**Elongation time dependent on length of insert

	Primer Sequence
C. thermarum PhoH2 Cloning Forward	TTGAAAAAAATATATGTTCTGGATACCAATGTTCT
C. thermarum PhoH2 Cloning Reverse	GTTCCCCTCTGGCCCAACTTGCTGCCCACGTTCTT
C. thermarum PhoH2 PstI Forward	AAAACTGCAGATGAAAAAAATATATGTTCTGGAT
C. thermarum PhoH2 XbaI Forward	CTAGTCTAGAATGAAAAAAATATATGTTCTGGAT
C. thermarum PhoH2 HindIII Reverse	CCCAAGCTTAAGAACGTGGGCAGCAAGTTGGGC
C. thermarum PhoH2 HindIII Reverse with Native Stop	CCCAAGCTTTTAAAGAACGTGGGCAGCAAGTTGGGC
pET28b-Htb Sequencing 4877 Forward	CATCGGTGATGTCGGCGATA

T7 Forward	TAATACGACTCACTTTAGGG
T7 Reverse	GCTAGTTATTGCTCAGCGG
pET28b-Htb Insert a Forward	TTAAGAAGGAGATATACTGCAGATGAAAAAAATAT
pET28b-Htb Insert b Forward	TAATTTTGTTTAACTTTAAGAAGGAGATAT
pET28b-Htb Insert c Forward	CTAGTCTAGAAATAATTTTGTTTAACT
pPROEX-Htb 30 EcoRI Forward	CGGAATTCAAATGAGTCATACACATAAAAATAAAA
pPROEX-Htb 18 EcoRI Forward	CCGGAATTCAAATGTATATTTTTCCAAGGCCAGTT
pET28b-Htb-pstI 30 EcoRI Forward	CGGAATTCATGAGTCATACACATAAAAATAAAACA
pET28b-Htb-pstI 18 EcoRI Forward	CGGAATTCATGTATATTTTTCCAAGGCCAGTTGAT

Table 2.3: Primers designed for cloning and sequencing C. thermarum *PhoH2, ordered from IDT technologies.*

2.1.6 Purification of PCR product in solution

A commercial High Pure PCR Product Purification Kit (Roche Applied Science, Switzerland) was used to purify PCR or enzymatic reaction products following pooling of 15 µL reactions.

2.1.7 Restriction Enzyme Digest

Restriction enzyme digests were performed by the Cutsmart double digest system following the manufacturer's instructions. Buffers used were chosen as recommended by the manufacturer to maximise the efficiency of the DNA digest. Incubations were carried out at 37 °C for three hours. Restriction enzymes were purchased from Invitrogen (USA). Double digestions were used when the reaction buffers were compatible for both enzymes. Following incubation, reactions were purified according to the method in section 2.1.6.

2.1.8 DNA Ligation

DNA Ligation reactions were conducted following the manufacturer's recommendations. One uL of T4 DNA Ligase (Invitrogen, USA), 1x Ligase buffer and a 1:3 ratio of insert:vector, and MilliQ H2O to a final volume of 20 µl. DNA Ligase reactions were incubated at 18 °C overnight.

2.1.9 DNA Transformation

2.1.9.1 Preparation and transformation of chemically competent *E. coli*.

The following method was used for the preparation and transformation of both DH5α and BL21 chemically competent *E. coli* strains.

A single colony of DH5α or BL21 cells was cultured overnight in 10 mL of LB media at 37 °C, 200 rpm. One mL of this starter culture was used to inoculate 100 mL of LB at 37 °C at 200 rpm until OD_{600}= 0.5-0.7 was reached. The 100 mL culture was then split into 2x 50 mL and chilled on ice for 30 minutes followed by centrifugation at 4000x g for 10 minutes at 4 °C. The cell pellet was then resuspended in 10 % of the original culture volume of ice cold TSB media. Aliquots of 100 µL were then snap frozen in liquid nitrogen and stored at -80 °C.

Aliquots of chemically competent cells were thawed on ice and 1-10 µL of the desired plasmid or ligation was added and gently pipetted. Cells were left on ice for 30 minutes incubation. Aliquots were then heat shocked at 42 °C for 45 seconds and then immediately put back on ice for 2 minutes. Following the addition of 1 mL of SOC media at room temperature aliquots were then incubated at 37 °C for 60 minutes.

The resulting cultures were then centrifuged at 4000xg on a bench top centrifuge for 30 seconds. The resulting pellet was then resuspended in 100 µL of SOC media and spread on pre-dried agar plates supplemented with the appropriate antibiotic for incubation at 37 °C overnight.

2.1.9.2 Glycerol Stocks

Glycerol stocks for the long term storage of bacterial cells were made in the ratio of 1:1 with sterile 50% glycerol. Glycerol stocks were stored at -80 °C.

2.1.10 DNA Sequencing

The correct integration of the *PhoH2* gene insert into the expression plasmid was confirmed by Sequencing at Massey Genome Service, Massey University, Manawatu Campus. Plasmid samples were sent at 250 ng with 4 µM primers, up to a final volume of 20 µL with MilliQ water.

2.2 Protein Expression and Purification

2.2.1 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) Protein Analysis

SDS-PAGE gels were comprised of a stacking gel (5 % acrylamide, pH 6.8) above a resolving gel layer (12 % acrylamide, pH 8). Acrylamide was added from a stock solution of 30 % acrylamide/bis (BioRad Laboratories, USA). Gels were prepared in a Hoefer multiple gel casting system. Gels included 0.1 % (w/v) SDS, 0.05 % (w/v) ammonium persulfate (APS) and 0.05 % (v/v) Tetramethylethylenediamine (TEMED).

Protein samples were mixed with 4x SDS loading buffer in a 3:1 ratio (4 % SDS, 20 % glycerol, 250 mM Tris HCL pH 6.8, 10 % β-mercaptoethanol, 0.025 % (w/v) bromophenol blue) and heated for 5 minutes at 95 °C. Following loading onto SDS-PAGE, gels were run at 70 V until the dye front had passed through the stacking layer, then at 150 V until through the resolving layer in 1x SDS-PAGE running buffer (25 mM Tris, 250 mM glycine, 0.1 % (w/v) SDS).

2.2.3 Coomassie Blue Staining for Protein Gel electrophoresis

Following electrophoresis, gels were individually incubated in 100 mL of Coomassie Blue stain (0.05 % Coomassie® R-250 in 25% isopropanol, 10% acetic acid). Gels in staining solution were micro waved for 30 seconds, then left on an orbital gel shaker at room temperature for 15 minutes. Coomassie stain was decanted, and the gel rinsed in MilliQ water before incubated in 100 mL destain solution (10% acetic acid) until desired background achieved.

2.2.4 Measurement of Protein concentration

The concentration of protein was measured using the Nano-drop ND-1000 spectrophotometer. This measures absorbance at 280 nm and the protein concentration is calculated using the Beer-Lambert equation: $A=\varepsilon.c.l$

A = absorbance at 280 nm

ε = theoretical molar extinction coefficient (M^{-1} cm^{-1})

c = concentration (M)

l = pathlength (cm)

The theoretical molar extinction coefficients (ε) were calculated by entering the amino acid sequence into the online programme ProtParam (http://web.expasy.org/protparam).

2.2.5 Concentration of Protein Samples

Fractions containing the Protein samples were pooled and spun in either 15 mL or 2 mL Vivaspin concentrators (Sartoris AG, Germany) with a molecular weight cut off of 10 kDa.

Concentrators were first equilibrated with 0.05 % (v/v) Tween-80 and suitable purification buffer. Pooled samples were added to the top reservoir and spun at 3300 g at 4 °C in 20 minute intervals until the desired concentration or volume was reached.

2.2.6 Small Scale Protein expression trial

Small Scale 100 mL LB cultures of *E. coli* DH5α or BL21 containing the plasmid for the desired protein expression were induced at OD_{600} = 0.4-0.6 by the addition of IPTG and grown overnight at 37 °C. Three mL of expression culture was spun down at 4000xg on a bench-top centrifuge for 1 minute.

The supernatant was discarded, and the pellet resuspended in 250 µL lysis buffer, 50 mM Tris pH 8.0, 200 mM NaCl and 20 mM Imidazole, unless otherwise stated. Following sonciation using an Ultra sonic processor XL 2020 (Misonix

incorporation, SA), the sample was centrifuged at 4000xg for 20 minutes at 4 °C. The supernatant was then loaded onto Ni^{2+} Sepharose beads pre-equilibrated with lysis buffer and incubated at room temperature for 15 minutes. Following light centrifugation, the supernatant was then discarded and the beads washed in 1 mL lysis buffer. Fifteen µL samples of the Whole Cell, Supernatant, Insoluble, Wash and resin were incubated at 95 °C for 5 Minutes in Qx4 quencher and then visualised on a 12 % SDS PAGE gel.

2.2.7 Purification of His - Tagged proteins using IMAC/Nickel chromatography

Large scale expression cultures (1 L) were lysed by sonication, and the resulting supernatant was filtered through 1.2, 0.45 and 0.2 µm Minisart filters (Sartorius AG, Germany). The Filtered supernatant was then loaded onto a 5 mL IMAC column (GE Sciences) following pre-equilibration with the appropriate lysis buffer in a 'drop-to-drop' method. The column was then connected to an AKTA Purifier™ FPLC system (GE Sciences).

The 5 mL column was washed with 25 mL of lysis buffer at a flow rate of 1 mL/min to remove any excess unbound proteins. Bound proteins were removed and collected in 2 mL fractions by an increasing gradient of 0-50% elution buffer (Lysis buffer + 1 M Imidazole) over 50 mL at a flow rate of 1 mL/min. Fractions corresponding to the protein were analysed by SDS-PAGE (See section 2.2.1).

2.2.8 Size Exclusion chromatography

Following IMAC chromatography, further purification of the desired protein was achieved by Size exclusion chromatography. IMAC fractions containing the desired protein were pooled, and concentrated (see section 2.2.5) until a final volume of 500 µL was reached. A S200 10/300 column was connected to AKTA Purifier™ pre-equilibrated in size exclusion buffer (50 mM Tris pH 8.0, 200 mM NaCl, 1 mM EDTA and 1 mM DTT). Protein was then loaded into pre-equilibrated loop through 0.2 µm filter. Following elution of the protein by size, fractions were collected in 0.5 mL volumes, and subsequently run on SDS-PAGE (See section 2.2.1) to identify desired protein and their corresponding chromatograph peaks.

2.3 Protein ATPase Assays

ATPase Assays were conducted to determine active populations of desired protein. A standard curve was determined using concentrations of Potassium dihydrogen phosphate ranging from 0 - 50 mM, prepared in assay buffer (20 mM Hepes, pH 8.0, 20 mM NaCl) and 1 mM $MgCl_2$. Control reactions were included that used an assay buffer that contained 2 mM EDTA.

Pilot studies were performed to determine appropriate ATP and protein concentrations. ATP ranged between 1 - 10 mM and protein concentration between 0.25 - 1 µM.

Reactions of 100 µL were pre-equilibrated at 37 °C for 10 minutes before enzyme was added. The assay reaction was left for a further 10 minutes at 37 °C. Thirty µL of malachite green stop solution (1.5% ammonium molybdate, 0.18% Tween-80, 0.1% malachite green, 14% sulfuric acid) was added to the reaction and held at room temperature for 10 minutes. Absorbance was read at 620 nm using Multiskan Go (Thermofisher Scientific).

3 Results

3.1 Cloning, Expression and Purification of *C. thermarum* PhoH2

3.1.1 Cloning and Expression

To achieve the first objective of cloning, expressing and purifying PhoH2 protein, the PhoH2 gene from *C. Thermarum* was amplified and cloned into plasmids pET28b-*Pst*I and pPROEX-HtB for protein over-expression in *E. coli*.

Firstly, C. thermarum DNA was successfully extracted according to methods by Olsson (2002).

Gradient PCR amplification from genomic DNA was conducted for optimum annealing temperature. Initial PCR used primers not containing restriction digesting sites, and resulting annealing temperature kept at 56 °C for subsequent amplifications.

Many attempts were made to ensure successful cloning and thus expression of PhoH2 *C. thermarum*. Initially, *Pst*I restriction endonuclease was used at the forward position, and *Hind*III restriction endonuclease at the reverse position for cloning into pET28b-*Pst*I. Cloning was successful however DNA gels showed a smaller sized gene than expected following a post-cloning screen using the Cutsmart Double Digest system with *Pst*I and *Hind*III (Figure 3.1).

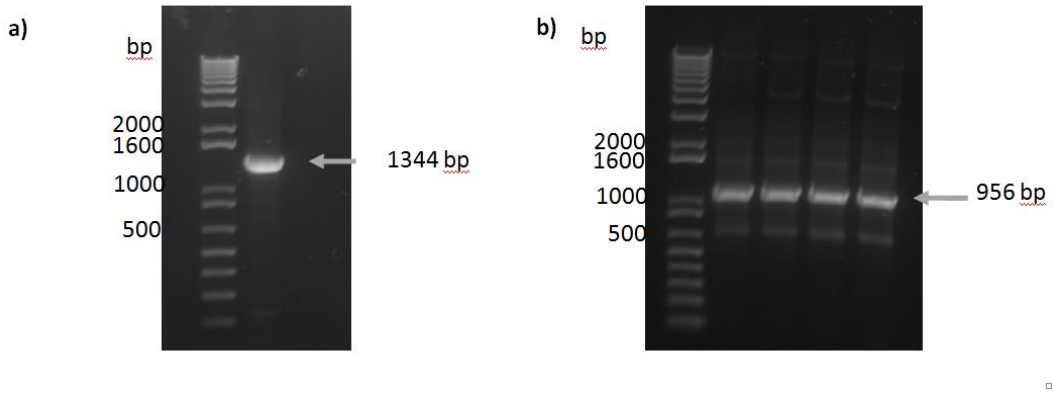


Figure 3.1: 1% agarose gels of the PhoH2 gene a) prior to restriction endonuclease digestion, and b) following restriction endonuclease digestion

Upon investigation to the cause of this smaller band it was discovered that *phoH2* from *C. Thermarum* contained a cutting sequence for *Pst*I internally. Therefore, when the gene construct was digested by the Cutsmart double digest system prior to ligation and transformation the gene was also being cleaved internally.

Thus, the forward restriction endonuclease was changed to *Xba*I, and the reverse kept as *Hind*III. The entire PhoH2 gene was successfully cloned into pET28b-*Pst*I plasmid (Figure 3.2) and transformed into BL21 *E.coli* in preparation for protein expression, after confirmation of correct insertion by sequencing at the Massey Genome Service (MGS).

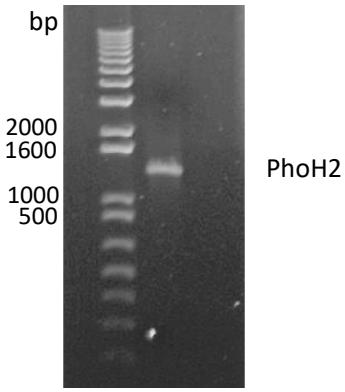


Figure 3.2: PhoH2 insert following RE digest dropout to confirm the presence of PhoH2 in BL21 E.coli

Small scale temperature expression trials were conducted post-induction with IPTG at 18 °C, 25 °C, 30 °C and 37 °C overnight alongside empty pET28b--*Pst*I plasmids. No expression of the PhoH2 *C. thermarum* protein was observed with any of the expression temperatures (Figure 3.3).

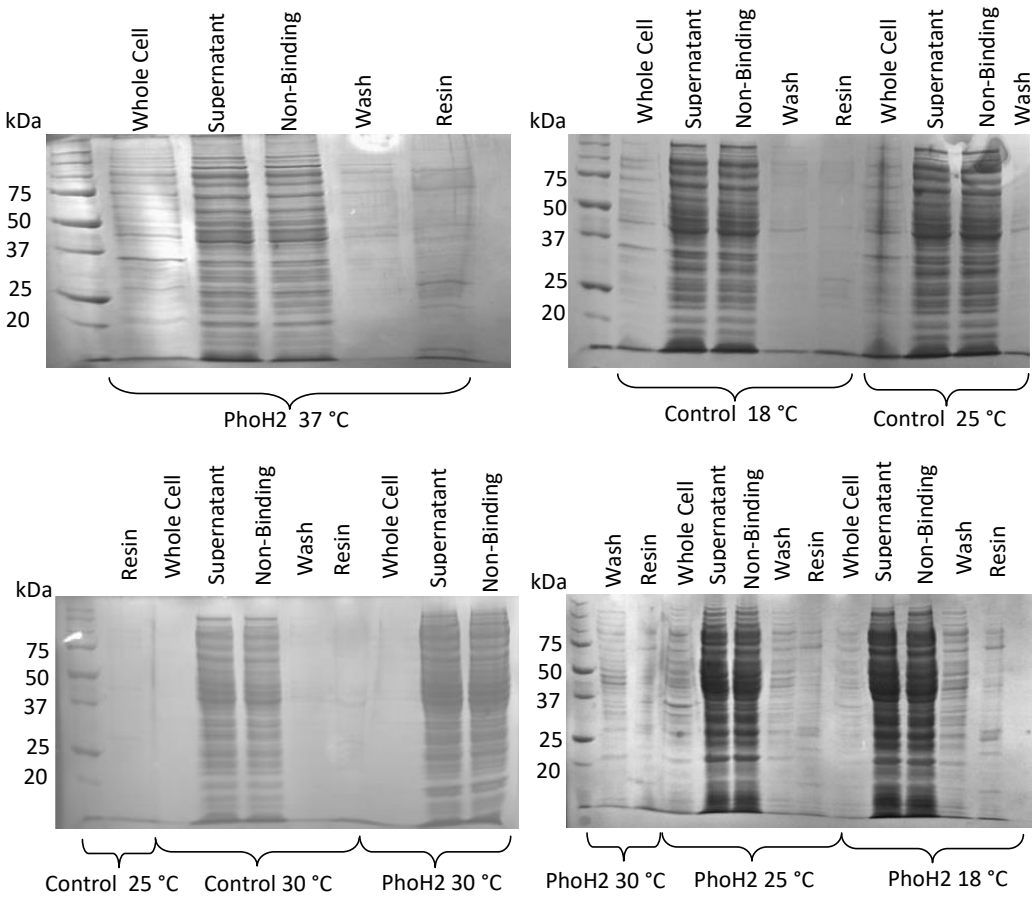


Figure 3.3: 12 % SDS-PAGE gels showing the small scale expression trials of PhoH2 cloned into pET28b-PstI for expression in E. coli BL21 at 37 °C, 30 °C, 25 °C and 18 °C. Controls containing empty pET28b-PstI were tested at 30 °C, 25 °C and 18 °C in parallel. Expected band size 56 kDa not present in control or PhoH2 samples.

Despite numerous attempts to obtain expression of PhoH2 protein, including Terrific Broth media no protein was observed. Revisiting the cloned construct sequence revealed the *Xba*I site in pET28b-*Pst*I was downstream of the ribosome binding site (RBS). Thus, when the plasmid was digested with *Xba*I, and the PhoH2 gene product ligated into pET28b-*Pst*I, the RBS was removed, reducing the ability of *E. coli* BL21 to express the inserted PhoH2 protein despite the phoH2 gene sequence being correct.

Upon this discovery, three new cloning constructs were designed to overcome this issue (Figure 3.4). Construct A used the same restriction endonucleases (*Xba*I and *Hind*III), and therefore the *PhoH2* gene product used previously was directly inserted into the pPROEX-Htb plasmid for expression with an N-terminal His-tag in

either Dh5α or BL21 *E. coli*. While this was straight-forward to clone, and could be used to check for possible expression of PhoH2 protein, the C-terminus did not house a native stop codon, therefore an additional 8 amino acids were included at the C-terminus upon expression. Construct B removed these additional 8 amino acids by adding the native stop codon immediately following PhoH2. Construct C was more difficult to produce as 3 forward primers were used to re-introduce the RBS upstream of the PhoH2 start codon for re-cloning into pET28b-*Pst* for expression with a C-terminal His-tag in BL21 *E. coli.*

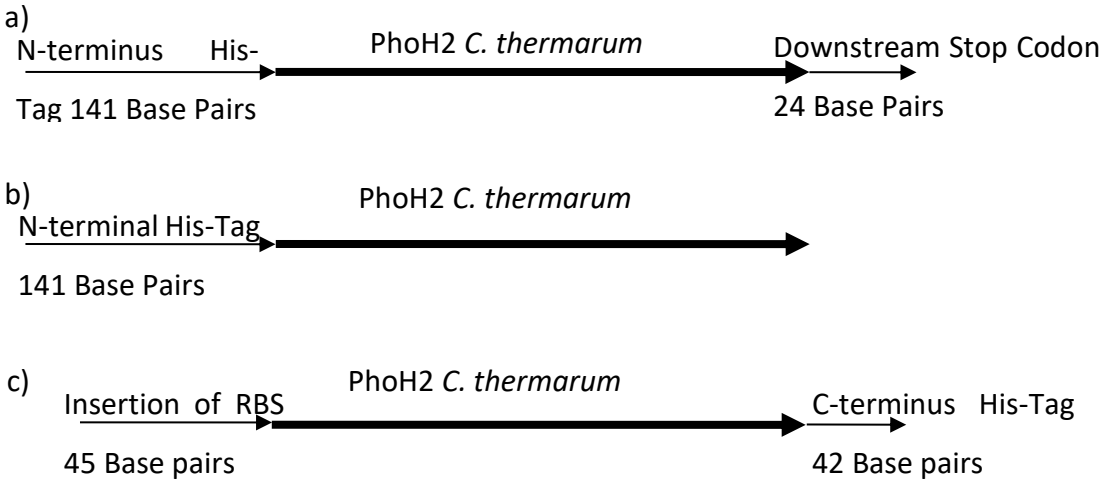


Figure 3.1: Schematic diagram of the 3 cloning constructs

Cloning Construct A

Whilst the cloning of the PhoH2 gene into pPROEX-Htb plasmid (Construct A) was successful as determined by MGS sequencing, subsequent expression was problematic. Small scale expression trials initially showed a protein band of approximately the expected size (56 kDa) present in the whole cell and the resin samples for both trials conducted at 37 °C and 30 °C post-induction with IPTG, however it is notably lacking an enrichment of desired protein in the supernatant sample (Figure 3.5).

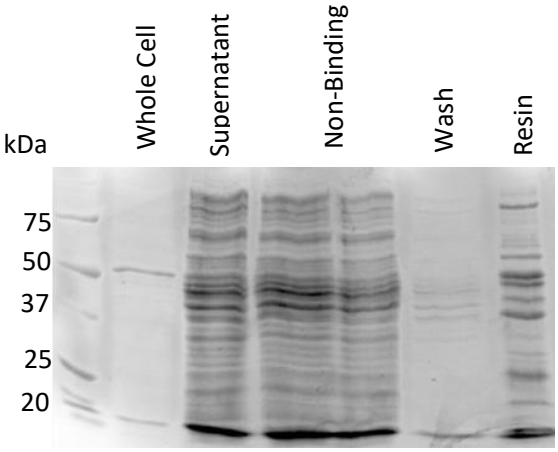


Figure 3.5: Small scale expression of PhoH2 in pPROEX-Htb plasmid from DH5α E. coli at 37 °C. Whole Cell, Supernatant, Insoluble, Wash 1, Resin.

Given the enrichment on the Ni resin large scale (1 L) cultures in LB media were conducted at both 37 °C and 30 °C. Following lysis of the cells by sonciation, the soluble fraction was loaded onto a Nickel column for IMAC chromatography thereby separating the proteins based on their binding affinity to the Nickel resin (Figure 3.6).

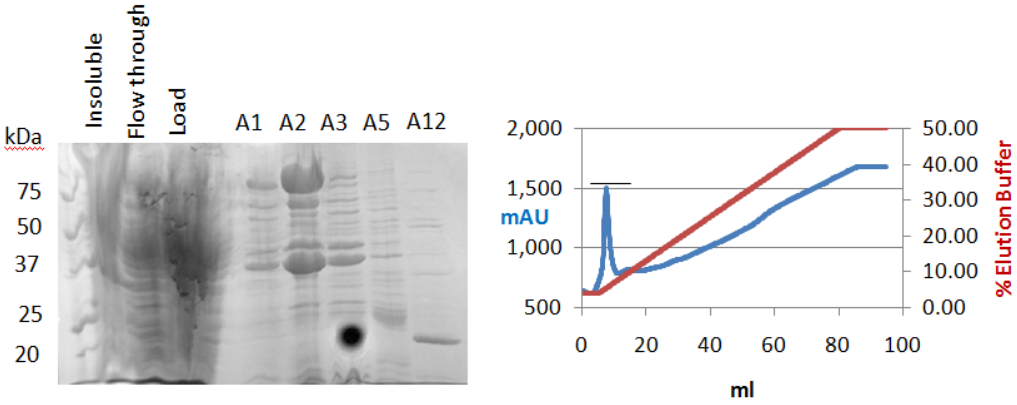


Figure 3.6: IMAC chromatography curve and corresponding 12 % SDS-PAGE of the fractions covering the main peaks as shown by the black line across the chromatography curve corresponding to fractions A1 – A12, grown at 37 °C.

The main peak came off the Nickel column very early in the elution gradient (fractions A1-A3) as seen in Figure 3.6. The corresponding SDS-PAGE (Figure 3.6) shows two main proteins present in this primary peak, one at approximately 80 kDa, another at approximately 48 kDa. There is also the presence of a slightly less

expressed protein, which was assumed to be PhoH2 at 56kDa. Fractions were also run of smaller secondary peaks at 13 mL (A5) and 27 mL (A12) respectively. No protein bands corresponding to the expected size of PhoH2 were clearly observed, suggesting that PhoH2 may not be successfully expressed and purified similarly to what was seen in the small scale test (Figure 3.5).

As these primary bands seen on the SDS-PAGE were not expected, a His-Tag stain was conducted on a SDS-PAGE with fractions taken from expression at both 37 °C and 30 °C following IMAC Ni chromatography. A control protein with a known His-Tag from the Proteins and Microbes Lab was also included (Pfu) (Figure 3.7).

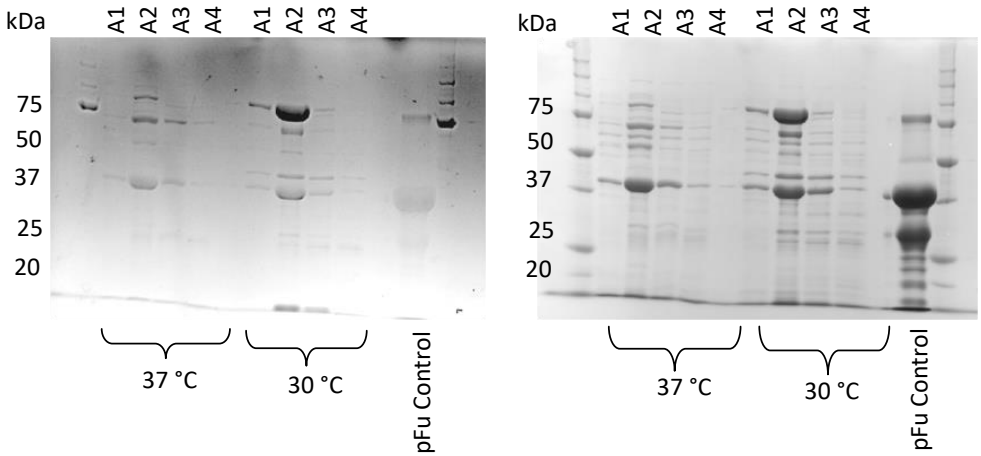


Figure 3.7: His-tag stain 12% SDS-PAGE, followed by a standard Coomassie Stain. pFu protein was used as a control for size (47 kDa) and presence of His-tag. Fractions A1-A4 for both expression temperatures (37 °C and 30 °C) were loaded respectively.

The His-tag stain showed that both the higher MW and lower MW bands and the band assumed to be PhoH2 contained a high percentage of histidine residues, indicating the likely presence of a His-tag and therefore the desired PhoH2 protein.

Subsequent purification using size exclusion chromatography (Figure 3.8: Size Exclusion Chromatography Curve and Corresponding 12% SDS-PAGE a) 37 °C and b) 30 °C (Figure 3.8) separated these populations of protein for individual analysis by ATPase assays to confirm which population may or may not correspond to PhoH2. The elution volumes were entered into a calibration equation deduced for

this SE column and were calculated to correspond to the predicted MW of PhoH2 if PhoH2 is oligomerising.

The ATPase assays showed a consistent lack of activity (figures not shown), therefore the two protein populations were prepared for Mass Spectrophotometry which was carried out by commercial Lab MS3 Solutions.

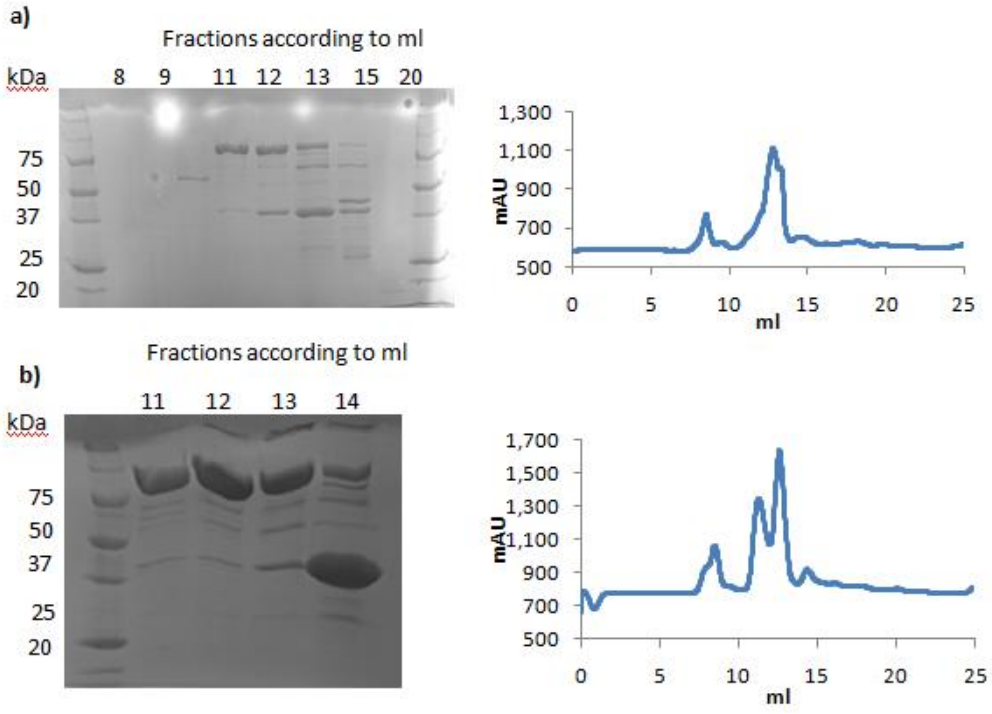


Figure 3.8: Size Exclusion Chromatography Curve and Corresponding 12% SDS-PAGE a) 37 °C and b) 30 °C.

The Mass Spectrophotometry results concluded that neither of these protein populations were PhoH2, but instead were identified as *E.coli* proteins Catalyse at 84 kDa and Lac Repressor at 38.5 kDa.

Cloning Construct B

Construct B, in which there is an added native stop directly at the end of the PhoH2 gene, was confirmed by MGS sequencing, and small scale expression trials were carried out in DH5α *E.coli* cells.

A similar pattern to the first construct (A) was observed with this second construct (B). Small scale expression trials over a range of temperatures (37°C, 30 °C and 28 °C) showed the presence of a PhoH2-sized band in the whole cell and resin samples

for all temperatures (Figure 3.9). However, large 1 L expression failed to produce visible PhoH2 band while again the *E.coli* Catalyse and Lac Repressor proteins were over-expressed. In the chromatography curves, a second minute peak was observed at approximately 60% Elution buffer (approximately 120 mM imidazole). Whilst very weak, initial SDS-PAGE shows this to be of consistent size to the expected size of PhoH2 (56 KDa).

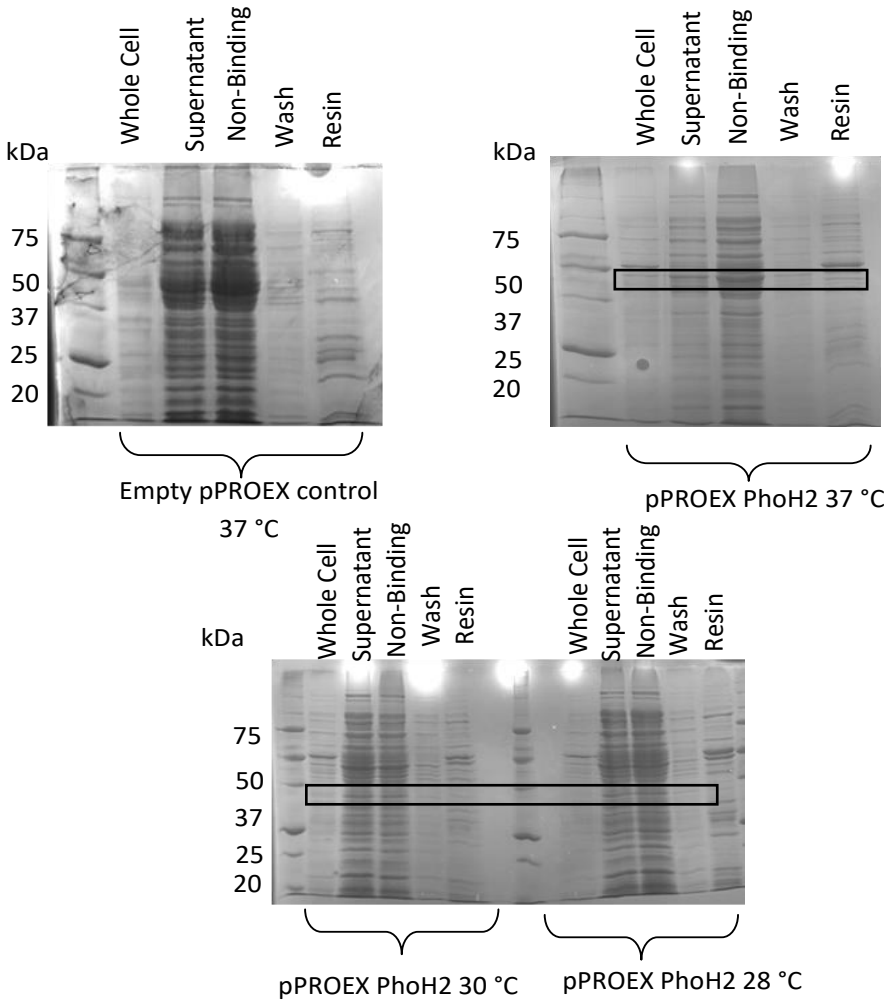


Figure 3.9: Small scale expression of PhoH2 in pPROEX-Htb plasmid in DH5α E.coli cells over 37 °C, 30 °C and 28 °C post-induction. The black boxes highlight the weak protein band of expected size present in Resin samples over all temperatures. Note there is no enrichment of PhoH2 in the 37 °C control (empty pPROEX-HtB).

3.1.1.1 Purification

Due to the above results being unsuccessful in yielding soluble protein multiple buffer conditions were explored. In order to carry out the prescribed biochemical characterisation assays, and eventually X-ray crystallography, the PhoH2 protein

is required to be as pure as possible, to attribute findings as a direct action of PhoH2 rather than other contaminant proteins.

Typically, purification of recombinant proteins requires the desired protein to remain in solution following the lysis or rupture of the cell. The soluble protein is then separated according to biochemical property such as size, charge or binding ability by several chromatography techniques.

Purification of the PhoH2 protein proved to be difficult, although the yield of PhoH2 appeared to be enriched in the whole cell. Upon lysis of the cell by sonication, PhoH2 was insoluble and was therefore retained in the pellet when centrifuged. An in depth buffer screen was conducted to determine the conditions in which PhoH2 remained in solution in order for downstream purification methods.

Several small scale trials were conducted in order to determine the solubility of PhoH2 and therefore the ability for PhoH2 to bind to Nickel chromatography resin. Buffers were varied in pH, with constant NaCl and imidazole concentrations (Table 3.1) with construct transformed into both DH5α (Figure 3.10) and BL21 *E. coli* cells (Figure 3.11).

50 mM MES pH 6.0, 200 mM NaCl, 20 mM Imidazole
50 mM HEPES pH 7.5, 200 mM NaCl, 20 mM Imidazole
50 mM Tris pH 8.0, 200 mM NaCl, 20 mM Imidazole
50 mM BisTris Propane pH 9.5, 200 mM NaCl, 20 mM Imidazole

Table 3.1: Composition of purification buffers varying pH conditions.

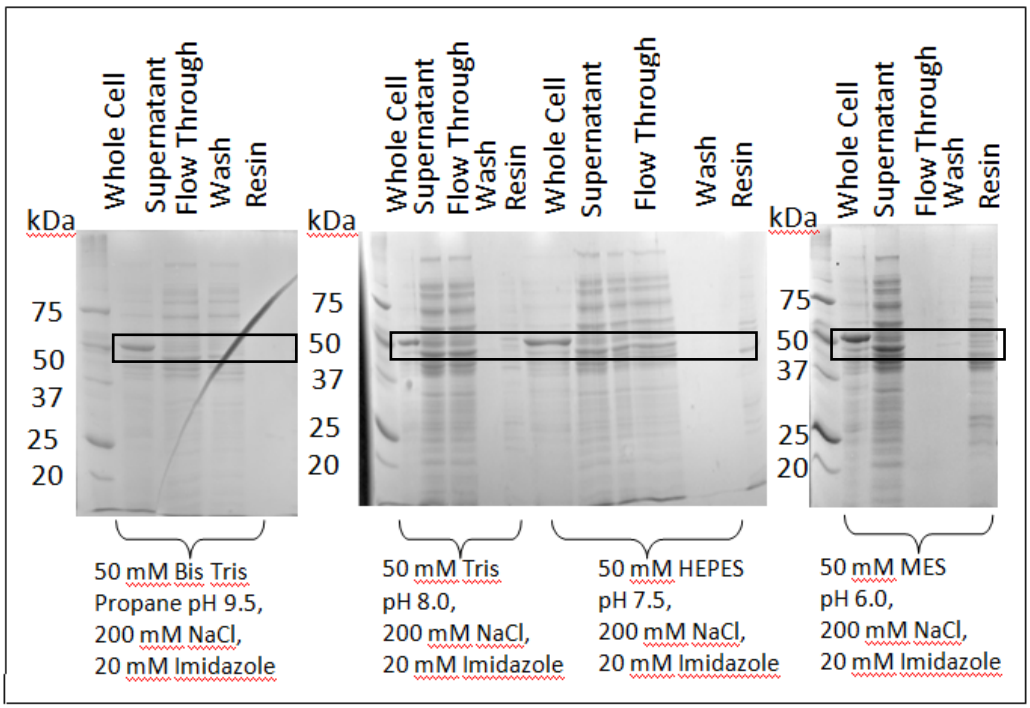


Figure 3.10: SDS-PAGE gels of small scale expression trials screening increased salt concentrations and the addition of 10 % glycerol to lysis buffers on PhoH2 expressed in DH5α E. coli cells.

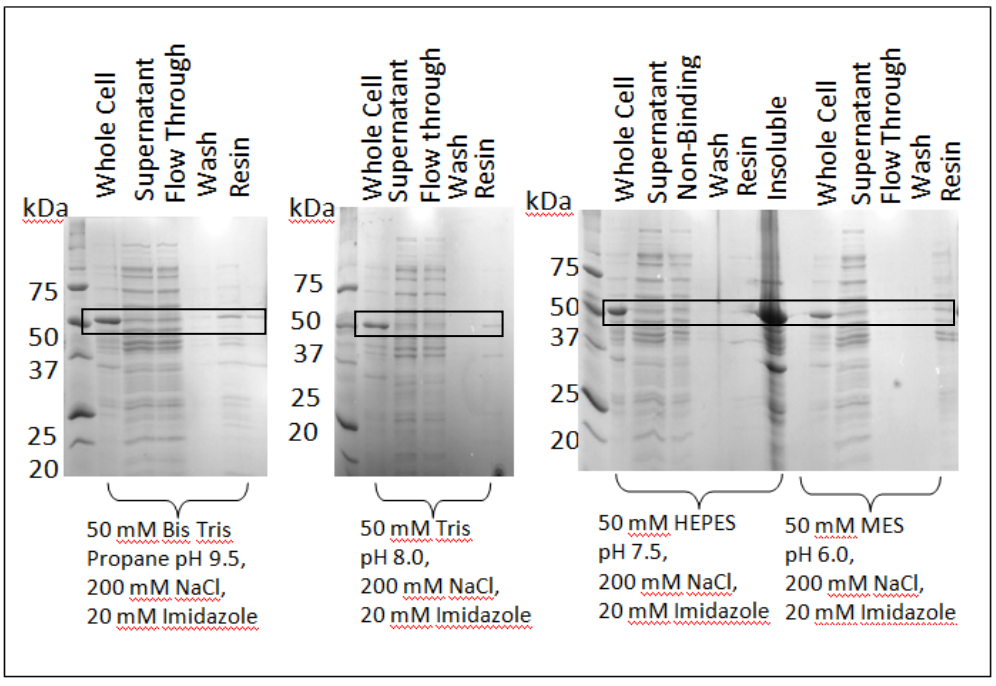


Figure 3.11: SDS-PAGE gels of small scale expression trials screening increased pH of lysis buffers on PhoH2 expressed in BL21 E. coli cells.

It appeared there was no significant improvement in the solubility of PhoH2 across the buffer pH range, however BisTris Propane pH 9.5 and Tris pH 8.0 buffers did show faint PhoH2 bands on the nickel resin, particularly with BL21 strain. Growth of the *C. thermarum* organism, appears in the literature to favour higher NaCl concentrations due to its halophilic properties, and with the purification of other *C. thermarum* proteins requiring glycerol to be included in the purification buffers [49]. A second buffer screen using Nickel pull down experiments investigated the solubility of PhoH2 in increased NaCl, both with and without 10% glycerol in attempt to increase the solubility and stability of PhoH2 upon cell lysis (Table 3.2). Results indicated enrichment of PhoH2 protein remaining in the insoluble pellet (Figure 3.12).

50 mM Tris pH 8.0, 1.5 % NaCl, 20 mM Imidazole
50 mM Tris pH 8.0, 0.5 M NaCl, 20 mM Imidazole
50 mM Tris pH 8.0, 1.5 % NaCl, 20 mM Imidazole, 10 % Glycerol
50 mM Tris pH 8.0, 0.5 M NaCl, 20 mM Imidazole, 10 % Glycerol

Table 3.2: Composition of purification buffers varying Salt and glycerol conditions.

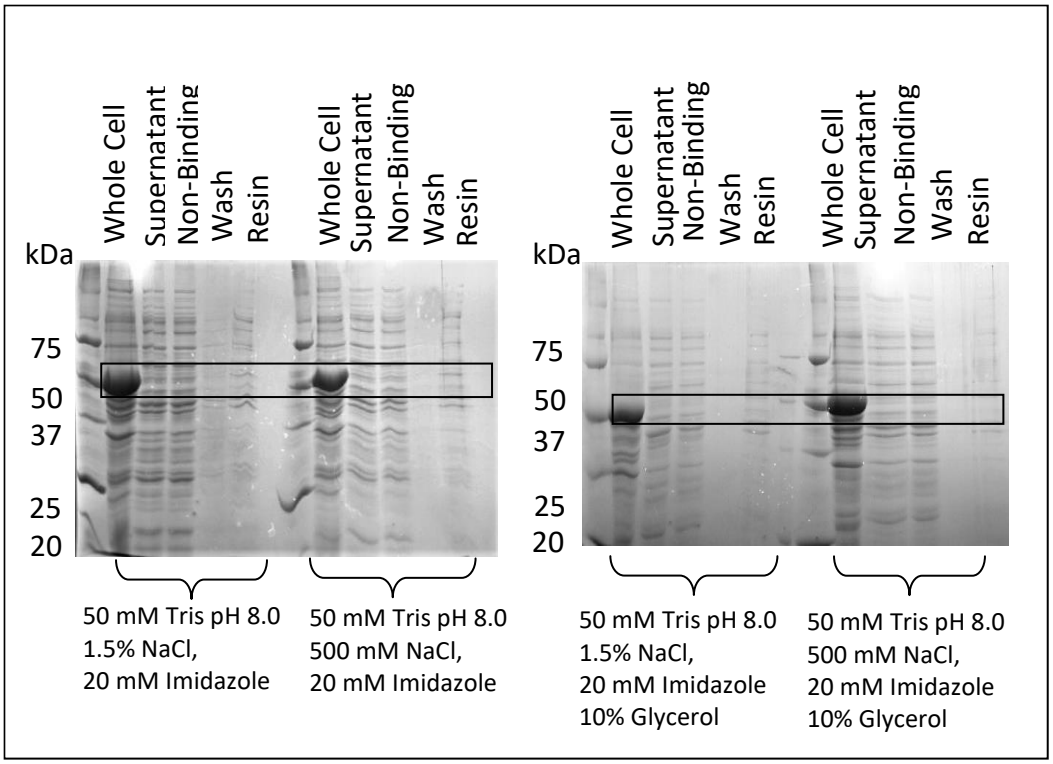


Figure 3.12: SDS-PAGE gels of small scale expression trials screening increased salt concentrations and the addition of 10 % glycerol to lysis buffers on PhoH2.

Due to the above attempts towards gaining soluble protein, consultation with Dr Emma Andrews and Prof. Vic Arcus regarding the previous refolding trials on the domains of PhoH2 from *M. tuberculosis*, and the time constraints for this thesis, refolding trials were not attempted.

Cloning Construct C

Construct C in which the PhoH2 gene is inserted into pET28b-*Pst*I with primers which also add the RBS ahead of the PhoH2 gene was kindly cloned by Dr Emma Andrews and transformed into *E.coli* BL21 cells. Dr Andrews also carried out initial small scale expression trials in LB media with IPTG induction over a range of temperatures which resulted in little to no evident expression.

This construct was not further pursued. Additional expression trials would be required to test for expression using alternate media, such as auto-expression, terrific broth and 2xYT media.

The identification of alternative start sites

Further inspection of *C. thermarum* TA2.A1 genome including the 300 bp region upstream of PhoH2 provided curious options. Initial searches investigated possible antitoxin binding partners similar to those which bind and stabilize PIN domain VapC toxins. Options presented here were discounted as suitable antitoxins due to:

- Presence of a stop codon in the antitoxin sequence
- Not a multiple of 3, therefore unable to translate into amino acids
- Very short, only 30 aa long

Further searches identified alternative starting sites for the PhoH2 gene which may have previously been overlooked as *C. thermarum* TA2. A1 has not been entirely annotated (the initial sequence for PhoH2 from *C. thermarum* was identified with a BLAST search using PhoH2 from *M. tuberculosis* as a reference sequence). A short 30aa sequence and 18aa sequence immediately upstream of PhoH2 surprisingly showed homology to sequences found in DEAD/H box helicase proteins. Due to this and fortunately as we have access to the TA2. A1 genomic

DNA, these 2 alternative start sites were investigated in the hope that the additional upstream sequences would stabilize and fold PhoH2 correctly, yielding soluble PhoH2 protein.

3.1.1.2 PhoH2 + 30 amino acids

This new construct consists of the original *C. thermarum* PhoH2 sequence with an additional 30 amino acids upstream which are natively found in the *C. thermarum* TA2.A1 genome (Figure 3.13). These 30 amino acids begin with a Methionine residue, which is recognised by *E. coli* as a translation start site.

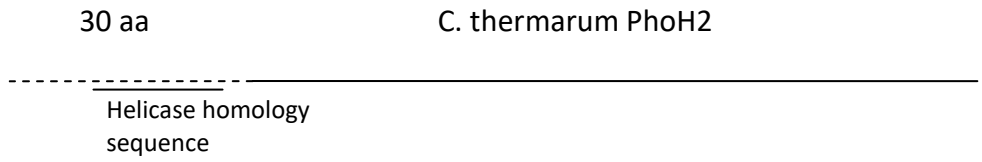


Figure 3.13: Schematic diagram of 30-PhoH2 construct.

3.1.1.3 PhoH2 + 18 amino acids

The second construct contains an additional 18 amino acid sequence upstream of PhoH2 (Figure 3.14). These 18 amino acids contain only the sequence identified in BLAST searches that has homology to DEAD/H box helicase proteins.

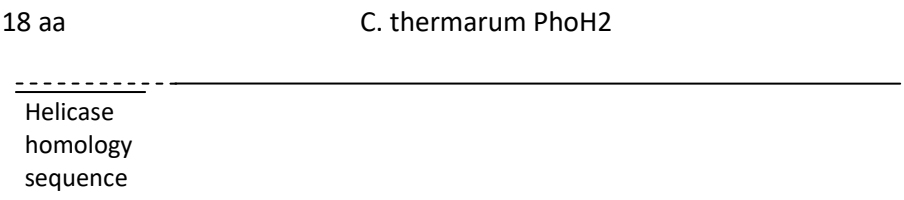


Figure 3.14: Schematic diagram of 18-PhoH2 construct.

New forward primers corresponding to the two alternative start sites were designed with an EcoRI restriction site and ordered from IDT technologies. EcoRI was used as it would be suitable for insertion into both pPROEX-Htb and pET28b-*Pst*I plasmids, therefore allowing studies with the His tag on either the N-terminus or C-terminus respectively. The reverse primers remained the same as previously used, for pPROEX-Htb this primer included the native stop codon and *Hind*III restriction site and for cloning into pET28b-PstI this excluded this stop codon but

included the *Hind*III restriction site. For the purpose of continuity, pPROEX-Htb plasmid was focused on initially while the pET28b-Htb-PstI plasmid was prepared for future use.

Using HotFire Polymerase, an initial gradient annealing temperature PCR was performed with temperatures ranging from 50-60 °C. Both the 30 and 18 PhoH2 constructs had good amplification across all temperatures, however 58 °C showed the cleanest product (Figure 3.15). Therefore 58 °C was chosen as the annealing temperature for subsequent PCRs.

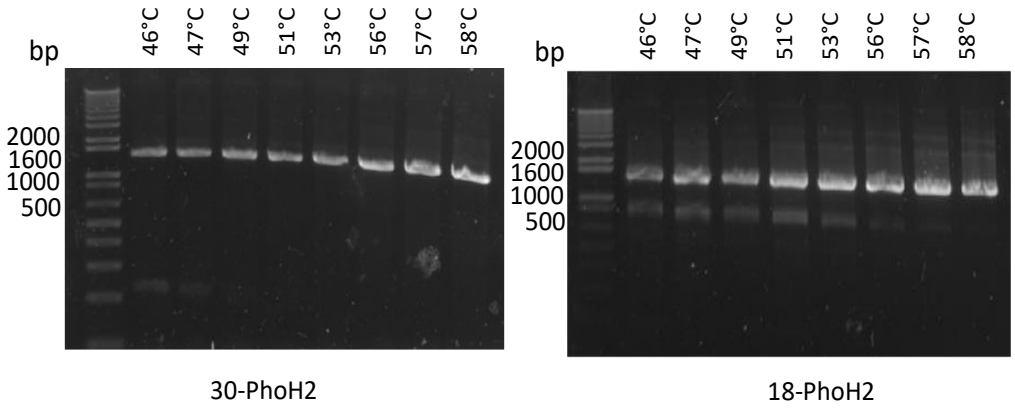


Figure 3.15: Gradient PCR from C. Thermarum gDNA using cloning primers on 30-PhoH2 and 18-PhoH2 constructs respectively.

Following the correct cloning of the 30-PhoH2 insert into pPROEX-Htb, confirmed by MSG sequencing, once again, the pPROEX-Htb 30- PhoH2 was expressed in small scale 100 mL LB media, and underwent small scale expression trials in order to try to gain soluble protein for further biochemical characterisations.

Temperatures following induction with IPTG were trialled at 28 °C, 30 °C and 37 °C. There appeared to be no observable difference in the solubility of the 30- PhoH2 protein across the temperatures (Figure 3.16). There was also no difference in expression, therefore subsequent small scale expression trials in various buffers were carried out using cells which were grown and induced at 37 °C.

Subsequent lysis buffer screens across pH (Figure 3.17) and NaCl and glycerol (Figure 3.18) addition also shows the desired 30-PhoH2 protein remains insoluble.

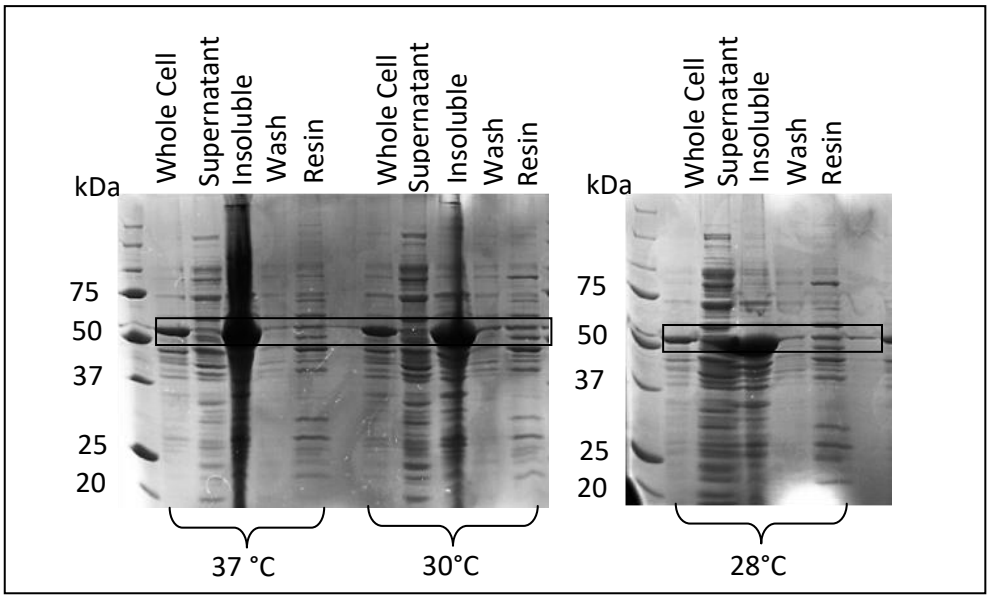


Figure 3.16: SDS-PAGE gels of small scale expression trials screening variable expression temperatures 37°C, 30°C and 28°C using lysis buffer 50 mM Tris pH 8.0, 200 mM NaCl, 20 mM Imidazole in 30-PhoH2.

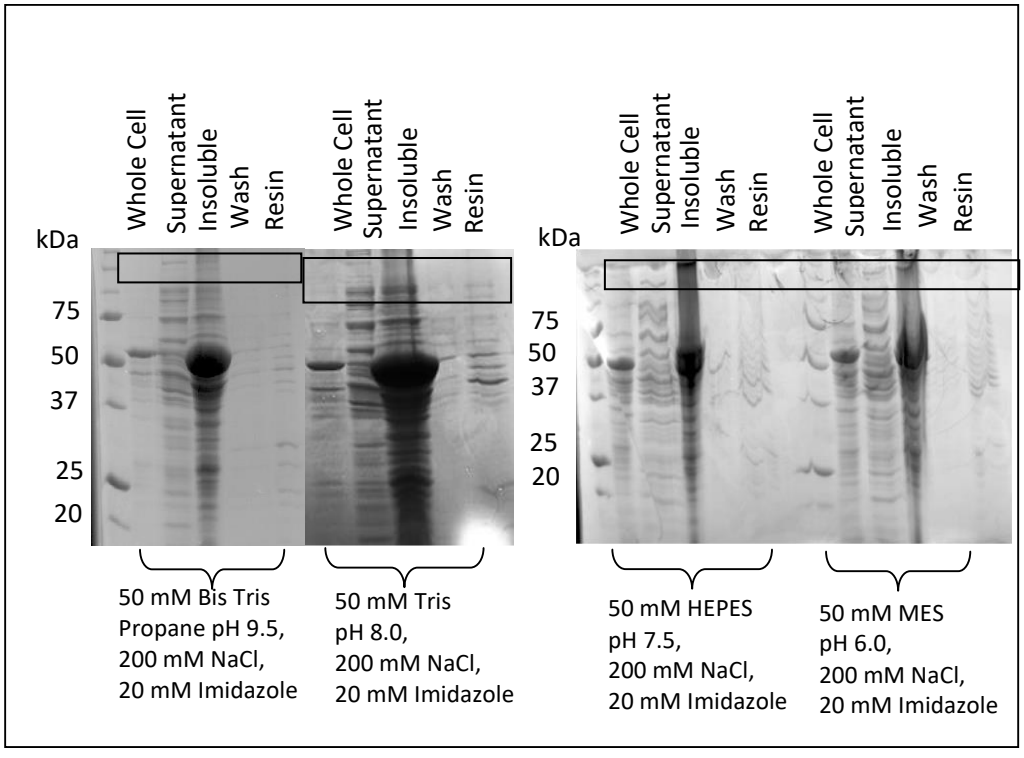


Figure 3.17: SDS-PAGE gels of small scale expression trials screening increased pH of lysis buffers in 30-PhoH2.

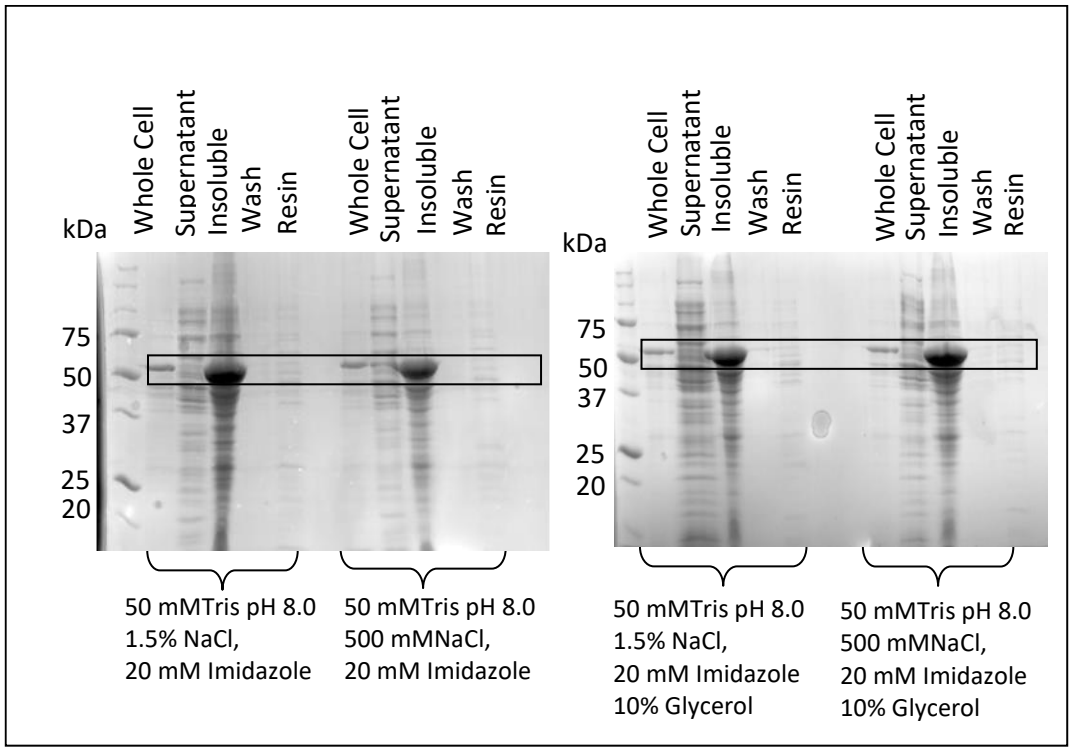


Figure 3.18: SDS-PAGE gels of small scale expression trials screening increased salt concentrations and the addition of 10 % glycerol to lysis buffers in 30-PhoH2.

The insert for 18-PhoH2 were also cloned by gradient PCR (Figure 3.15), and successfully ligated into pPROEX-Htb plasmid and transformed into DH5α *E. coli* cells. The above small scale expression trials were also conducted on this construct (Figure 3.19, Figure 3.20, Figure 3.21).

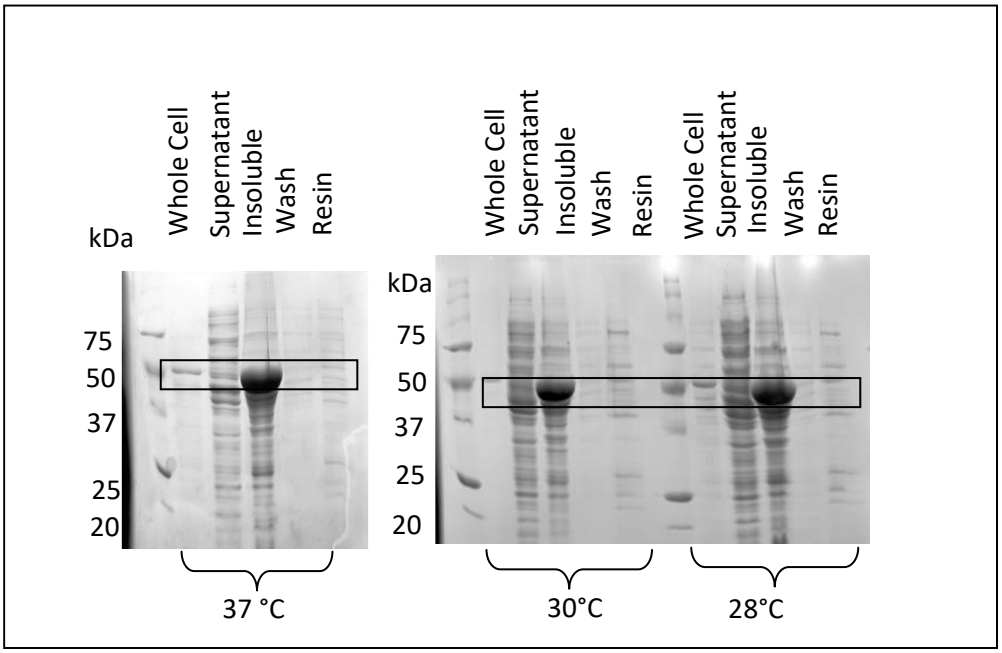


Figure 3.19: SDS-PAGE gels of small scale expression trials screening variable expression temperatures 37°C, 30°C and 28°C using lysis buffer 50 mM Tris pH 8.0, 200 mM NaCl, 20 mM Imidazole in 18-PhoH2.

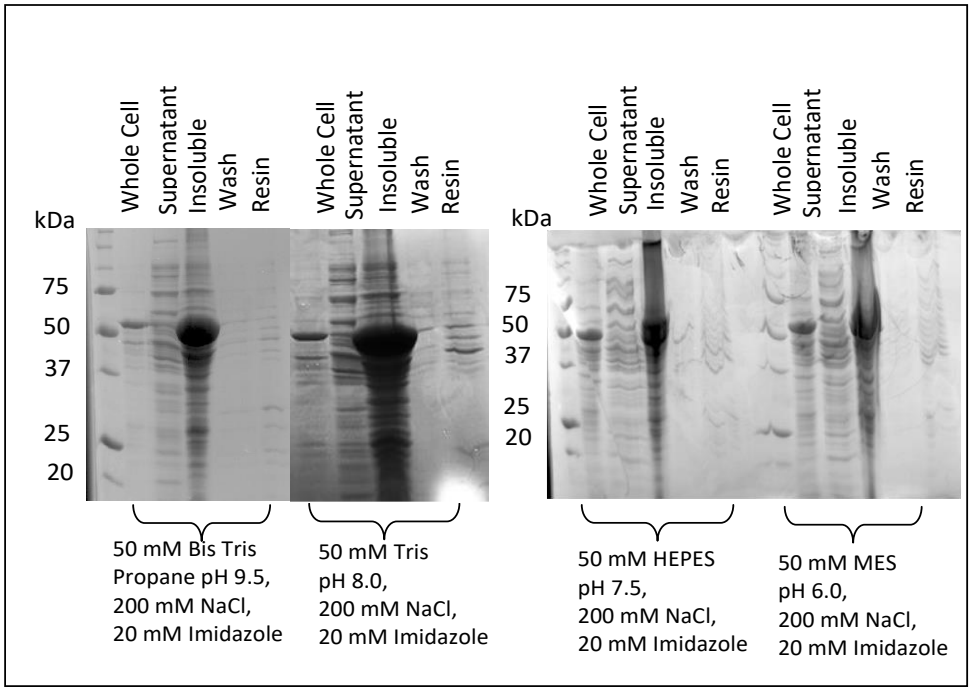


Figure 3.20: SDS-PAGE gels of small scale expression trials screening increased pH of lysis buffers in 18-PhoH2.

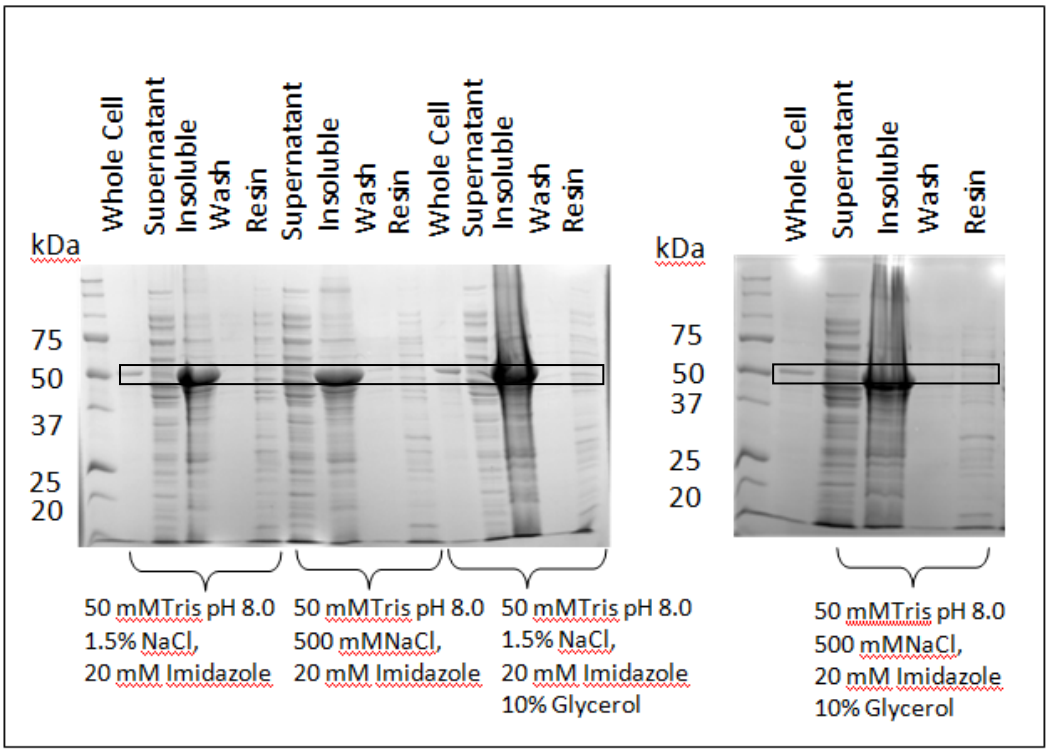


Figure 3.21: SDS-PAGE gels of small scale expression trials screening increased pH of lysis buffers in 18-PhoH2.

Additionally, work was carried out which aimed to clone the two alternative start sites into pET28b-*Pst*I plasmid. Both 30-PhoH2 and 18-PhoH2 inserts were

amplified from gDNA using HotFire Polymerase PCR successfully, the 30-PhoH2 insert was then ligated into pET28b-*Pst*I.

While for this construct cloning was achieved, the 18-PhoH2 insert was more difficult to ligate into the same digested pET28b-PstI plasmid.

18-PhoH2 amplification of the insert was repeated, and following the ligation and transformation into DH5α *E. coli* cells, there were minimal (5) colonies present. Colony PCR of these colonies using cloning primers showed a lack of 18-PhoH2 insert.

Discussion

Previous efforts into crystallising and structure solving of PhoH2 using X-ray crystallography have produced data to only modest resolution ~4 Å, but have been unsuccessful in determining the structure of PhoH2. Based on previous studies solving the structure of another large protein; ATP synthase, *C. thermarum* was predicted to be an advantageous candidate for steps towards solving the protein structure of PhoH2. The intended objectives for this study were to produce PhoH2 protein crystals for X-ray crystallography and biochemical characterisation using previously worked up assays, the structure of which would then be used to model subsequent PhoH2 structures from other organisms. In order to achieve this goal, the protein PhoH2 needs to be expressed and remain soluble in large quantities, and purified away from other cellular proteins through multiple steps to gain soluble pure PhoH2.

The above results describing the expression and purification trials undertaken of *C. thermarum* PhoH2 have not been able to produce conclusive conditions in which PhoH2 can be expressed and purified in a soluble form in preparation for downstream structural studies.

Across all three possible start site constructs cloned into pPROEX-HtB for expression in DH5α *E. coli*; PhoH2, 18-PhoH2 and 30-PhoH2, numerous buffer conditions in which the induced cells containing the over-expressed *C. thermarum* PhoH2 were lysed by sonication (100 mL). These buffer conditions were chosen to replicate the *in vivo* environmental ionic cellular conditions in order to maintain the stability of the PhoH2 protein of interest [47], while other buffers were chosen directly from conditions used by other research groups using proteins also cloned from *C. thermarum* [48, 50].

Lowering the temperature for protein expression following induction by the addition of IPTG failed to show any variation in the quantity of expressed protein, or any increase in solubility. The pH of the lysis buffer also did not improve the solubility. The addition of glycerol (5 - 10%) in the lysis buffer often increases the solubility of the protein, by decreasing the mobility of proteins and cellular debris

upon cellular lysis [51]. Addition of 10 % glycerol, in an attempt to stabilize the protein upon cell lysis was also unsuccessful, as was a range of salt concentrations in the lysis buffer.

Increasing the salt concentration of the buffer is thought to screen out the strength of ionic interactions [52]. The salt concentration of the expression medium (LB) influences the aggregation of proteins in solution. The likelihood of protein aggregation differs between proteins due to the amino acids side chains present and exposed in a natively folded protein. Therefore if the presence of acidic residues within the protein of interest is considered high, the increase of salt in the lysis buffer is likely to increase the stability of the protein, preventing the aggregation and precipitation of the protein of interest. In *C. thermarum* PhoH2, contains approximately 13 % acidic residues (ProtParam), therefore buffers trialled contained a range of increased salt, ranging from 200 mM NaCl to 500 mM NaCl, mimicking the halophilic environment *C. thermarum* tolerates. However, this did not appear to show any considerable improvement in the stability and thus the solubility of PhoH2.

Solubility additives not explored in this study are primarily the use of detergents in the lysis buffer. Traditionally, detergents are used where the protein of interest is an integral membrane protein [53]. The addition of detergents, such as Triton X-100 or deoxycholate up to 1%, are thought to increase the solubility of some proteins, by disrupting the cell membranes or releasing the proteins from aggregating into insoluble cellular debris. Given that the current studies into PhoH2 strongly suggest PhoH2 is not an integral membrane protein, but a cytoplasmic protein, the addition of a detergent to the lysis buffer is unlikely to result in significantly increased protein solubility.

Trialling various media in which the cells containing the recombinant PhoH2 protein is another option to improve the solubility upon lysis. Media conditions have the potential to induce cellular mechanisms such as an increased expression of chaperones which assist in the folding of the protein. Similarly, transforming the recombinant plasmid into Rosetta BL21 *E.coli* cells, or other strains which co-

expresses Gro ES/EL chaperones may increase the solubility of recombinant proteins.

Another factor in protein expression and purification is the presence of the linker region between the His-tag and the beginning of the PhoH2 protein when cloned into pPROEX. In pPROEX recombinant PhoH2, there are 47 amino acids between the N-terminal His-tag and the start codon of PhoH2. For the alternative start sites, 30-PhoH2 and 18-PhoH2 the length between the His-tag and the start of PhoH2 is reduced to 30 amino acids due to the position of EcoRI cutting sequence in pPROEX compared to the cutting sequence of XbaI, however no increase in expression or solubility of the desired PhoH2 protein was observed. Regardless of the length of linker region, it is plausible this region interferes with native folding of the protein, preventing optimum stability and thus affecting the solubility of the protein upon cell lysis. This would have substantial effect on the stability particularly on the oligomerisation of multi-domain proteins, as PhoH2 is reported to be a hexamer [34]. Secondly, lysing the cells at a high temperature may maintain the solubility of PhoH2, while denaturing and precipitating the *E.coli* cellular proteins.

Future studies following the above results may explore multiple options. Persistence with the current constructs using various expression media, cell lines and detergent containing media may be worthwhile. However, confirmation of cloning of PhoH2 into pET28b-*Pst*I plasmid for expression with a C-terminal His-tag removes the N-terminal linker between the His-tag and the beginning of the PhoH2 protein, may be a more successful option, given the previous studies by Andrews and Arcus (2015) showing soluble expression of PhoH2 with a C terminal His-tag with shorter linker region and adopted a hexamer conformation [34]. In addition to potentially gaining some degree of solubility, the lack of additional linker region would allow PhoH2 to be readily compared to PhoH2 from other organisms that have already been characterised, as was the original intention with the objectives of this study. Following the complete investigation into the pET28b-*Pst*I PhoH2 constructs if no improvement in the solubility of the protein is observed, investigations into studying a PhoH2 fusion protein from another viable thermophilic organism may be worthwhile.

Secondly, cloning, expressing and purifying the two domains (PIN and PhoH) individually in order to solve the separate structures and carry out basic biochemical characterisations could be a step towards solving the fusion PhoH2 protein. This has partially been done where the PhoH domain taken from the PhoH2 fusion protein has been structurally solved and is currently the only published structure available for PhoH on the Protein data bank [34]. The corresponding PIN domain from the same organism does not have a structure available. Individually, these domains may be structurally solved, leading to information which may be applicable to the fusion PhoH2 protein. Considering this PhoH domain from *C. glutamicum* has already been characterised, it may be worthwhile perusing this organism for an initial PhoH2 structure.

References

1. Yamaguchi, Y., J.H. Park, and M. Inouye, *Toxin-antitoxin systems in bacteria and archaea.* Annu Rev Genet, 2011. **45**: p. 61-79.
2. WHO, *Global tuberculosis report 2015*. 2015.
3. Wallis, R.S., *et al*., *Tuberculosis—advances in development of new drugs, treatment regimens, host-directed therapies, and biomarkers.* The Lancet Infectious Diseases, 2016. **16**(4): p. e34-e46.
4. Ribeiro, F.M. and T. Goldenberg, *Mycobacteria and autoimmunity.* Lupus, 2015. **24**: p. 374-381.
5. Pearce-Duvet, J.M., *The origin of human pathogens: evaluating the role of agriculture and domestic animals in the evolution of human disease.* Biol Rev Camb Philos Soc, 2006. **81**(3): p. 369-82.
6. Russel, D.G., *Mycobacterium Tuberculosis: Here today, and here tomorrow.* Molecular Cell Biology, 2001. **2**.
7. Gengenbacher, M. and S.H. Kaufmann, *Mycobacterium tuberculosis: success through dormancy.* FEMS Microbiol Rev, 2012. **36**(3): p. 514-32.
8. Keren, I., *et al*., *Characterization and Transcriptome Analysis of Mycobacterium tuberculosis Persisters.* mBio, 2011.
9. Vissa, V.D., *et al*., *Defining mycobacteria: Shared and specific genome features for different lifestyles.* Indian J Microbiol, 2009. **49**(1): p. 11-47.
10. organization, W.h., *National action for global change of antimicrobial resistance*. 2016: Geneva, World Health Organization.
11. Health, M.o., *Guidelines for the control of multi-drug resistant organisms in New Zealand*. 2007: Wellington: Ministry of Health.
12. Cox, R.A. and G.M. Cook, *Growth Regulation in the Mycobacterial Cell.* Current Molecular Medicine, 2007. **7**: p. 231-245.
13. Parrish, N.M., J.D. Dick, and W.R. Bishai, *Mechanisms of latency in Mycobacterium tuberculosis.* Trends in microbiology, 1998. **6**(3): p. 107-112.
14. Daffe, M. and G. Etienne, *The capsule of Mycobacterium tuberculosis and its implications for pathogenicity.* Tubercule and Lung Disease, 1999. **78**(3): p. 153-169.
15. Hatzios, S.K. and C.R. Bertozzi, *The regualtion of sulfur metablism in Mycobacterium tuberculosis.* PLoS Pathogens, 2011. **7**(7).
16. Liu, M., *et al*., *Mycobacterium tuberculosis effectors interfering host apoptosis signaling.* Apoptosis, 2015. **20**(7): p. 883-91.
17. Bartralot, R., *et al*., *Cutaneous infections due to nontuberculous mycobacteria: histopathological review of 28 cases. Comparative study between lesions observed in immunosuppressed patients and normal hosts.* Journal of cutaneous pathology, 2000. **27**(3): p. 124-129.
18. Sharp, J.D., *et al*., *Growth and translation inhibition through sequence-specific RNA binding by Mycobacterium tuberculosis VapC toxin.* J Biol Chem, 2012. **287**(16): p. 12835-47.
19. Sala, A., P. Bordes, and P. Genevaux, *Multiple toxin-antitoxin systems in Mycobacterium tuberculosis.* Toxins (Basel), 2014. **6**(3): p. 1002-20.

20. Van Melderen, L. and M. Saavedra De Bast, *Bacterial Toxin-Antitoxin Systems: More than Selfish Entities?* PLOS Genetics, 2009. **5**(3).
21. Jurenaite, M., A. Markuckas, and E. Suziedeliene, *Identification and characterization of Type II toxin-antitoxin systems in the opportunistic pathogen Acinetbacter baumannii.* Journal of Bacteriology, 2013. **195**(14): p. 3165-3172.
22. Wen, Y., E. Behiels, and B. Devreese, *Toxin-Antitoxin systems: their role in persistence, biofilm formation, and pathogenicity.* Pathog Dis, 2014. **70**(3): p. 240-9.
23. Schuster, C.F. and R. Bertram, *Toxin-antitoxin systems are ubiquitous and versatile modulators of prokaryotic cell fate.* FEMS Microbiol Lett, 2013. **340**(2): p. 73-85.
24. Ramage, H.R., L.E. Connolly, and J.S. Cox, *Comprehensive functional analysis of Mycobacterium tuberculosis toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution.* PLoS Genet, 2009. **5**(12): p. e1000767.
25. Ghafourian, S., *et al*., *Toxin-antitoxin Systems: Classification, Biological Function and Application in Biotechnology.* Curr. Issues. Mol. Biol., 2013. **16**: p. 9-14.
26. Brantl, S., *Bacterial type I toxin-antitoxin systems.* RNA Biol, 2012. **9**(12): p. 1488-90.
27. Leplae, R., *et al*., *Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families.* Nucleic Acids Res, 2011. **39**(13): p. 5513-25.
28. Sala, A., *et al*., *TAC from Mycobacterium tuberculosis: a paradigm for stress-responsive toxin-antitoxin systems controlled by SecB-like chaperones.* Cell Stress Chaperones, 2013. **18**(2): p. 129-35.
29. McKenzie, J.L., *et al*., *A VapBC toxin-antitoxin module is a posttranscriptional regulator of metabolic flux in mycobacteria.* J Bacteriol, 2012. **194**(9): p. 2189-204.
30. Cook, G.M., *et al*., *Ribonucleases in bacterial toxin-antitoxin systems.* Biochim Biophys Acta, 2013. **1829**(6-7): p. 523-31.
31. Arcus, V.L., P.B. Rainey, and S.J. Turner, *The PIN-domain toxin-antitoxin array in mycobacteria.* Trends Microbiol, 2005. **13**(8): p. 360-5.
32. Arcus, V.L., *et al*., *The PIN-domain ribonucleases and the prokaryotic VapBC toxin-antitoxin array.* Protein Eng Des Sel, 2011. **24**(1-2): p. 33-40.
33. Clissold, P.M. and C.P. Ponting, *PIN domains in nonsense-mediated mRNA decay and RNAi.* Current Biology, 2000. **10**(24): p. 888-890.
34. Andrews, E.S. and V.L. Arcus, *The mycobacterial PhoH2 proteins are type II toxin antitoxins coupled to RNA helicase domains.* Tuberculosis (Edinb), 2015. **95**(4): p. 385-94.
35. Finn, R.D., *et al*., *The Pfam protein families database: towards a more sustainable future.* Nucleic Acids Research, 2015. **44**(D1): p. D279-D285.
36. Min, A.B., *et al*., *The crystal structure of the Rv0301-Rv0300 VapBC-3 toxin-antitoxin complex from M. tuberculosis reveals a Mg(2)(+) ion in the active site and a putative RNA-binding site.* Protein Sci, 2012. **21**(11): p. 1754-67.

37. McKenzie, J.L., *et al*., *Determination of ribonuclease sequence-specificity using Pentaprobes and mass spectrometry.* RNA, 2012. **18**(6): p. 1267-78.
38. Winther, K.S., *et al*., *VapC20 of Mycobacterium tuberculosis cleaves the sarcin-ricin loop of 23S rRNA.* Nat Commun, 2013. **4**: p. 2796.
39. Koonin, E.V. and K.E. Rudd, *Two domains of superfamily I helicases may exist as separate proteins.* Protein Sci, 1996. **5**(1): p. 178-180.
40. Tanner, N.K. and P. Linder, *DExD/H Box RNA Helicases: Review From Generic Motors to Specific Dissociation Functions.* Molecular Cell, 2001. **8**: p. 251-262.
41. Andrews, E., *The biology and biochemistry of PhoH2 proteins*, in *Biological Sciences*. 2013, The University of Waikato. p. 195.
42. Kim SK, *et al*., *Molecular Analysis of the phoH Gene, Belonging to the Phosphate Regulon in Escherichia coli.* Journal of Bacteriology, 1993: p. 1316-1324.
43. Anderson, A., L. Ljungqvist, and M. Olsen, *Evidence that protein antigen b of Mycobacterium tuberculosis is involved in phosphate metabolism.* Journal of General Microbiology, 1990. **136**: p. 477-480.
44. Espitia, C., *et al*., *Phosphate starvation enhances expression of the immunodominant 38-kilodalton protein antigen of Mycobacterium tuberculosis: demonstration by immunogold electron microscopy.* Infect Immun, 1992. **60**(7): p. 2998-3001.
45. Kalamorz, F., *et al*., *Draft genome sequence of the thermoalkaliphilic Caldalkalibacillus thermarum strain TA2.A1.* J Bacteriol, 2011. **193**(16): p. 4290-1.
46. Peddie, C.J., G.M. Cook, and H.W. Morgan, *Sodium-dependent glutamate uptake by an Alkaliphilic thermophilic Bacillus strain, TA2.A1.* Journal of Bacteriology, 1999. **181**(10): p. 3172-3177.
47. Olsson, K., *et al*., *Bioenergetic Properties of the Thermoalkaliphilic Bacillus sp. Strain TA2.A1.* Journal of Bacteriology, 2003. **185**(2): p. 461-465.
48. Stocker, A., *et al*., *Purification, crystallization, and properties of F1-ATPase complexes from the thermoalkaliphilic Bacillus sp. strain TA2.A1.* J Struct Biol, 2005. **152**(2): p. 140-5.
49. Stocker, A., *et al*., *The structural basis for unidirectional rotation of thermoalkaliphilic F1-ATPase.* Structure, 2007. **15**(8): p. 904-14.
50. Keis, S., *et al*., *Cloning and molecular characterization of the atp operon encoding for the F1F0–ATP synthase from a thermoalkaliphilic Bacillus sp. strain TA2.A1.* Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression, 2004. **1676**(1): p. 112-117.
51. Timasheff, S.N., *The control of protein stability and association by weak interactions with water: How do solvents affect these processes?* Annu. Rev. Biophys. Biomol. Struct., 1993. **22**: p. 67/97.
52. Pace, C.N., G.R. Grimsley, and J.M. Scholtz, *Protein ionizable groups: pK values and their contribution to protein stability and solubility.* J Biol Chem, 2009. **284**(20): p. 13285-9.
53. Schuck, S., *et al*., *Resistance of cell membranes to different detergents.* Proc Natl Acad Sci U S A, 2003. **100**(10): p. 5795-800.