The thermo-responsive regulation of *Yersinia pestis* immune protective protein (F1) by the Caf1R transcription factor

Abdulmajeed Dhafer Majeed Al-jawdah

Thesis submitted in partial fulfillment of the requirements of the regulations for the degree of Doctor of Philosophy

Newcastle University

Faculty of Medical Sciences

Institute for Cell and Molecular Biosciences

August 2019

Acknowledgements

First of all, I would like to express my very great gratitude to my lord, Allah who guided me and helped me to finish this study.

I would like to express my deep gratefulness to my best supervisor Professor Jeremy H. Lakey for his patient supervision, enthusiastic encouragement and valuable assessments at all stages of this research work.

I am more than grateful to Dr Daniel T. Peters for his smart advice and beneficial assistance at all stages of this study.

Many thanks to Dr Helen Waller for her continuous professional technical support and for Dr Jeremy Brown and Dr Kevin Waldron for their wise advice and worthy evaluations.

Many thanks for my sponsor, Higher Committee for Education Development in Iraq for their generous funding of my study.

I wish to express my warm thanks to my parents, my wife, my brothers, my sister and my friends for their continuous support and bright encouragement.

Finally, special thanks for Dr Azzeldin Madkour, Dr Nico Paracini, Dr Iglika G. Ivanova and Prof Neil D Perkins and other lab members for their appreciated assistance during this work.

Abstract

Pathogenic bacteria can sense the temperature of the host body to provoke the production of required virulence factors that allow them to defend themselves from the host immune system. The molecular mechanism of such temperature-sensitive regulation might be at the transcriptional or translational level involving DNA, RNA or protein. In one example, when *Yersinia pestis*, (the plague causative bacterium) infects the flea where the temperature ≈ 25^{o}C it does not produce the Capsular Antigen F1 (Caf1). However, the bacterium can sense the change in temperature when it invades the human body and produces the Caf1 polymer as a coat to avoid the phagocytosis. The gene transcription of the *caf1* operon is increased in response to body temperature (37^{o}C).

Although Caf1R is thought to be the responsible factor for such transcriptional upregulation, there are no clear results explaining the mode of action of Caf1R. In this study it was shown that *caf1R* gene deletion stops the secretion of Caf1 polymer. Caf1R is the transcription factor responsible for the immediate thermal upregulation of *caf1* operon genes. Using molecular biology techniques, it was found that the *caf1* operon genes *caf1M*, *caf1A* and *caf1* are transcribed as a single long mRNA. The deletion of one of three predicted promoter regions upstream of this region (between *caf1R* and *caf1M*) was seen to have a significant effect on the transcription of *caf1* operon genes and protein production.

Finally, Caf1R was exploited to generate a novel thermal responsive expression system, which can be induced by switching the temperature from 25^{o}C to 35^{o}C instead of using a chemical inducer like IPTG or L-arabinose. Although this expression system needs more development to be usable in biotechnology applications, *gfp* expression from this system was shown to be highly temperature dependent.

List of abbreviations

LPS	lipopolysaccharides
CT	cytotoxin
FGS	F1-G1 loop Short
FGL	F1-G1 loop Long
Caf1	Capsular antigen F1
Ig	Immunoglobulin
SDS	Sodium Dodecyl Sulfate
PAGE	Poly Acrylamide Gel Electrophoresis
IPTG	Isopropyl β-D-1-thiogalactopyranoside
CAPS	*N*-cyclohexyl-3-aminopropanesulfonic acid
OD_{600}	Optical Density at 600 nm
RT-PCR	Real-Time Polymerase Chain Reaction
TB medium	Terrific Broth medium
LB medium	Luria-Bertani medium
nt	Nucleotide
bp	Base pair
RBS	Ribosome Binding Site
RACE	Rapid Amplification of cDNA Ends
GSP	Gene Specific Primer
GSP-RT	Gene Specific Primer for Reverse Transcription
HTH	Helix-Turn-Helix
SHS2	Strand-helix-Strand-Strand
WT	Wild-Type

3D	Three Dimensional
Y. pestis	*Yersinia pestis*
E. coli	*Escherichia coli*
EDTA	Ethylenediamine tetraacetic acid
DTT	Dithiothreitol
PBS	Phosphate Buffered Saline
SOC medium	Super Optimal broth with Catabolite repression medium
PMF	Peptide Mass Fingerprinting
6His-tag	Six-Histidine-tag

List of contents

List of figures

Chapter One: Introduction

1. Chapter One: Introduction

1.1. The plague causing bacterium *Yersinia pestis*

1.1.1. *Y. pestis* biology

Y. pestis is a Gram-negative, facultative aerobe, rod-shaped and capsule forming bacterium (Madigan, 2015). It is a plague causing bacterium which infects rodents as disease reservoirs and fleas can transfer the infection from infected animals to humans by biting. Bubonic, pneumonic and septicaemic are the most common severe forms of human plague. Plague was spread to Europe from Asia in the thirteenth century causing what is known as the Black Death, approximately a quarter of the European population succumbed to this epidemic infection (Greenwood, 2012).

Y. pestis can grow as single cells, in pairs and in chains in liquid culture as well as, it appears, in different morphologies like pear-shaped or globular cells especially in old cultures. When it is stained with Giemsa stain or methylene blue the smears will show characteristic bipolar staining (safety pin appearance). The optimum temperature for the bacteria to grow is 27^{o}C. The colonies that develop on blood agar within 24 hours will be small, slightly sticky, translucent and non-hemolytic. Additionally, it is sensitive to drying but could remain viable for many months especially at low temperatures. Heating to 55^{o}C for 5 minutes or using 0.5% phenol for 15 minutes is enough to kill the bacteria. *Y. pestis* is sensitive to many types of antibiotic including aminoglycosides, fluoroquinolones, chloramphenicol, co-trimoxazole and tetracyclines (Greenwood, 2012).

1.1.2. *Y. pestis* in the flea body

Y. pestis, the plague causative bacterium undergoes two phases in their pathogenicity. In the first phase, the bacteria infect the flea as a vector before passing into the human body. Therefore, bacteria have two kinds of factors used in each phase, the transmission factors which are important in flea vector and the virulence factors are crucial in the mammalian host. The ability of *Y. pestis* to form a biofilm is considered as one of the important transmission factors (Vadyvaloo *et al.*, 2007). The midgut of infected fleas is the place where *Y. pestis* multiplies to create a thick biofilm between the spines of the proventricular valve (Darby *et al.*, 2002; Jarrett *et al.*, 2004). In normal situations, the proventriculus closes and opens in a synchronised way during feeding, in order to push the blood into the midgut and to prevent it from returning back (Vadyvaloo *et al.*, 2007).

Y. pestis biofilm in the infected flea prevents the blood from reaching the midgut and consequently the contaminated blood will be returned into a mammalian host causing an infection (Vadyvaloo *et al.*, 2007). The main essential operon involved in biofilm formation in *Y. pestis* is *hmsHFRS*, the strain containing this operon can bind to Congo red (CR) or hemin and produce biofilm *in vitro* at 26^{o}C but not at 37^{o}C (Perry *et al.*, 1990; Hinnebusch *et al.*, 1996; Darby *et al.*, 2002; Jarrett *et al.*, 2004). The temperature range of 21^{o}C to 28^{o}C is the preferred temperature for biofilm formation *in vitro* which matches the body temperature in the flea, while the biofilm production is prevented at human body temperature 37^{o}C (Perry *et al.*, 1990; Jones *et al.*, 1999). Therefore, the biofilm formation by *Y. pestis* is not significant in the development of bubonic plague in the mammalian body (Kutyrev *et al.*, 1992; Lillard *et al.*, 1999).

The other known factors that *Y. pestis* uses during flea infection are the hemin storage genes (*hms*), this locus is responsible for the foregut blockage of the flea and then it is crucial for infection transmission, so the *hms* locus is required for biofilm formation (Hinnebusch *et al.*, 1996). The phosphoheptoisomerase is another transmission factor which is involved in synthesis of LPS and then in the formation of *Y. pestis* biofilm (Darby *et al.*, 2005). In addition, the intracellular phospholipase (murine toxin), which is produced by *Y. pestis* to survive in the flea midgut, encoded by *ymt* gene. The mechanism of action of this toxin is not known but it is necessary for bacteria to initiate the infection (Hinnebusch *et al.*, 2002).

1.1.3. *Y. pestis* in the mammalian host

The wild rodent species are considered as the main reservoir of the *Y. pestis* epidemics throughout the world (Vadyvaloo *et al.*, 2007). *Y. pestis* as pathogenic bacteria have many virulence factors used by them during their pathogenicity in the mammalian host. For example, lipopolysaccharides are a component of the *Y. pestis* outer membrane like other Gram-negative bacteria which are considered as one of the important virulence factors released as endotoxins after bacterial death (Brooks *et al.*, 2013). The type III secretion system is also used by *Y. pestis* to inject their virulence proteins (YopB and YopD) into the cytoplasm of the host cells directly. These proteins have the ability to disrupt the actin cytoskeleton and subsequently will help the bacteria to kill phagocytic cells and avoid triggering an inflammatory response in a host body (Slonczewski, 2011).

LcrV is a translocator protein which assists in the delivery of the effectors (YopB and YopD) across the target cell membrane (Cornelis and Wolf-Watz,

1997; Marenne *et al.*, 2003; Broz *et al.*, 2007), since it is essential for the correct assembly of the translocation pore (Goure *et al.*, 2005). Using Type III secretion systems, *Yersinia* can inject six other effector proteins YopH, YopE, YopJ, YopM, YopO and YopP and all these proteins have a role in deactivating the antibacterial activities of the host cells including signalling cascades, proinflammatory cytokines and the innate immune system (Cornelis, 2002).

The type III secretion system and effectors are encoded by about 50 genes located in the pCD1 plasmid. *Y. pestis* can produce the Pla surface protease which is a multipurpose virulence factor encoded by the pPCP1 plasmid. Pla can activate the plasminogen (the distributing human zymogen) to protease plasmin and inactivate the plasmin inhibitor, α_2-antiplasmin. Plasmin is able to degrade laminin, the major component of the basement membranes in mammals. In addition, complement proteins are proteolyzed by the Pla protease and by such proteolysis processes Pla mediates the invasion of *Y. pestis* through the human endothelial cell lines. *In vivo* studies showed that Pla is needed to establish bubonic plague (Suomalainen *et al.*, 2007).

As this study is investigating the regulation of the F1 capsule biogenesis, which is an important extracellular structure for the plague causing bacterium *Y. pestis* to defend itself from phagocytosis, it is useful here to understand the structural differences between Gram-negative and Gram-positive bacteria and to give a brief introduction about the pathogenicity, biotechnological uses and the outer membrane structure of Gram-negative bacteria.

1.2. Gram-negative bacteria

1.2.1. Gram-negative vs. Gram-positive bacteria

Based on the Gram stain method, which was developed in 1884 by Christian Gram, most bacteria can be divided into two major groups, Gram-positive bacteria which look purple under microscope and Gram-negative bacteria which are stained pink to red. The development of transmission electron microscope clarified the differing in cell wall structure between these two groups (Beveridge, 2001). One of the main different features is the peptidoglycan layer (**Figure 1-1**), the cell wall of Gram-negative bacteria is composed of a thin peptidoglycan layer (2 to 7 nm) coated by the outer membrane. Whereas, the Gram-positive bacterial cell wall contains a thick peptidoglycan layer (20 to 80 nm) which is located outside of the plasma membrane.

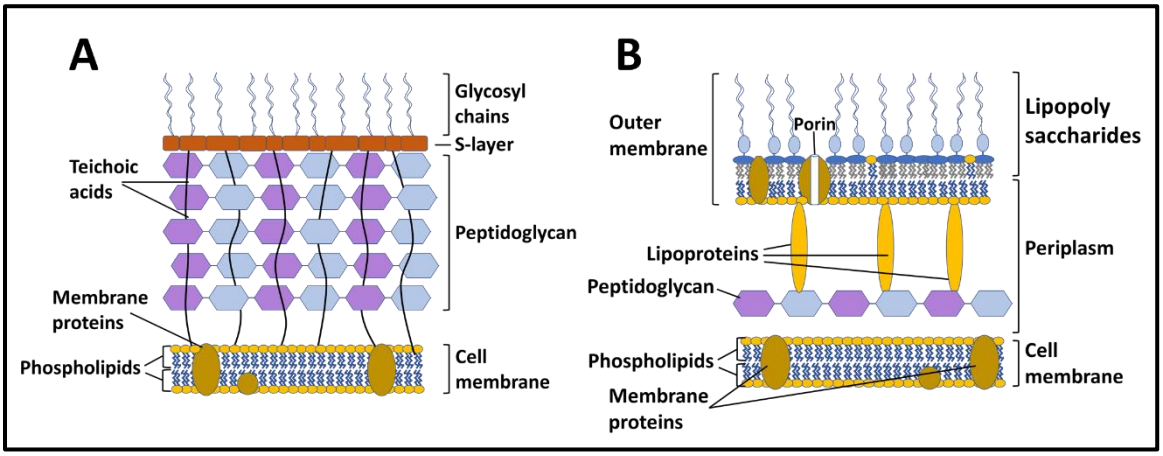

Figure 1-1. The structure of cell envelope in Gram-positive bacteria (A) and Gram-negative bacteria (B). Showing the multiple layer of peptidoglycan with teichoic acids threads in the thick cell wall of Gram-positive bacteria and thin cell wall with only one layer of peptidoglycan in Gram-negative bacteria. Adapted from (Slonczewski, 2011).

The greater thickness of peptidoglycan layer in Gram-positive bacteria as compared with Gram-negative bacteria make them more resistance to the osmotic pressure (Stock *et al.*, 1977; Willey *et al.*, 2008). Although the greater part of the cell wall in Gram-positive bacteria is peptidoglycan, it also contains a large quantity of teichoic acids which are attached covalently to either plasma membrane lipids or to the peptidoglycan surface (Heptinstall *et al.*, 1970). Therefore, the negative charge of the cell wall in Gram-positive bacteria comes from charged of teichoic acids. Although the presence of teichoic acids is crucial for conserving of cell wall structure, the molecular function of teichoic acids is unclear.

In terms of complexity, the Gram-negative cell wall structure is more complicated than the cell walls in Gram-positive bacteria. The periplasm of Gram-negative bacteria, which lies between the plasma membrane and the outer membrane (**Figure 1-1 A**) is unique to this group. However, the periplasm in *E. coli*, as an example for Gram-negative bacteria, contains only one or two peptidoglycan sheets of about 2 nm thick. The lipopolysaccharides are the most usual components of the Gram-negative bacterial outer membrane. They are large complicated molecules composed of three main parts: lipid A, the core oligosaccharides and the O-antigen. The lipid A is composed of two linked derivatives of the glucosamine sugar each one attached normally to three fatty acids and a phosphate or pyrophosphate group. The core polysaccharides are connected to the lipid A from one side and O side chain from other side which in turn extend to the outside (Willey *et al.*, 2008). The Gram-negative family of bacteria includes a number of bacteria species such as *Haemophilus*, *Salmonella*, *Shigella*,

Pseudomonas, *Proteus*, *Escherichia* and *Yersinia* (Willey *et al.*, 2008; Slonczewski, 2011).

1.2.2. Pathogenicity of Gram-negative

Pathogenicity is the ability of microorganism to create pathologic changes or disease in the host body and a pathogen is any disease-producing microorganism (Willey *et al.*, 2008; Foster, 2017). The magnitude of pathogenicity (virulence) is determined by the level of damage caused by a particular microorganism including invasiveness and infectivity. The pathogens, including Gram-negative bacterial pathogens, have often special structural or soluble products which increase bacterial pathogenicity and these bacterial tools are called virulence factors (Willey, 2014). Bacterial virulence factors include endotoxins, exotoxins, attachment proteins, capsules and other mechanisms used by bacteria to resist the host immune system (Willey *et al.*, 2008; Slonczewski, 2011).

Endotoxins are the structural components of the Gram-negative outer membrane otherwise known as lipopolysaccharides (LPS). As mentioned above LPS are composed of three structural components which are the membrane-distal O-specific polysaccharides, membrane-proximal core polysaccharides and lipid A. The toxicity of LPS originates from the lipid A portion, whereas, the water-soluble polysaccharide subunits are responsible for the activation of protective immunogenic reactions in the host body (Madigan, 2015). Lipid A causes endotoxic shock by binding to specific proteins called Toll like receptors or TLRs, on the surfaces of host cells promoting the production of cytokines (such as interleukins, tumor necrosis factor and platelet-activating factor) and nitric oxide leading to cell death

(Wilson and Salyers, 2011). A range of physiological effects can result from different types of Gram-negative bacteria due to endotoxin effects including fever, shock, blood coagulation, weakness, diarrhea, inflammation, fibrinolysis and death (Willey *et al.*, 2008; Foster, 2017).

Most Gram-negative pathogens also have the ability to produce specific types of proteins to kill the host cells such as cytolytic enzymes and receptor-binding proteins which are known as exotoxins (Murray, 2015). Many bacterial exotoxins, but not all, are composed of two subunits which are termed A and B. The A part is responsible for toxic properties, while B subunit binds to host cell receptors (Slonczewski, 2011). The exotoxins play an important role as a virulence factors in many Gram-negative bacteria. For example, the gastroenteritis-causing *E. coli* produces an enterotoxin called Shiga-like toxin which has an ability to inhibit protein synthesis in the host cells, *Salmonella* spp. produces cytotoxin (CT) which induces fluid loss from intestinal cells (Madigan, 2015). This is similar to heat labile enterotoxin of *E. coli* and cholera toxin from *Vibrio cholerae* (Spangler, 1992). In addition, *Yersinia pestis*, the plague-causing factor uses Yop proteins to disrupt the actin cytoskeleton of the host phagocytic cells to protect itself from phagocytosis (Slonczewski, 2011).

Other Gram-negative bacteria produce pore-forming toxins which permeabilise target cells such as aerolysin produced by *Aeromonas* sp. (Degiacomi *et al.*, 2013). Mimicking the host is another mechanism used by Gram-negative bacteria as a virulence factor, since such bacteria have the ability to produce capsules which resemble host polysaccharides. For example, *Neisseria meningitidis* has capsules consisting of sialic acid which is a common part of host cell glycoproteins and this type of capsule is not

immunogenic. The host immune cells do not produce antibodies to opsonize the surface of this capsule (Wilson and Salyers, 2011). Some other Gram-negative bacteria such as *Salmonella* can protect themselves from the hazardous effects of lysosomal enzymes by preventing lysosomal fusion with the phagosome (Slonczewski, 2011).

1.2.3. Uses of Gram-negative bacteria in biotechnology

Although Gram-positive bacteria are the most common bacteria used to produce numerous useful products for many different applications, Gram-negative bacteria can also be used for such applications. For example, organic acids and solvents, amino acids, exopolysaccharides, commercial and food processing enzymes, pharmaceuticals, biosurfactant, bioremediation and biofuels. The organic acids and solvents could be produced from genetically engineered *E. coli* in high levels such as succinic acid, polyhydroxyalkanoic acid (Rogers *et al.*, 2013), acetone, isopropanol (Chen *et al.*, 2013), glycerol and ethanol (Chen *et al.*, 2013). Furthermore, amino acid production was achieved using immobilized *E. coli* on special resin (for example, acrylamide gel) like L-aspartate, L-proline and L-threonine (Kumagai, 2013). Gram-negative bacteria (like *Klebsiella* spp. and *E. coli*) were also utilised in the production of commercial lipopolysaccharides with relatively simple structure (Sutherland, 2001).

Moreover, bacterial enzymes can be utilised for different purposes including commercial uses, detergents, food processing and textiles. For example, *E. coli* bacteria were used to manufacture bovine chymosin. Bovine chymosin is a protease used in the clotting initiation of the milk to make a cheese (Mohanty *et al.*, 1999). The production of cellulases has been performed

using different kinds of Gram-negative bacteria as expression hosts like *E. coli*, *Proteus vulgaris*, *Klebsiella pneumonia*, *Serratia liquefaciens*, *Pseudomonas fluorescens*. Briefly, the main applications of cellulases are in food processing, textiles manufacture and the production of cheap sugars for fermentation processes (Juturu and Wu, 2014). Additionally, numerous pharmaceutical products have been produced by recombinant Gram-negative bacteria including insulin, antibiotics, enzymes inhibitors, antitumor agents and immunomodulators (Lancini and Demain, 2013).

The microorganism can be employed in detoxification of different chemical compounds which contaminate waters and soils in process known as bioremediation. Both *Pseudomonas indigofera* and *E. coli* were used to detoxify the Cr^{6+} through reduction reactions which greatly decreases the toxicity this metal. As well, natural and artificial polymers can be enzymatically degraded using different types of microorganisms including *Pseudomonas syringae* (Gu and Mitchell, 2013). There are many other applications of Gram-negatives in biotechnology not mentioned above, one of them is the production of natural polymers for synthetic biology applications. For example, cellulose produced by Gram-negative bacteria was used in paper industry, biofuels, textiles and fibre industry (Kukoyi, 2016).

1.2.4. The outer membrane of Gram-negative bacteria

The outer membrane of Gram-negative bacteria gives protective abilities and toxigenic properties for many pathogens of this family such as *Salmonella* species and *E. coli*. The outer membrane of Gram-negative bacteria consists of bilayers with inward-facing and outward-facing leaflets.

The inward-facing sheet of the outer membrane contains phospholipids and lipoproteins that link the outer membrane to the cell wall peptidoglycan, while, the outward-facing sheet of the outer membrane is composed of a distinctive type of phospholipid called lipopolysaccharides (LPS) which also function as “endotoxin” that has additional medical significance, due to their ability to induce potentially lethal shock (Slonczewski, 2011).

Additionally, the outer membrane also includes other proteins not found in the inner membrane called porins (Willey *et al.*, 2008; Foster, 2017), which creates channels that allow nutrients such as sugars and peptides to enter. Depending on environmental conditions, cells produce different types of porins, for example in a poor environment cells express porins with a large pore size to increase the amount of nutrients uptake, whereas, in a fertile condition the cells downregulate the production of such porins to avoid the uptake of toxins (Willey *et al.*, 2008; Slonczewski, 2011; Foster, 2017). Most ions and small organic molecules are allowed to pass through the outer membrane selectively, since there is a cut off for soluble molecules of ≈ 600 Da. On the other hand, lipophilic toxins such as bile acids, proteins and other macromolecules are not permitted to enter (Nikaido and Vaara, 1985; Foster, 2017).

Furthermore, the outer membrane blocks proteins whose activities take place in the periplasm from spreading away from the cell (Madigan, 2015). The periplasm is about 15 nm wide (Leduc *et al.*, 1989), between the outer surface of the cytoplasmic membrane and the inner surface of the outer membrane. The periplasm could include several different sets of proteins according to the type of microorganism giving it a gel-like conformation. These involve hydrolytic enzymes, periplasmic binding proteins and

chemoreceptors (Nikaido and Vaara, 1985) and most of these proteins are transferred from the cytoplasm by a protein-exporting system, the general secretion system, which are two types in Gram-negative bacteria, type II and type V secretion pathways (Willey *et al.*, 2008; Foster, 2017).

1.3. Surface structures of Gram-negative bacteria

Depending on their changing environment bacteria sometimes need to attach to surfaces. At other times it is better to move toward or away from something to protect itself from hazards. Subsequently, many bacteria need to have specific structures that extend from the outer membrane to mediate the attachment of bacteria to the host cell (Willey, 2014) such as Type 1 pili in enterotoxigenic and uropathogenic *E. coli* (Lindberg *et al.*, 1987; Gong and Makowski, 1992; Jones *et al.*, 1995), since the attachment process is considered as the first step in the infectious disease (Willey *et al.*, 2008). While others need to have specific types of surface structures to use them in advantageous movements during their infectivity such as *Vibrio cholerae* flagella (Willey *et al.*, 2008). Other surface structures are considered as important virulence factors for Gram-negative pathogenic bacteria to resist the immune response in human body such as toxins, attachment proteins and capsules (Foster, 2017).

Pili or fimbriae are rod-shaped filamentous protein surface structures which bacteria use in adherence. Most Gram-negative bacteria have proteinaceous pili or fimbriae which vary in thickness and length. The most well-known pili in Gram-negative bacteria are Type 1 and P pili (**Figure 1-2**) in *E. coli*. The operons which express these pili are *fim* and *pap* respectively. The distinct feature of Type 1 pili is to induce hemagglutination. Type 2 and 3 fimbriae

are commonly found in *Enterobacteriaceae* including *E. coli*, *Klebsiella pneumoniae* and *Salmonella* species. Strikingly, type 1, 2 and 3 pili in Gram-negative bacteria are all assembled by the chaperone-usher pathway, while type 4 fimbriae are assembled by the type II secretion system (Wilson and Salyers, 2011).

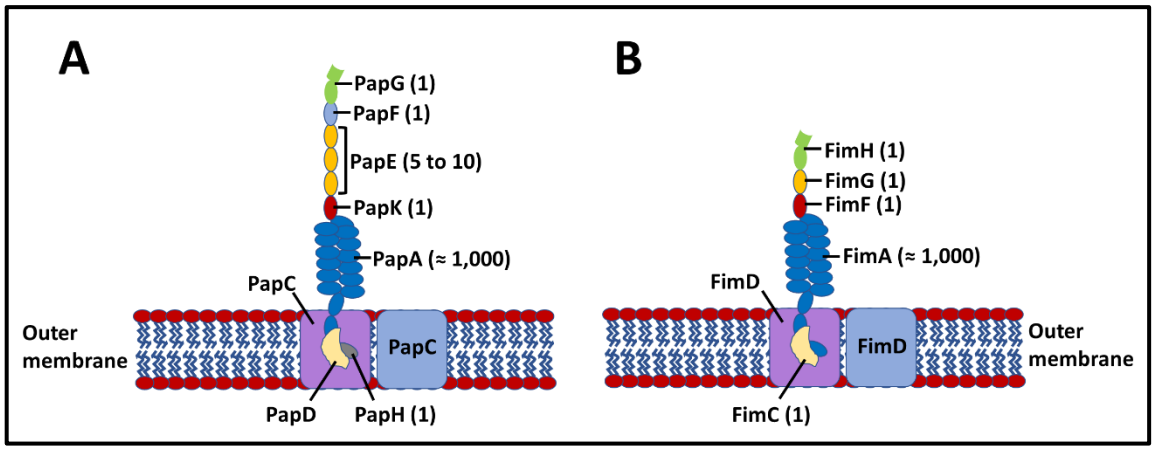

Figure 1-2. Schematic illustration of Pap and Fim systems as examples of P pilus and type I pilli in *E. coli*. For P pilli (A), PapD represents chaperone protein, PapC is usher outer membrane protein. Whereas, PapA, PapK, PapE, PapF and PapG are all different subunits (the building blocks) of the pili. For type I pilli (B), FimC represents chaperone protein, FimD is usher outer membrane protein. Whereas, FimA, FimF, FimG and FimH are all different subunits (the building blocks). Adapted from (Waksman and Hultgren, 2009).

As mentioned above some pathogenic Gram-negative bacteria have other types of structures which they use in swimming toward food and away from poisons, such structures are called flagella. Flagella are long and thin appendages attached to the cell surface at one end (Madigan, 2015) and they are composed of helically coiled subunits of a protein called flagellin (Murray, 2015). Flagella have the ability to rotate like a propeller on a boat generating rapid cell movement. For example, *E. coli* flagella can rotate at a rate of 270 revolutions per second (rps), the *Vibrio alginolyticus* motor can rotate at up to 1100 rps (Willey *et al.*, 2008). The movement of the cells

depends on the flagellar rotation. For example, when bacteria sense a chemical attractant, the flagellar motor will rotate counter-clockwise causing the bacteria to swim forward. If the bacteria need to move away from the attractant, the flagella will rotate clockwise against the twist of the helix causing the cell to turn (Foster, 2017).

Additionally, some Gram-negative bacteria can produce well organized layers of polysaccharides and protein called capsules. Even though capsules are not essential for cell growth and reproduction (Willey, 2014), they are considered as important virulence factors, for example, encapsulated *Enterobacteriaceae* have the ability to repel the hydrophobic surface of the phagocytic cells by the hydrophilic capsular antigens thus protecting itself from phagocytosis process (Foster, 2017). In one example, *Y. pestis* can produce the F1 capsule which has antiphagocytic activity (Du *et al.*, 2002). As the biogenesis of most pili, fimbriae and other non-plus structures are assembled by chaperone-usher secretion system, then, it is important to comprehend this secretion system.

1.4. Chaperone-usher secretion system

Many of the surface structures in different Gram-negative pathogens are made by chaperone-usher secretion system (Hung *et al.*, 1996). The chaperone-usher pathway consists of two main proteins, the chaperone protein which is responsible for transport of subunits through the periplasm to the outer membrane usher (Knight *et al.*, 2000), which in turn has a crucial role in assembling and exporting the growing polymers to the surface of the bacterium by forming a large pore in the outer membrane for both adhesive and non-adhesive surface organelles (Dodson *et al.*, 1993). The outer

membrane usher protein is thus the second main component of the chaperone-usher system, which binds to chaperone-subunit complexes encouraging both dissociation (Dodson *et al.*, 1993; Jones *et al.*, 1995) and polymerization of the structural subunits into a linear fibre. The growing polymer is then translocated extracellularly through the usher pore (Dubnovitsky *et al.*, 2010). The regulation of usher permeability is achieved by the middle plug domain, which has similar structure to fimbrial subunits (Yu *et al.*, 2009).

Pap pili in uropathogenic *E. coli* are assembled by the chaperone-usher pathway with the assistance of the PapD protein as a chaperone and PapC as a membrane usher protein. The assembly process of this pilus is regulated by 2 genes *papB* and *papE* in the *pap* operon which also encodes for PapD chaperone, PapC usher, PapA the major pilus subunit, PapF and PapG the initiator and terminator subunits and PapH the mannose-binding adhesin (Slonczewski, 2011). Furthermore, the atypical fimbriae Saf in *Salmonella typhi* the causative agent of enteric typhoid fever is another example of a chaperone-usher protein. The assembly of this pilus is mediated by the SafB chaperone and SafC usher which are encoded by the *saf* operon. This operon also encodes the subunit and the alternative subunit of the *Salmonella* atypical pili SafA and SafD respectively (Folkesson *et al.*, 1999). Recently, it was shown that *S. typhi* possesses 12 chaperone-usher gene clusters (Dufresne *et al.*, 2018).

Chaperone-usher pathways are divided into two subfamilies, FGS and FGL, according to the primary structure of the chaperone subunits (Hung *et al.*, 1996). FGS subfamily chaperones are responsible for the assembly of the rod-like adhesive organelles such as P and type-1 pili in *E. coli* (Knight *et al.*,

2000). These types of chaperones have short F1-G1 loops (**Figure 1-3**) which can reach to 10 - 20 residues in length and their name is derived from the loop length (F1-G1 Short). In the FGS subfamily, there are two invariant residues Tryptophan 36 and Asparagine 89 and the residue number 110 is always a positively charged amino acid (Hung *et al.*, 1996). In addition, the β-zipper motif in FGS chaperones is characterized by having three features: (i) a conserved configuration of alternating hydrophobic residues at positions 4, 6 and 8 from the carboxyl terminal; (ii) a penultimate tyrosine; (iii) a glycine at position 14 from carboxyl terminus (Kuehn *et al.*, 1993). The β-zipper is the extended β-sheet between PapD (chaperon) and the anchoring peptide of PapG (subunit) which is formed as a result of subsequent positioning of the PapG peptide along the G1 β-strand of the PapD (Kuehn *et al.*, 1993).

On the other hand, the FGL chaperone subfamily **(Figure 1-4)** is responsible for assembly of non-pilus, amorphous or capsule-like structures such as the F1 capsule in *Y. pestis* (Hung *et al.*, 1996; Zavialov *et al.*, 2003). As a result of having a long sequence between F1 and G1 strands which ranges from 21-29 residues, this chaperone subfamily is called FGL (F1-G1 long) (Zav'yalova *et al.*, 1995; Hung *et al.*, 1996). In the FGL subfamily chaperones, position 39 is occupied by either a polar or charged amino acid but not tryptophan as in FGS chaperones. Additionally, positions 89 and 110 are invariantly cysteine residues which are predicted to make a disulphide bond with each other (Hung *et al.*, 1996). The β-zipper motif in most FGL chaperones has a tyrosine 3 residues away from the carboxyl terminus and each FGL subunit has at least two alternating hydrophobic residues, most being at positions 6 and 8 residues from the carboxyl end (Hung *et al.*, 1996). Moreover, there are no

specific differences between the usher proteins involved in FGS chaperone-assembled adhesive pili and FGL chaperone-assembled fimbriae (Zavialov *et al.*, 2007).

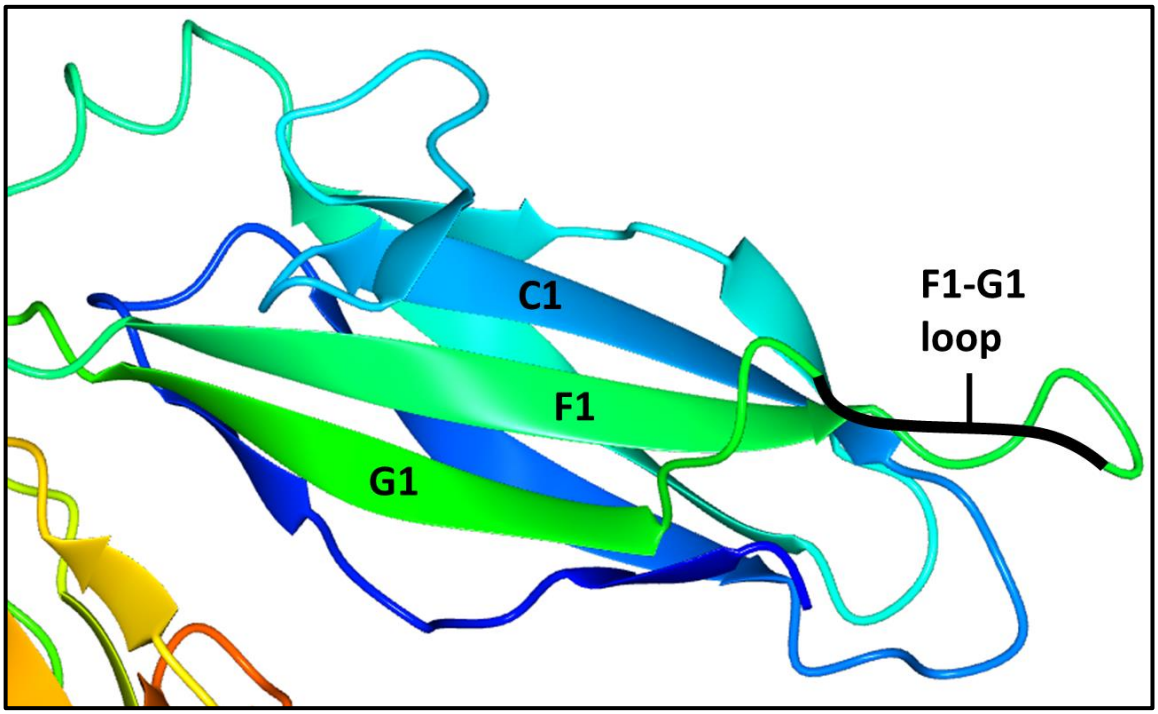

Figure 1-3. Crystal structure of PapD (chaperone) in complex with a pilus biogenesis inhibitor, pilicide 2c (PDB ID: 2J7L). Showing the F1 and G1 strands and the F1-G1 loop in green colour, which is short in this chaperon family (FGS) (Pinkner *et al.*, 2006). The structure was visualised using CCP4mg molecular-graphics software.

1.4.1. Caf1M (chaperone) from *Y. pestis* as an example of FGL subfamily

The chaperone protein, Caf1M, is one of the FGL-chaperone family which possess a long loop between F1 and G1 β-strands (**Figure 1-4**). It is encoded by the *caf1M* gene in the *caf1* operon with a molecular weight of 26.5 kDa (Chapman *et al.*, 1999). The main functions of the Caf1M protein are to protect Caf1 subunits from protease digestion, prevent subunit aggregation and transport Caf1 subunits to the usher protein where polymerization and secretion are achieved (Galyov *et al.*, 1991a; Zavyalov *et al.*, 1997; Chapman *et al.*, 1999). Caf1M chaperone like other FGL chaperones are responsible for

the assembly of simple surface structures which are formed from the polymerization of a single type of subunit (Zav'yalova *et al.*, 1995). Caf1M is composed of two Ig-like domains connected at approximately right angles to create a profound hydrophobic cleft between the domains (**Figure 1-4**) (Zavialov *et al.*, 2003).

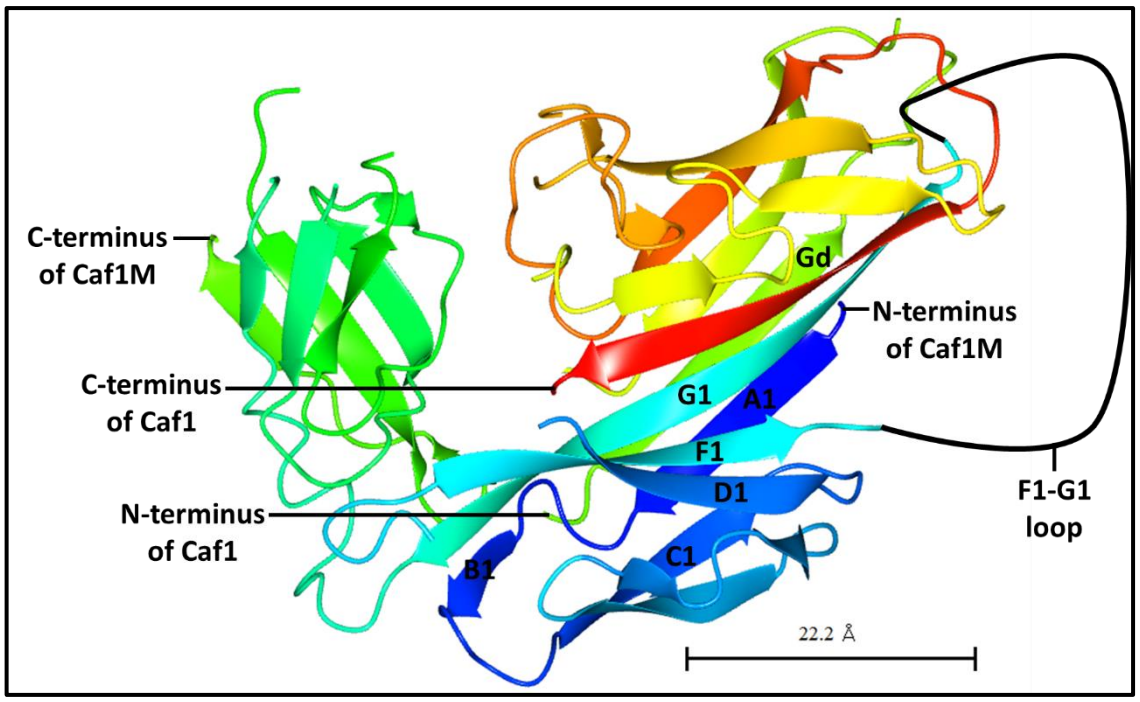

Figure 1-4. Crystal structure of Caf1M-Caf1 complex (PDB code: 4AZ8). Showing the two Ig-like domains connected at right angles with a large cleft between them (where the subunit C-terminus in red is anchored into) and G1 strand is inserted in the acceptor cleft of the Caf1. N and C-terminus of both Caf1M and Caf1, G1, F1, A1, C1, D1 β-strands, F1-G1 loop of Caf1M and Gd strand of Caf1 (Di Yu *et al*., 2012). The structure was visualised using CCP4mg molecular-graphics software.

The immunoglobulin-like domain (which is β-sandwich structure that occurs widely in biology and is named after the example found in immunoglobulins, the blood serum glycoproteins produced by plasma cells as antibodies in human immune system. It consists of two layers of 7-9 β-sheets with a Greek key topology and is found in many cell surface proteins) is constructed by linking two protein subunit chains. Although the seventh G strand of the Ig-

like bend is provided by the N-terminal (donor strand) of the anterior subunit in the fibre, all the first six strands originate from a single chain (Di Yu *et al.*, 2012). Since the donor strand of the anterior subunit takes the place of the missing Ig domain G strand to complete the Ig-like fold, the name of this interaction is donor strand complementation (Choudhury *et al.*, 1999; Sauer *et al.*, 1999). The Caf1M protein has two cysteine residues close to the subunit binding pocket and these cysteine residues can make disulphide bond. The disulphide bond is crucial for Caf1M-Caf1 complex formation. The Caf1M-Caf1 dissociation constant increases when the disulphide bond is either reduced or alkylated. The dissociation constant of the wild type complex was 4.77×10^{-9} M, whereas, the dissociation constant of the Caf1-Caf1M complex using modified Caf1M protein with either reduced or alkylated disulphide bond was 3.68×10^{-8} M (Zavyalov *et al.*, 1997). Though the disulphide bond is not important to maintain the Caf1M overall structure, it seems to affect the fine properties of the subunit-binding pocket structure. (Zav'yalov *et al.*, 1997). The G_1 β-strand of Caf1M is also significant for the subunit binding, as the hydrophobic side chains of the G_1 β-strand is inserted into the hydrophobic cleft of the subunits to be at the heart of the hydrophobic core by two positively charged amino acids (Zavialov *et al.*, 2003).

1.4.2. Caf1A outer membrane protein (usher) from *Y. pestis* as an example of FGL subfamily

Caf1A is the outer membrane protein (usher) which plays a crucial role in the polymerization of Caf1 subunits and secretion of the growing polymers from the bacterial cell surface. The usher monomer is composed of 811 residues

with a molecular weight of 90.5 kDa (Dubnovitsky *et al.*, 2010). The main structural components of Caf1A are: The N-terminal periplasmic domain, the central β-barrel domain which contains the transmembrane usher region, the carboxyl terminal periplasmic domain and the usher middle domain (**Figure 1-5**) (Remaut *et al.*, 2008; Yu *et al.*, 2009). The functions of each component have been clarified by several studies. The N-terminal domain has the ability to bind Caf1-loaded Caf1M during capsule assembly but not free Caf1M (Yu *et al.*, 2009). Even though the middle domain is not necessary for binding the Caf1M-subunit, it is required for the assembly process.

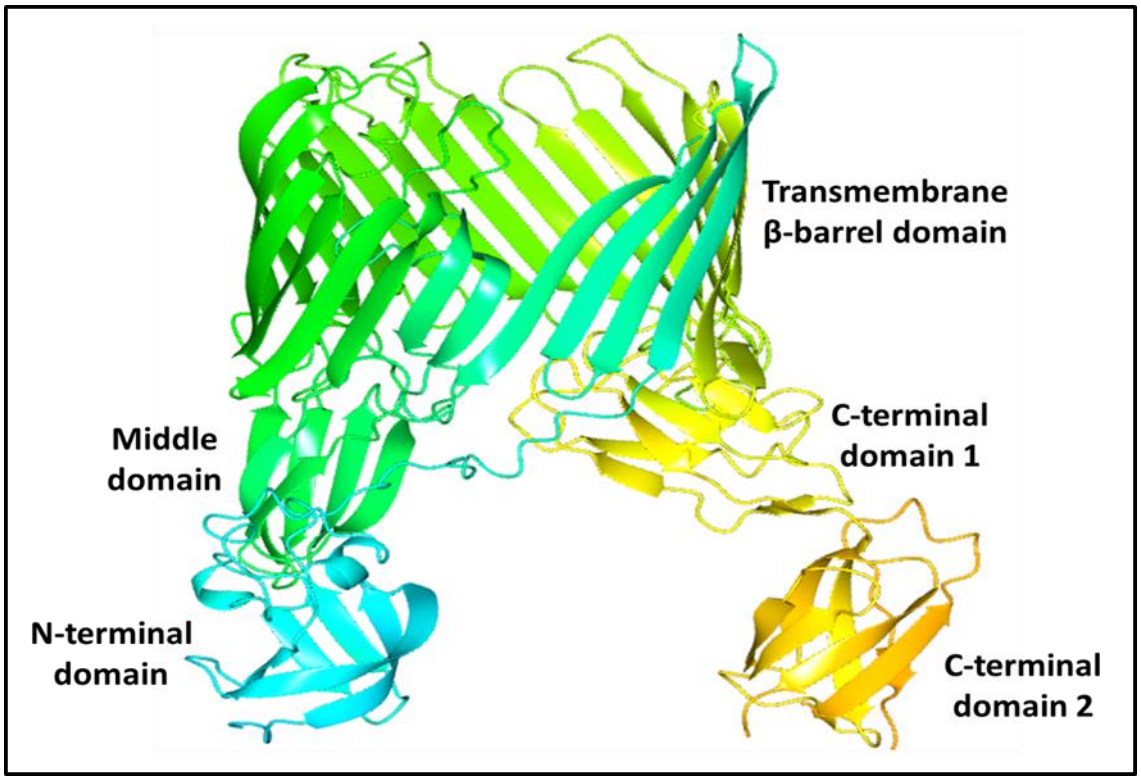

Figure 1-5. Structural modelling of the Caf1A usher outer membrane protein. Based on the crystal structure of FimD usher (PDB code: 3RFZ) (Phan *et al.*, 2011). Showing the structural components. The structure was visualised using CCP4mg molecular-graphics software.

Due to the similarity which the middle domain structure shows with the subunit to be exported, it seems that the usher middle domain is acting as a dummy subunit which plugs the secretion pore. While in the presence of Caf1M-subunit complex, the middle domain will be released from the pore to initiate the assembly and secretion processes (Yu *et al.*, 2009). The central β-barrel forms a suitable secretion channel to export the growing fibre to the cell surface (Remaut *et al.*, 2008; Yu *et al.*, 2009). The C-terminal periplasmic domain of Caf1A is composed of seven β-barrel strands. Although the C-terminal domain is not important for initial binding of Caf1M-subunit complex to the usher, the F1 capsule assembly process requires this domain to be achieved. The surface of the C-terminal domain contains clusters of hydrophobic residues which are responsible for its function (Dubnovitsky *et al.*, 2010).

1.5. Capsular Antigen Fraction 1 (Caf1) in *Y. pestis*

1.5.1. F1 capsule biogenesis

Y. pestis expresses large quantities of Caf1 subunits to form a thick gel-like capsule, which is deemed to be a distinctive feature of this bacteria (Du *et al.*, 2002). The F1 capsule as a non-pilus (atypical) adhesin is assembled by the FGL chaperone-usher system (Hung *et al.*, 1996) to form an amorphous structure composed of the Caf1 subunit protein as a building block (Galyov *et al.*, 1990; Vorontsov *et al.*, 1990). Capsule biogenesis is encoded by the *caf1* operon (**Figure 1-6**), which contains four structural genes (*caf1 R*, *caf1M*, *caf1A* and *caf1*) (Runco *et al.*, 2008). The F1 capsule is expressed efficiently at 35-37^{o}C to form a gelatinous surface covering the bacteria (Knight, 2007) through polymerization of Caf1 subunits on the cell surface to

make a mass of thin linear Caf1 fibres approximately 2 nm in diameter (Zavialov *et al.*, 2002; Zavialov *et al.*, 2003).

After the expression of *caf1* in the cytoplasm, the Caf1 subunits with their N-terminal signal sequence will be translocated across the plasma membrane to the periplasm by the general secretion pathway. The signal sequence is cleaved off and the subunits are then protected from aggregation and proteolytic action by the chaperone Caf1M during the transportation process across the periplasm (Knight, 2007). The Caf1 protein uses its C-terminal carboxyl group to bind in the subunit-binding cleft between two Caf1M domains specifically to Arg 20 and Lys 139 at the bottom of the binding cleft. These two residues are critical for the chaperone's function, as these two residues are highly conserved across all periplasmic chaperones. In Caf1M-Caf1 complexes, a super-barrel of β-strands in both subunit and chaperone are generated through the formation of hydrogen bonding interactions between the A and F edge strands of Caf1 subunit and the A1 and G1 strands of the chaperone (Zavialov *et al.*, 2003).

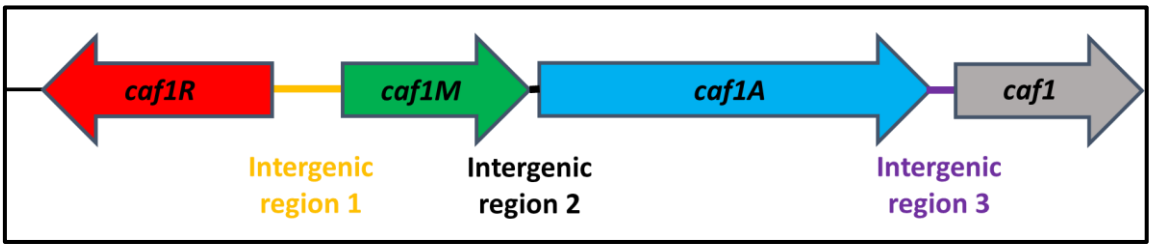

Figure 1-6. A diagram of *caf1* operon. It is present naturally in the pFra plasmid from *Y. pestis*. *caf1R* encodes the transcriptional regulator in the opposite direction, *caf1M* encodes the chaperone, *caf1A* encodes the usher outer membrane protein and *caf1* encodes the Caf1 subunit (the building block of the capsule). These genes are separated by three intergenic regions shown in different colours, yellow for intergenic region 1, black for intergenic region 2 and purple for intergenic region 3.

Consequently, the polymerization cleft will be coated to prevent both premature assembly and undefined subunit aggregation, as well as to protect the Caf1 subunit from proteolytic degradation. In the super-barrel, the five large residues of chaperone G1 donor strand are implanted between the two β-strands of the Caf1 subunit and become an integral portion of the hydrophobic core in the subunit (Knight, 2007). The folded Caf1 subunits will be then transported to the Caf1A usher in the outer membrane by Caf1M (Knight *et al.*, 2000; Sauer *et al.*, 2000a; Sauer *et al.*, 2000b) and chaperone-subunit interaction will then be replaced by subunit-subunit interaction. The exchange process is achieved by donor strand complementation (**Figure 1-7**), since the chaperone's A1 and G1 strands, which cover the subunit, need to be released, allowing the N-terminal sequence of the next subunit to be interleaved into the polymerization cleft of the growing Caf1 polymer on the periplasmic side of the usher (Choudhury *et al.*, 1999; Sauer *et al.*, 1999; Zavialov *et al.*, 2003). The subunit-free Caf1M predominantly occurs as a tetramer, each Caf1M molecule uses its subunit binding residues (the A1 and G1 strands) to form this tetramer (Zavialov and Knight, 2007). The tetramerization of Caf1M provides a protection from protease digestion (Zavialov and Knight, 2007).

Zavialov *et al.* (2003) proposed two models to clarify the process of donor strand exchange. In the first model, the N-terminal Gd donor strand of a chaperone-subunit complex is ordered and inserted into the unoccupied polymerization cleft after the liberation of chaperone-bound Caf1 subunit at the root of the growing fibre. The second model includes the tandem release of chaperone G1 donor strand and insertion of the subunit G1 donor strand in a zip-out-zip-in mechanism. The polymerization and secretion of Caf1

polymers to the surface is performed by Caf1A usher (Zavialov *et al.*, 2003; Remaut *et al.*, 2008). The growing fibres then extend through the secretion pore forming an amorphous structure on the cell surface (gel-like capsule) as in **Figure 1-8** (Vorontsov *et al.*, 1990; Galyov *et al.*, 1991b). The subunit-binding loop of the free Caf1M (chaperone) contains a pair of proline residues at its beginning called a proline lock (Di Yu *et al.*, 2012). This proline lock blocks the binding of chaperone to usher by occluding the usher-binding surface of the chaperone (Di Yu *et al.*, 2012). When the chaperone binds to the subunit, the proline lock will rotate away from the usher-binding surface to allow the subunit-chaperone complex to bind usher (Di Yu *et al.*, 2012). The protein material of the capsule is water soluble and has an ability to separate from the bacterium during *in vitro* cultivation and enter the culture medium (Du *et al.*, 2002).

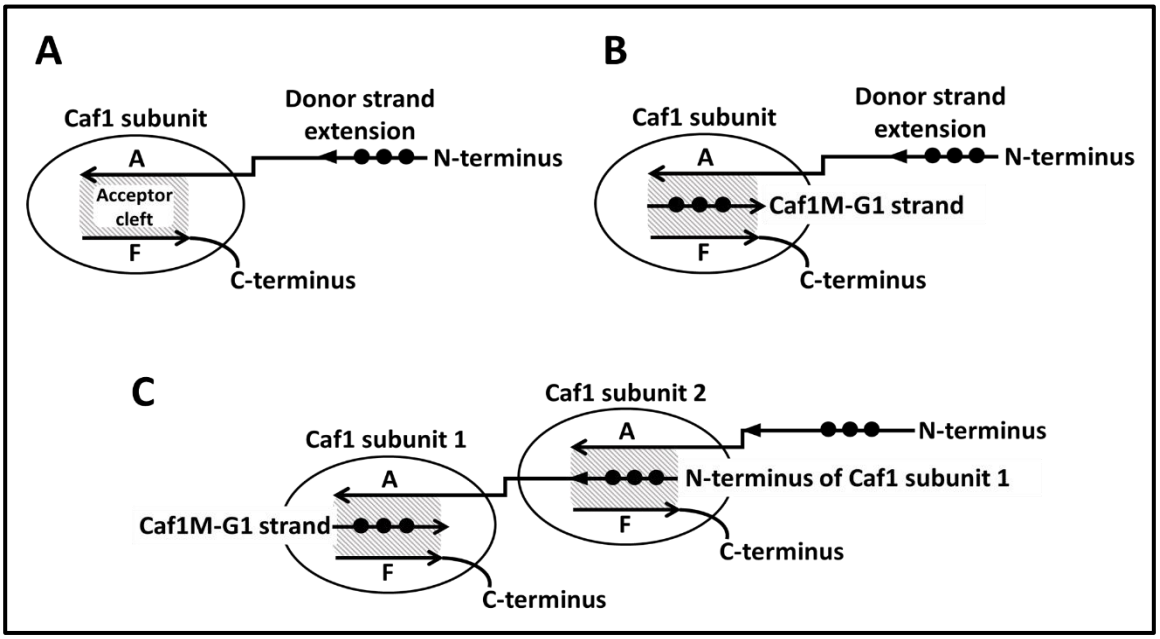

Figure 1-7. Schematic illustration of the donor strand complementation. (A) The Ig-like structure of subunit missing the seventh strand (G) to create the hydrophobic cleft between A and F strands. (B) The insertion of G1 β-strand from chaperone into the hydrophobic cleft during chaperone-subunit interaction. (C) The N-terminal donor strand of the subunit from the chaperone-subunit complex is inserted in the acceptor cleft of the second subunit. Adapted from (Zavialov *et al.*, 2002).

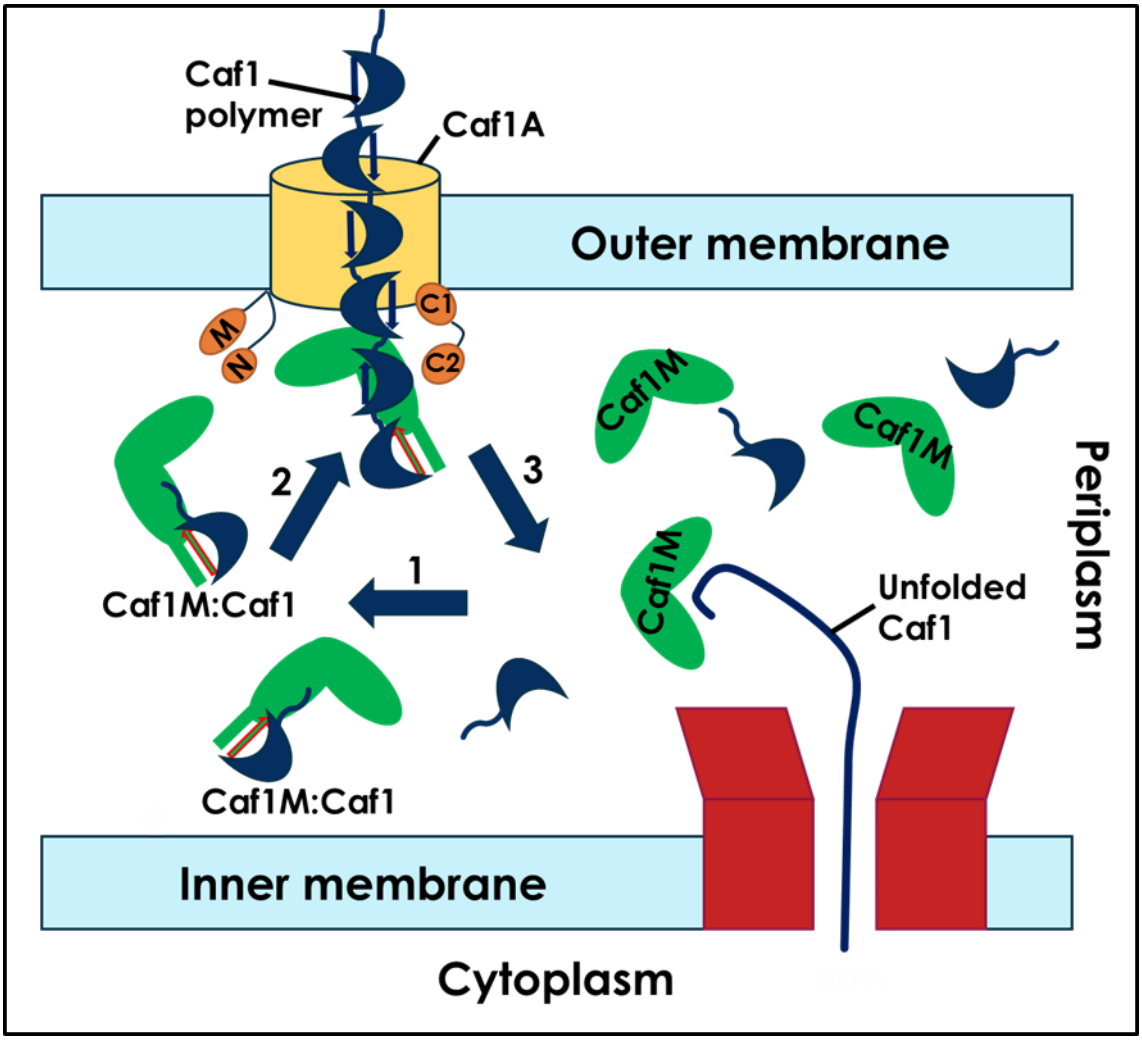

Figure 1-8. The schematic illustration of F1 capsule biogenesis. The main steps: (1) Caf1M binds to Caf1 subunits by donor strand complementation; (2) Caf1M:Caf1 complex bind to the activated Caf1A; (3) Caf1M release and the elongation of growing Caf1 polymer through the secretion channel of Caf1A. N, the N-terminal domain of the Caf1A; M, the middle domain of the Caf1A; C1 and C2, the C-terminal domains of Caf1A. Adapted from (Di Yu *et al.*, 2012).

1.5.2. The role of F1 capsule in virulence and immunity

The F1 capsule is often claimed to be essential for full bacterial virulence (Welkos *et al.*, 2004). Du *et al.* (2002) stated that the loss of F1 capsule caused a six-fold increase in bacterial uptake by macrophages and suggested that the F1 capsule has the ability to prevent this uptake by interfering with receptor interactions in the phagocytosis process. Friedlander *et al.* (1995)

assumed that the F1 capsule protein has a significant role in phagocytosis resistance. Furthermore, the expression of F1 capsule plays a part in the resistance of phagocytosis without any major effect on macrophage function, since the ability of macrophages (J774 cells) to digest yeast particles was not affected by the expression of F1 (Du *et al.*, 2002). This capsule also prevented the interaction between *Y. pestis* and macrophage-like cells (J774) (Du *et al.*, 2002).

Additionally, in a separate study, Cowan *et al.* (2000) observed that the invasion of *Y. pestis* into HeLa cells was restrained in a temperature-induced manner. Therefore, the F1 capsule might be responsible for this influence. Consequently, the F1 capsule is likely to have a general antiphagocytic impact not only against phagocytes. These observations could be evidence for a mechanism in which the F1 protein prevents the interaction between *Y. pestis* and phagocytes by covering the adhesion receptors and this could possibly lead to the reduced uptake of bacteria by phagocytes (Du *et al.*, 2002). Another study conducted by Liu *et al.* (2006) showed the ability of the F1 capsule to inhibit phagocytosis by three different types of human respiratory tract epithelial cells by inhibiting both adhesion and internalization of bacteria.

The significance of the F1 capsule in immunity was confirmed by several studies. The first isolation and characterization of F1 capsule as a protective immunogen was by Baker *et al.* (1952). Simpson *et al.* (1990) used F1 capsule expressed by recombinant *E. coli* to induce a protective immune response and showed that the high titres of F1 antibody enable immunized mice to resist infection by the plague. Therefore, it was assumed that recombinant F1 produced in *E. coli* is appropriate for vaccine development and

serodiagnostics of the plague (Simpson *et al.*, 1990). The protective immunogenicity of the F1 capsule was confirmed when mice were immunized against plague using both *Y. pestis* F1 antigen and recombinant F1 capsule from *E. coli* (Friedlander *et al.*, 1995). A next generation vaccine for plague has been developed by Williamson and Oyston (2013) by combination of both F1 capsular- and V-antigens. The circularly permuted monomeric vaccine was found to be less protective (Chalton *et al.*, 2006).

1.6. The Caf1R protein and its role in Caf1 expression

Karlyshev *et al.* (1992b) proposed that the Caf1R protein is a positive regulator of DNA transcription involved in F1 capsule biogenesis, due to the homology of Caf1R with the AraC proteins family, which are well-known DNA-binding proteins. The Caf1R protein is composed of 301 residues with a molecular weight 36 kDa. The calculated isoelectric point of Caf1R is 9.5, therefore, it is predicted to be highly basic protein, like other DNA-binding proteins (Karlyshev *et al.*, 1992b). The intergenic region between *caf1M* and *caf1A* genes could be the most likely regulatory region for Caf1R, since this sequence has conserved repeated elements ATGCTAA and TAGCTTT, which are repeated five times in this region and spread at equal distances from each other. The distances between elements are constant, 33 bp, between the 1-2, 2-3 and 3-4 repeats and 36 bp between the 4-5 repeats and the orientation of these sequences is on the same DNA side (Karlyshev *et al.*, 1992b).

Although Caf1R protein was required for *caf1* operon expression, using only the 81 N-terminal amino acids of Caf1R (truncated Caf1R) led to an increase the F1 capsule production (Karlyshev *et al.*, 1992b). Furthermore, it has been

suggested that the positive regulatory function could be achieved by the Caf1R N-terminus (Karlyshev *et al.*, 1992b), whereas, the C-terminus of Caf1R is responsible for another unknown regulatory role. The gene expression of both *caf1M* and *caf1A* is induced by Caf1R in response to a signal transduction which could be temperature and / or other environmental factors like the concentrations of Mg^{2+} or Ca^{2+} ions (Karlyshev *et al.*, 1992b). Finally, the Caf1R N-terminus shows high homology with the non-canonical transcriptional regulators (Rob, MarA and SoxS) from the AraC family and the prediction of 3D structure of Caf1R by SWISS-MODEL Workspace based on MarA and Rob structures (PDB code: 1D5Y and PDB code: 1BL0, respectively) displayed two potential helix-turn-helix motifs which could be the DNA-binding domains of Caf1R. The crystal structures of both MarA Rob are shown in **Figure 1-9**. It should be noted, however, that the exact molecular details of the Caf1R protein's role in the control and facilitation of F1 capsule biogenesis remain poorly understood.

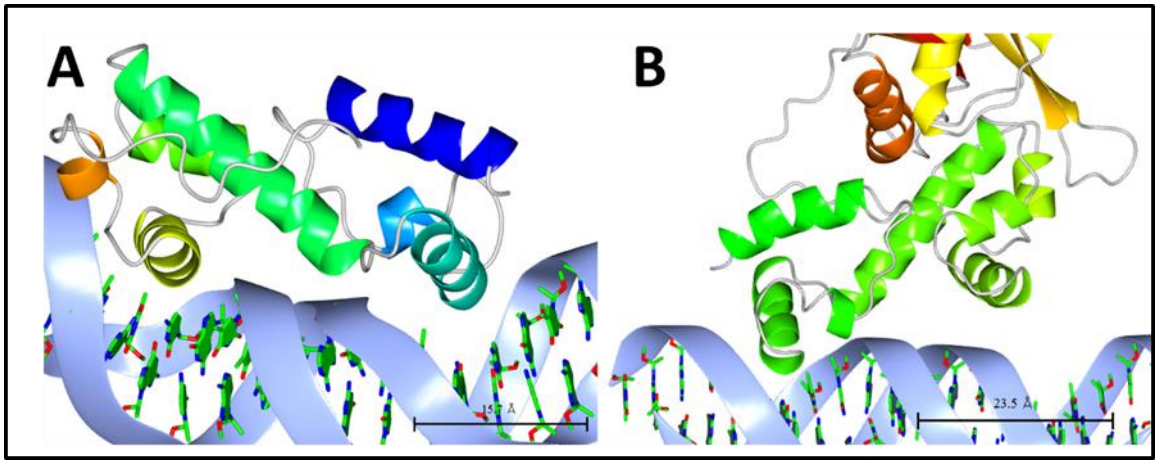

Figure 1-9. The crystal structures of Caf1R homologues. (A) The crystal structure of MarA (PDB code: 1BL0) (Rhee *et al.*, 1998). (B) The crystal structure of Rob (PDB code: 1D5Y) (Hyock Joo *et al.*, 2000). Showing the positioning of N-terminal helix-turn-helix domains (the DNA-binding domains) of both the MarA and Rob proteins close to the major grooves and backbones of DNA. The structure was visualised using CCP4mg molecular-graphics software.

1.7. AraC family transcriptional regulators

The AraC transcription factors is one of the most well studied family of the positive transcriptional regulators. The AraC family members show a significant homology especially in the N-terminal DNA-binding domain which can bind to specific promoters to trigger the transcription process of the target gene. The distinct feature of AraC family transcription factors is having the helix-turn-helix DNA-binding motifs. For example, MarA has two helix-turn-helix motifs which are involved in the regulation of many genes related to Multiple Antibiotic Resistance in *E. coli* including the activation of 25 genes (Martin and Rosner, 2002) and the repression of 3 promoters (Schneiders *et al.*, 2004; Schneiders and Levy, 2006). The position and orientation of the MarA-binding site (marbox) determines the kind of regulation achieved by MarA protein. When the marbox sequence is overlapped the promoter it acts as a repressor for the transcriptional activator protein, Rob (McMurry and Levy, 2010), on the other hand, the transcriptional activation (Dangi *et al.*, 2004) and repression (McMurry and Levy, 2010) by MarA protein can be also achieved by direct interaction with RNA polymerase.

The Rob protein homologue of MarA is another example of an AraC family transcriptional regulator in *E. coli* which is responsible for the activation of gene expression necessary for resistance to antibiotics, organic solvents and heavy metals (Hyock Joo *et al.*, 2000). In addition, the SoxS protein is also a MarA homologue which is involved in the transcriptional activation of the *soxRS* locus in response to the stress of nitric oxide or superoxide (Li and Demple, 1994). Another transcriptional activator from the AraC family is RhaS, which is responsible for the activation of L-rhamnose transport (Tate

et al., 1992) and catabolism as an alternative carbon source in the presence of L-rhamnose in *E. coli* (Power, 1967; Moralejo *et al.*, 1993) by the transcription activation of both *rhaBAD* (Egan and Schleif, 1993) and *rhaT* (Via *et al.*, 1996) operons. The *rhaBAD* involves three structural genes: *rhaA* coding for rhamnose isomerase, *rhaB* coding for rhamnulose kinase and *rhaD* coding for rhamnulose-1-phosphate aldolase (Power, 1967). Whereas, *rhaT* is coding for the transport system of the rhamnose (Garciamartin *et al.*, 1992; Via *et al.*, 1996)

The most common and well-known transcriptional regulator in the AraC family is the AraC protein, which gives its name to the family. AraC regulates the uptake of L-arabinose as a carbon source when it is available in the growth medium. The AraC protein is functional in a homodimer form, the monomer is composed of a dimerization domain which is also the L-arabinose-binding and DNA-binding domain (**Figure 1-10 A**). In the absence of L-arabinose, the DNA loop will be created to prevent RNA polymerase from binding to P_{BAD} regulatory site (**Figure 1-10 B**). The DNA loop is formed as a result of the orientation of DNA-binding domains of the AraC dimer to bind to the I_1 and O_2 half-sites (the regulatory sites of the araCBAD genes) in the absence of L-arabinose. Since, it is energetically favoured for the N-terminal arm and the linker of AraC protein to be in combination with each other, this in turn will hold both DNA-binding domains moderately rigidly to their dimerization domains (Schleif, 2000).

In the presence of L-arabinose and due to its binding to the dimerization domains, the N-terminal arms will bind to the dimerization domains. Consequently, the DNA binding domains will be released to be free for reorientation relatively. The overall energy state for AraC protein dimer in

the presence of L-arabinose is in favour with binding of the DNA-binding domains to the adjacent I_1 and I_2 half-sites (**Figure 1-10 C**) rather than looping the DNA (Schleif, 2000).

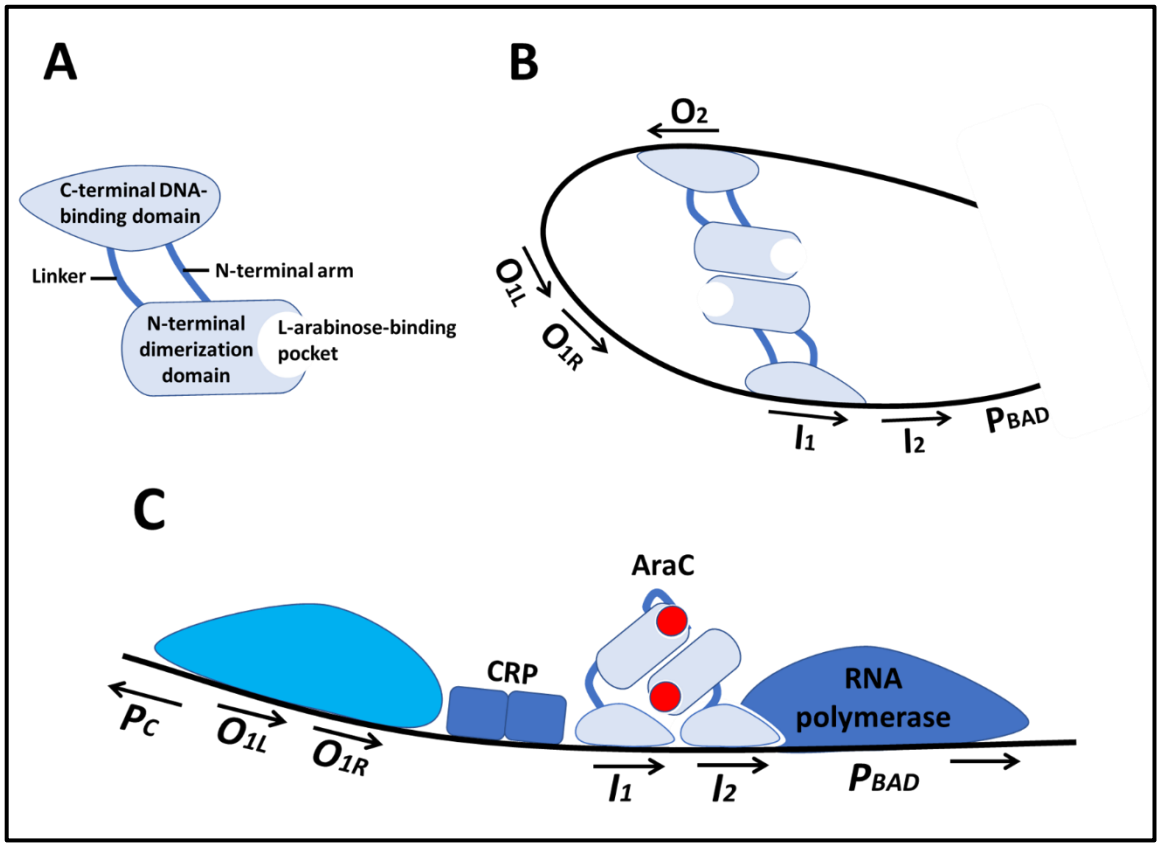

Figure 1-10. Schematic illustration of the regulation by AraC transcriptional regulator as an example for AraC family. (A) The composition of AraC protein as a monomer. (B) The repression effect of AraC on the *PBAD* promoter in the absence of L-arabinose. The overall energy state of the AraC dimer encourage both DNA-binding domains to contact *O2* and *I1* half-sites to form DNA loop and prevent RNA polymerase from binding. (C) The activation of transcription in the presence of L-arabinose. One DNA-binding domain L-arabinose-bound AraC protein release from *O2* regulatory site and bind to *I2* to recruit RNA polymerase and activate the transcription. *PC* is the promoter of cyclic AMP receptor protein. Adapted from (Schleif, 2010).

1.8. The F1 capsule (Caf1) protein and its special properties

Caf1 polymer is composed of 15.5 kDa subunits chain and can reach to 1.5 µm and up to 250 monomers in length (**Figure 1-11**) (Soliakov *et al.*, 2010) which are linked with each other through strong non-covalent interactions

(Zavialov *et al.*, 2005; Remaut and Waksman, 2006). Caf1 polymer is encoded by *caf1* operon which contains four structural genes encoding for four proteins. Caf1R the transcriptional regulator, Caf1M the chaperone, Caf1A the usher outer membrane protein and the Caf1 subunit (the building block of the capsule) (Runco *et al.*, 2008). Vitagliano *et al.* (2008) suggested that Caf1 can form a linear polymer of Ig-like structure with significant stability according to the structure models they obtained from MD simulations. The study of the conformational states of Caf1 monomer showed that the unbound forms of protein tend to be more compact than the folded subunits. Therefore, this structural feature could hinder the subunit-subunit association, preserving the high reactivity of the subunit which is important for firm association through the donor strand complementation process (Vitagliano *et al.*, 2008).

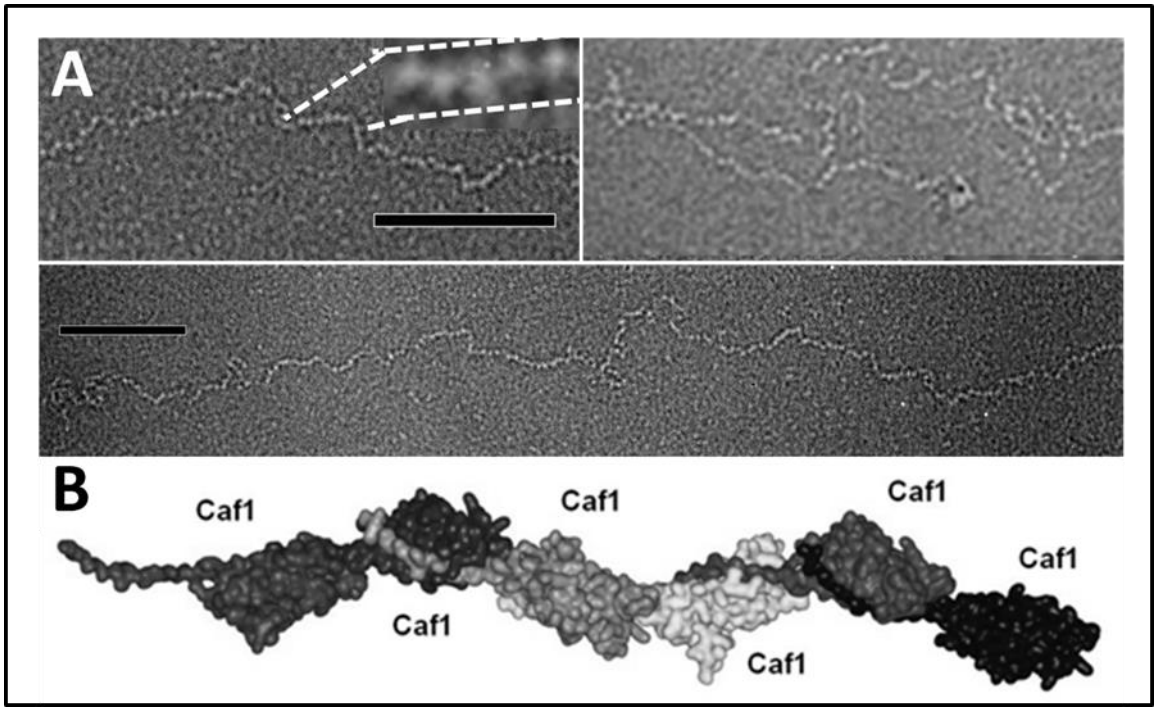

Figure 1-11. The structure of Caf1 polymer. (A) TEM of Caf1 polymer shows a long linear polymer. The black scale bar implies 100 nm in all cases. (B) The model of the linear Caf1 polymer based on (PDB code: 1Z9S). Taken from (Soliakov *et al.*, 2010).

Caf1 has several features which make it suitable to be utilized for biotechnological purposes. Caf1 has Ig-subunit-like structure which is similar to that of human fibronectin III and other cell surface proteins (Leahy *et al.*, 1996; Özkan *et al.*, 2013), which makes it useful in three-dimensional cell culture matrices. In addition, Caf1 can be produced (Miller *et al.*, 1998), engineered and modified proficiently in recombinant *E. coli* (Roque et al., 2014). Additionally, Caf1 subunits are very robust and non-adhesive for cells (Roque *et al.*, 2014). Caf1 polymer is highly thermostable even in extreme chemical conditions like different pH ranges and the presence of detergents and high salt concentrations (Ulusu *et al.*, 2017). Zavialov *et al.* (2005) also stated that Caf1 polymer is able to resist degradation by proteolysis in addition to its high thermostability. These properties suggest Caf1 as a promising biomaterial to produce a synthetic hydrogel for cell culture.

1.9. The aim of this study

The aim of this study is to illustrate the function of Caf1R in more detail, show how the *caf1* operon gene expression is activated in response to thermal change and the role of Caf1R in such regulation to determine the molecular mechanism of regulation by which Caf1R can induce the transcription of *caf1* operon at 35^{o}C not at 25^{o}C. Using molecular biology techniques such as RT-PCR and western blot, the number of mRNA transcript units produced from transcription of the *caf1* operon will be examined in order to estimate the number and positions of responsible promoters and/or Caf1R-binding sites for the transcription of *caf1* operon. The structural modelling of Caf1R will be achieved based on the crystal structures of Rob and MarA the Caf1R homologues to predict the structural details of Caf1R especially the helix-

turn-helix DNA-binding domains. The significance of Caf1R C-terminus in *caf1* expression will be investigated as well. The purification of Caf1R protein will be performed to characterise the pure Caf1R *in vitro* by biophysics and molecular biology techniques. Finally, the ability of Caf1R to activate the transcription in response to temperature change will be exploited to generate a novel thermal inducible expression system.

Chapter Two: Materials and Methods

2. Chapter Two: Materials and Methods

2.1. Introduction

The procedures used in experiments performed in chapters 3, 4 and 5 are described in this chapter. Particular features or procedures unique to certain experiments may be described in the relevant chapter.

2.2. Suppliers

All chemicals and reagents were supplied either from Sigma-Aldrich or Melford Laboratories, unless otherwise noted. Q5 Site-directed mutagenesis, Phusion DNA polymerase, terminal transferase, Monarch plasmid miniprep and Monarch DNA gel extraction kits were supplied from New England Biolabs. In-Fusion HD cloning kit was supplied from Clontech. QuantiTect Reverse Transcription Kit was supplied from Qiagen. EZ-10 Total RNA Mini-Preps Kit was supplied from Bio Basic Inc. GoTaq Flexi DNA Polymerase and PCR nucleotide mix were supplied from Promega. HisTrap HP His-tag purification column was supplied from GE Healthcare and Proteo SEC 3 - 70 kDa HR size exclusion column was supplied from Generon. Anti-Caf1 and anti-DnaK antibodies were supplied from Abcam.

2.3. Bacterial strains

Different *E. coli* strains were used in this study as listed in **Table 2-1**.

Table 2-1. The different *E. coli* strains used in this study

***E. coli* strain**	**Genotype**	**Supplier**	**Function**
BL21 (DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	New England Biolabs	Protein expression
Mach1™	F- *φ80(lacZ)ΔM15 ΔlacX74 hsdR (rK-mK+) ΔrecA1398 endA1 tonA*	Invitrogen	Cloning
Turbo	F´ *proA+B+ lacIq Δ lacZ M15/ fhuA2 Δ(lac-proAB) glnV gal R(zgb-210::Tn10)TetS endA1 thi-1 Δ(hsdS-mcrB)5*	New England Biolabs	Cloning
Tuner™ (DE3)	F- *ompT hsdSB (rB- mB-) gal dcm lacY1(DE3)*	Novagen	Protein expression
BL21(DE3) pLysE	F- *ompT hsdSB (rB-mB-) gal dcm (DE3) pLysE (CamR)*	Invitrogen	Protein expression
BL21(DE3) pLysS	F- *ompT hsdSB (rB-mB-) gal dcm (DE3) pLysS (CamR)*	Invitrogen	Protein expression
BL21 AI™	F- *ompT hsdSB(rB- mB -) gal dcm araB::T7RNAP-tetA*	Invitrogen	Protein expression
C41 (DE3)	F- *ompT hsdSB (rB- mB-) gal dcm (DE3)*	Lucigen Corporation	Protein expression

2.4. Plasmids construction

Plasmid constructs **(Table 2-2)** were derived from either pGEM-T Caf1, pBad33SD Caf1, described previously (Roque *et al.*, 2014), pTolT described by (Anderluh *et al.*, 2003) or pET28a vectors. pCOP was constructed by substitution of the T7 promoter with a random sequence of the same length in pGEM-T Caf1 using a Q5 Site-directed mutagenesis kit from New England Biolabs as stated in (2.8). All other plasmids were constructed using an In-Fusion HD cloning kit (Clontech) as stated in (2.9). Primers sequences are shown in **Table 2-3.**

Table 2-2. The plasmid constructs used in this study

Plasmid Name	Backbone	Genotype	Description
pCOP	pGEM-T	ΔT7, *caf1R, caf1M, caf1A, caf1, Amp^R, f1 ori*	T7 promoter was substituted with randomly selected sequence using pT7-COP
pCOPΔR	pGEM-T	ΔT7 *Δcaf1R, caf1M, caf1A, caf1, Amp^R, f1 ori*	*caf1R* gene was deleted from pCOP
pT7-COP	pGEM-T	T7, *caf1R, caf1M, caf1A, caf1, Amp^R, f1 ori*	Taken from (Roque *et al.*, 2014)

pT7-COPΔR	pGEM-T	T7, *Δcaf1R, caf1M, caf1A, caf1, Amp^R, f1 ori*	*caf1R* gene was deleted from pT7-COP
pT7-COPΔRSC	pGEM-T	T7, *Δ* start codon of *caf1R* with three stop codons inserted*, caf1M, caf1A, caf1, Amp^R, f1 ori*	The start codon of *caf1R* was substituted with randomly selected 3 bp and 3 stop codons were inserted using pT7-COP
pCOPF	pGEM-T	ΔT7, *caf1R^{FLAG}, caf1M^{FLAG}, caf1A^{FLAG}, caf1, Amp^R, f1 ori*	FLAG-tag sequence was added to the CT of *caf1R*, *caf1M* and *caf1A* using pCOP
pCOPFΔR	pGEM-T	ΔT7, *Δcaf1R, caf1M^{FLAG}, caf1A^{FLAG}, caf1, Amp^R, f1 ori*	FLAG-tag sequence was added to the CT of *caf1M* and *caf1A* using pCOPΔR
pCOPFΔM-RBS	pGEM-T	ΔT7, *caf1R, caf1M, caf1A, caf1, ΔRBS of caf1M, Amp^R, f1 ori*	The predicted RBS of *caf1M* was substituted with a

			randomly selected sequence using pCOPF
pCOPFΔA-RBS	pGEM-T	ΔT7, *caf1R, caf1M, caf1A, caf1, ΔRBS of caf1A, Amp^R, f1 ori*	The predicted RBS of *caf1A* was substituted with a randomly selected sequence using pCOPF
pCOPFΔF-RBS	pGEM-T	ΔT7, *caf1R, caf1M, caf1A, caf1, ΔRBS of caf1, Amp^R, f1 ori*	The predicted RBS of *caf1* was substituted with a randomly selected sequence using pCOPF
pCOPFΔP1	pGEM-T	ΔT7, *caf1R, caf1M, caf1A, caf1, Δ*predicted *promoter 1, Amp^R, f1 ori*	The predicted Promoter (P1) was deleted from pCOPF
pCOPFΔP2	pGEM-T	ΔT7, *caf1R, caf1M, caf1A, caf1, Δ*predicted *promoter 2, Amp^R, f1 ori*	The predicted Promoter (P2) was deleted from pCOPF

pCOPFΔP3	pGEM-T	ΔT7, *caf1R, caf1M, caf1A, caf1, Δ*predicted *promoter 3, Amp^R, f1 ori*	The predicted Promoter (P3) was deleted from pCOPF
pR-GFP	pGEM-T	ΔT7, *intergenic region 1, caf1R, gfp, Amp^R, f1 ori*	The sequences of *intergenic 1*, *caf1R* and *gfp* gene were inserted using pGEM-TΔT7
pR-T7 GFP	pGEM-T	T7, *intergenic region 1, caf1R, gfp, Amp^R, f1 ori*	The sequences of *intergenic 1*, *caf1R* and *gfp* gene were inserted using pGEM-T
pRC1-GFP	pGEM-T	ΔT7, *intergenic region 1,* insulated constitutive promoter*, caf1R, gfp, Amp^R, f1 ori*	Insulated constitutive promoter was inserted to pR-GFP
pRC2-GFP	pGEM-T	ΔT7, *intergenic region 1,* insulated constitutive promoter*, caf1R, gfp, Amp^R, f1 ori*	Insulated constitutive promoter was inserted to pR-GFP

pRC3-GFP	pGEM-T	ΔT7, *intergenic region 1,* insulated constitutive promoter*, intergenic 3, caf1R, gfp, Amp^R, f1 ori*	Intergenic region 3 was inserted to pRC2-GFP
pRC3-GFPΔR	pGEM-T	ΔT7, *intergenic region 1,* insulated constitutive promoter*, intergenic 3, Δcaf1R, gfp, Amp^R, f1 ori*	*caf1R* gene was deleted from pRC3-GFP
pBad Caf1R	pBad33	*caf1R, Cam^R, p15A ori*	*caf1R* gene was inserted in pBAD33
pBad Caf1RF	pBad33	*caf1R^{FLAG}, Cam^R, p15A ori*	FLAG-tag sequence was added to the CT of *caf1R* using pBad Caf1R
pBad Caf1M	pBad33	*caf1M, Cam^R, p15A ori*	*caf1M* gene was inserted in pBAD33
pBad Caf1A	pBad33	*caf1A, Cam^R, p15A ori*	*caf1A* gene was inserted in pBAD33

pBad Caf1	pBad33	*caf1, Cam^R, p15A ori*	*caf1* gene was inserted in pBAD33
pET28A Caf1R^{His}	pET28a	*caf1R^{CT-His}, Kan^R, Pbr322 ori*	His-tagged *caf1R* was inserted in pET28A
pTolT Caf1R	pET8c	*tolAIII^{NT-His}, caf1R, Amp^R, Pbr322 ori*	*caf1R* gene was added to pTolT (Anderluh *et al.*, 2003)
pTolT Caf1RCO	pET8c	*tolAIII^{NT-His}, codon optimized caf1R, Amp^R, Pbr322 ori*	Codon optimized caf1R gene were added to pTolT (Anderluh *et al.*, 2003)

Table 2-3. The sequences of primers used in this study

Name	Sequence (5' -3')	Function
Forward ΔT7	TCTAGCCGTTCGAATTGGGCCC GACGTC	Substitution of T7 promoter
Reverse ΔT7	AGAACAAGAACAATTCACTGG CC GTCGTTTTAC	Substitution of T7 promoter

Forward *Δcaf1R*	TAATCCTAATGTTACAGAATATAACCCAAATCAAAATAATAG	Deletion of *caf1R*
Reverse *Δcaf1R*	GTAACATTAGGATTACCAAAGAG	Deletion of *caf1R*
Forward *caf1R*^{FLAG}	TCATCATCATCTTTATAATCACTCTTTGGTAATCCTAATGTTACTGAC	Addition of C-terminal FLAG tag to *caf1R* in pCOP
Reverse *caf1R*^{FLAG}	AAAGATGATGATGATAAATAAAATTCCCGCGGCCGCCATG	Addition of C-terminal FLAG tag to *caf1R* in pCOP
Forward *caf1M*^{FLAG}	AAAGATGATGATGATAAATAATGATGTTTAAAGGGGACGGG	Addition of C-terminal FLAG tag to *caf1M* in pCOP
Reverse *caf1M*^{FLAG}	ATCATCATCATCTTTATAATCTAAAGTCACATTTTTGGAATACAAAC	Addition of C-terminal FLAG tag to *caf1M* in pCOP
Forward *caf1A*^{FLAG}	AAAGATGATGATGATAAATAAAACGGATGTTTATTTCAAACAGGACAC	Addition of C-terminal FLAG tag to Caf1A in pCOP
Reverse *caf1A*^{FLAG}	ATCATCATCATCTTTATAATCGTTATTTAAGATGCAGGTTGTGGATAAC	Addition of C-terminal FLAG tag to *caf1A* in pCOP
Forward *Δcaf1M-RBS*	GACGTACTAAGCTCATGATTTTAAATAGATTAAGTACG	Substitution of RBS of *caf1M* in pCOP
Reverse *Δcaf1M-RBS*	TGAGCTTAGTACGTCTACGGAATGGTGACAACACCTTC	Substitution of RBS of *caf1M* in pCOP
Forward *Δcaf1A-RBS*	GCAATCAGGGAATAATGAGGTATTCAAAGC	Substitution of RBS of *caf1A* in pCOP

Reverse *Δcaf1A-RBS*	TTATTCCCTGATTGCTTAAACA TCATCATAAAGTCACATTTTTG	Substitution of RBS of *caf1A* in pCOP
Forward *Δcaf1-RBS*	GCTGTACAATATATGAAAAAA ATCAGTTCCGTTATCGC	Substitution of RBS of *caf1* in pCOP
Reverse *Δcaf1-RBS*	CATATATTGTACAGCTCGAATA ATCCAATCCACGAACAAATTC	Substitution of RBS of *caf1* in pCOP
Forward *Δcaf1R start codon*	TAGTGATAATAGTAGACTGTA AATTCAATTATTCAATATATAG AAGAG	Substitution of *caf1R* start codon and addition of 3 stop codons
Reverse *Δcaf1R start codon*	CTACTATTATCACTAAGAATAT AACCCAAATCAAAATAATAGC ATTC	Substitution of *caf1R* start codon and addition of 3 stop codons
Forward *caf1A*^{FLAG} in pBAD33	ATCATCATCATCTTTATAATCA CTCTTTGGTAATCCTAATGTTA CTGAC	Addition of C-terminal FLAG tag to *caf1R* in pBAD33
Reverse *caf1A*^{FLAG} in pBAD33	AAAGATGATGATGATAAATAA AATTCCCGCGGCCGCCATG	Addition of C-terminal FLAG tag to *caf1R* in pBAD33
Forward pTolT amplification	GGATCCGCGCGGAACCAGAG ATCCCCCACCCGGTTTGAAGTC CAA	Amplification of pTolT vector for pTolT *caf1R* cloning
Reverse pTolT amplification	GGTACCTGATGAACGCGTGAG GAATTTTGAAGATCC	Amplification of pTolT vector for pTolT *caf1R* cloning
Forward *caf1R* amplification	CATGCATCACCATCACCATCAC	Amplification of *caf1R* for pTolT *caf1R* cloning

Reverse *caf1R* amplification	CTTATTAGCTTTTCGGCAGACCC	Amplification of *caf1R* for pTolT Caf1R cloning
Forward pTolT amplification	GGATCCGCGCGGAACCAGAGATCCCCCACCCGGTTTGAAGTCCAA	Amplification of pTolT vector for pTolT Caf1RCO cloning
Reverse pTolT amplification	GGTACCTGATGAACGCGTGAGGAATTTTGAAGATCC	Amplification of pTolT vector for pTolT Caf1RCO cloning
Forward codon optimised *caf1R* amplification	GTTCCGCGCGGATCCATGCTGAAACAAATGACCGTGAATAG	Amplification of *caf1R* for pTolT Caf1RCO cloning
Reverse Forward codon optimised *caf1R* amplification	CGTTCATCAGGTACCTTATTAGCTTTTCGGCAGACCCAG	Amplification of *caf1R* for pTolT Caf1RCO cloning
Forward *ΔPromoter1*	AATCCCCTTCATTTGTTACCCACC	Deletion of predicted promoter 1
Reverse *ΔPromoter1*	CAAATGAAGGGGATTCAATTTTATTTAAAAATGCACACAAAGTTTAGC	Deletion of predicted promoter 1
Forward *ΔPromoter2*	GAAATGATGGGGAGGGGGTGG	Deletion of predicted promoter 2
Reverse *ΔPromoter2*	CCTCCCCATCATTTCGGATTTTTATATCCGTAGCACAGCC	Deletion of predicted promoter 2

Forward *ΔPromoter3*	GGGGTGGGAAGGTGTTGTCAC	Deletion of predicted promoter 3 from pCOP
Reverse *ΔPromoter3*	ACACCTTCCCACCCCGTAACAAATGAAGGGGATTTTTATATCCG	Deletion of predicted promoter 3 from pCOP
Reverse transcription RACE	CCATAATTG	(5’ RACE) Reverse transcription reaction
Gene-specific primer RACE	GTAATTGGAGCGCCTTCCTTATATGTAAG	(5’ RACE)
(dT)17-adaptor RACE	GACTCGAGTCGACATCGATTTTTTTTTTTTTTT	(5’ RACE)
Adaptor RACE	GACTCGAGTCGACATCG	(5’ RACE)
Forward pGEM-T *caf1* amplification	TCCATATAGATAATAGATAAAGGAGGG	Amplification of pGEM-T vector including *caf1R* and *intergenic region 1* for pR-GFP
Reverse pGEM-T *caf1* amplification	GAGCTTAACCTCCTTACGGAATG	Amplification of pGEM-T vector including *caf1R* and *intergenic region 1* for pR-GFP
Forward *gfp* amplification	AAGGAGGTTAAGCTCATGAGTAAAGGAGAAGAACTTTTCAC	Amplification of *gfp* for pR-GFP
Reverse *gfp* amplification	TATTATCTATATGGACTATTTGTATAGTTCATCCATGCCATG	Amplification of *gfp* for pR-GFP
Forward	GCATTTTTAAATAAAATTGTTCTCAGTGAG	Amplification of pR-GFP for pRC1-GFP cloning

pR-GFP amplification		
Reverse pR-GFP amplification	ACACAAAGTTTAGCTTTTCGCGC	Amplification of pR-GFP for pRC1-GFP cloning
Forward amplification of insulated promoter	AGCTAAACTTTGTGTTTCTAGAGCACAGCTAACACCAC	Amplification of the insulated constitutive promoter for pRC1-GFP cloning
Reverse amplification of insulated promoter	TTTATTTAAAAATGCCTCTAGTAAAAGTTAAACAAAATTATTTGTAG	Amplification of the insulated constitutive promoter for pRC1-GFP cloning
Forward pR-GFP amplification	GGGAGGGGGTGGGAAGG	Amplification of pR-GFP for pRC2-GFP cloning
Reverse pR-GFP amplification	CATCATTTCATATCGACGATATGC	Amplification of pR-GFP for pRC2-GFP cloning
Forward amplification of insulated promoter	CGATATGAAATGATGTTCTAGAGCACAGCTAACACCAC	Amplification of the insulated constitutive promoter for pRC2-GFP cloning
Reverse amplification of insulated promoter	TTCCCACCCCCTCCCCTCTAGTAAAAGTTAAACAAAATTATTTGTAG	Amplification of the insulated constitutive promoter for pRC2-GFP cloning

Forward pRC2-GFP amplification	ATGAGTAAAGGAGAAGAACTTTTCAC	Amplification of pRC2-GFP for pRC3-GFP cloning
Reverse pRC2-GFP amplification	CTCTAGTAAAAGTTAAACAAAATTATTTGTAG	Amplification of pRC2-GFP for pRC3-GFP cloning
Forward *intergenic 3* amplification	TAACTTTTACTAGAGAACGGATGTTTATTTCAAACAGGAC	Amplification of intergenic region 3 for pRC3-GFP cloning
Reverse *intergenic 3* amplification	TTCTCCTTTACTCATATATTACCTCTATCGAATAATCCAATC	Amplification of intergenic region 3 for pRC3-GFP cloning

2.5. TB medium preparation

The TB medium was prepared by mixing 24 g of yeast extract, 12 g of tryptone and 4 ml glycerol. The mix was dissolved in deionised water up to 900 ml and autoclaved. 100 ml of sterile phosphate buffer pH 7.4 (0.17 M KH_2PO_4 and 0.72 M K_2HPO_4) was added prior to use directly.

2.6. LB medium preparation

The LB medium was prepared by adding 10 g of tryptone, 5 g of yeast extract and 5 g of NaCl, dissolved in deionised water up to 1 litre and autoclaved to be used in the experiments.

2.7. LB medium for agar plate

The medium was prepared by adding 10 g of tryptone, 5 g of yeast extract, 5 g of NaCl and 20 g of agar, dissolved in deionised water up to 1 litre and autoclaved. The required antibiotics were added to the media prior to pouring the plates. The final concentrations of antibiotics were as follows: ampicillin 100 µg/ml, chloramphenicol 20 µg/ml and kanamycin 30 µg/ml.

2.8. Q5-site directed mutagenesis

According to New England Biolabs protocol, the exponential amplification was performed by mixing of 12.5 µl Q5 hot start high-fidelity 2X master mix, 1.25 µl of 10 µM forward primer, 1.25 µl of 10 µM reverse primer, 1 µl of 10 ng/µl template DNA (pT7-COP) and 9 µl of nuclease-free water. The mix was placed in T100 Thermal cycler from BIO-RAD to perform the PCR reaction with the cycling conditions stated in **Table 2-4**. KLD (Kinase, Ligase and DpnI enzymes) reaction was achieved by mixing 1 µl of PCR product, 5µl of 2X KLD reaction buffer, 1 µl of 10 X KLD enzyme mix and 3 µl of nuclease-free water. The reaction mix was incubated for 5 min at room temperature. The transformation was performed by adding 5 µl of KLD reaction mix to 50 µl of chemically-competent cells. The cells plus mix were incubated on ice for 30 min, heat shocked at 42^{o}C for 1 min and then incubated on ice for 5 min. 950 µl of SOC medium was added to the mix and incubated at 37^{o}C with shaking for 1 h. 100 µl of the mix was spread onto LB agar medium in a petri Dish with appropriate antibiotics and incubated at 37^{o}C overnight. Three single colonies were grown in separate LB liquid medium with appropriate antibiotics at 37^{o}C overnight. The plasmids were purified from these cultures as stated in (2.11) to be sequenced by Eurofins Genomics as in (2.23).

Table 2-4. The cycling conditions of Q5 Site-directed mutagenesis

Step	Temperature	Time
Initial denaturation	98°C	30 s
25 Cycles	98°C	10 s
	50-72°C	30 s
	72°C	25 s/kb
Final extension	72°C	2 min
Hold	4-10°C	

2.9. In-Fusion HD cloning

In-Fusion cloning technique from Clontech offers a quick mutagenesis of DNA including insertion, deletion and substitution. First of all, for each vector we identified the site of insertion, deletion or substitution. For insertion cloning, PCR primers were designed for the target gene with 15-bp (5') tail complementary to the linearized vector's ends. As shown in **Figure 2-1** the target gene was amplified by PCR and purified from agarose gel using a Monarch kit as stated (2.12). Deletions were made by omitting the relevant region of the plasmid from the amplified region by using two primers designed to flank the target gene with 15-bp in one primer complementary to the other end of the linearized vector, then ligating the amplified region back together. Substitution reaction was performed as in deletion, but in substitution, one of the two primers contains the sequence to be inserted

and the other contains 15-bp complementary to the other end of the linearized vector. The PCR reaction was performed as mentioned in (2.10). The PCR product in each case was analysed by agarose gel electrophoresis. The target DNA band was extracted from the gel using Monarch Kit New England Biolabs as stated in (2.12). The In-Fusion HD cloning reaction mix was prepared by adding 2 µl of In-Fusion HD Enzyme Premix from Clontech, X µl of the linearized vector, X µl of the insert (the volume and concentration of both linearized vector and insert are calculated using In-Fusion HD cloning calculator) from Clontech and Nuclease-free water up to 10 µl as a final reaction volume. The cloning reaction mix was incubated at 50°C for 15 min. The reaction mix was diluted 5 times prior to use in transformation of chemically competent cells as in (2.13).

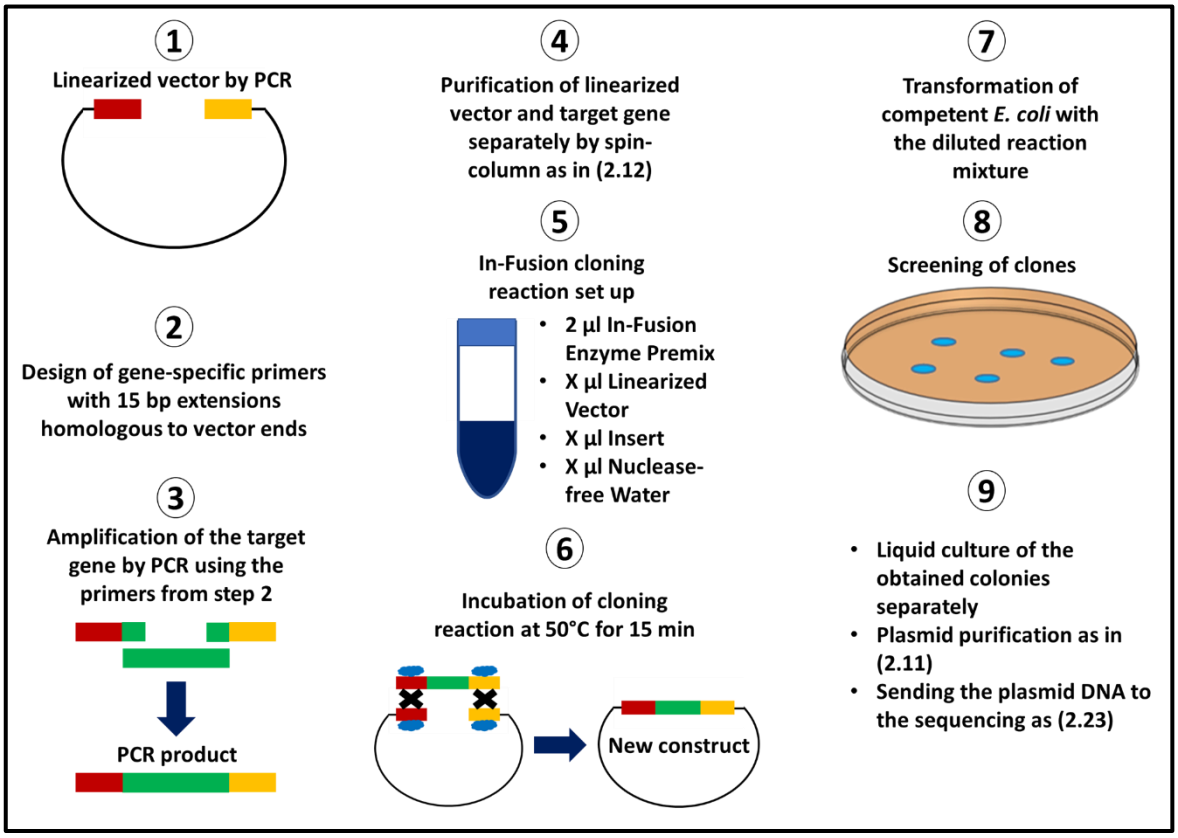

Figure 2-1. The workflow of the In-Fusion HD cloning technique.

2.10. Polymerase Chain Reaction (PCR) using Phusion DNA Polymerase

The PCR mixture was prepared by adding 10 µl of 5X Phusion GC buffer supplied with the kit from New England Biolabs, 1 µl of 10 mM dNTPs, 2.5 µl of 10 µM forward primer, 2.5 µl of 10 µM reverse primer, 1.5 µl of DMSO, 2 µl of template DNA (10 ng/µl), 0.5 µl of Phusion DNA polymerase enzyme and 30 µl of nuclease-free water. The cycling conditions are shown in **Table 2-5**.

Table 2-5. The cycling conditions of the PCR using Phusion DNA polymerase

Step	Temperature	Time
Initial denaturation	98^{o}C	30 s
25 Cycles	98^{o}C	10 s
	50-72^{o}C	30 s
	72^{o}C	25 s/kb
Final extension	72^{o}C	10 min
Hold	4-10^{o}C	

2.11. Plasmid DNA purification from liquid culture medium

This was performed according to the manufacturer's protocol which is New England Biolabs. Briefly, 4 ml of a liquid culture of the required transformed *E. coli* cells were pelleted in a 5 ml Eppendorf tube by centrifugation for 1 min at 16000 x g and supernatant discarded. The cell pellet was resuspended completely in 200 µl of plasmid resuspension buffer (B1) by pipetting up and

down several times. Cells were lysed by adding 200 µl of lysis buffer (B2) and the tube was inverted gently 5-6 times until the solution become dark pink, clear and viscous. The solution was incubated for 1 min at room temperature. Next, 400 µl of Plasmid Neutralization Buffer (B3) was added to the mixture to neutralize the solution and the tube was inverted gently until the formation of precipitate and the solution colour change to yellow.

The solution was incubated at room temperature for 2 min and then centrifuged at 16000 x g for 5 min to clarify the lysate. The supernatant was transferred to the spin column supplied in the kit, centrifuged at 16000 x g for 1 min and flow-through was discarded. 200 µl of Plasmid Wash Buffer 1 was added to the spin column and centrifuged at 16000 x g for 1 min to remove RNA, protein and endotoxin and flow-through was discarded. 400 µl of Plasmid Wash Buffer 2 was added to the spin column and centrifuged at 16000 x g for 1 min. The column was transferred to a clean 1.5 ml tube and 40 µl of nuclease-free water was added to column in order to elute the plasmid DNA.

2.12. Plasmid DNA purification from agarose gel

According to the manufacturer's protocol (New England Biolabs), the DNA fragment was excised from the gel using a scalpel, dissolved in 4 volumes dissolving buffer and incubated at 50^{o}C for 10 min with vortexing periodically to dissolve the gel completely. The solution was transferred to a spin column to be centrifuged at 16000 x g for 1 min and the flow-through was discarded. The bound DNA was washed by adding 200 µl of DNA Wash Buffer and centrifugation at 16000 x g for 1 min (this step was repeated two times). The

column was transferred to a clean 1.5 ml tube and 12 µl of Nuclease-free water was added to column in order to elute the plasmid DNA.

2.13. Transformation of bacterial cells

According to the manufacturer's protocol (New England Biolabs), the competent *E. coli* cells were thawed on ice for 10 min, 2 µl of the plasmid DNA (100 ng/µl) were added to the thawed competent cells with a gentle flicking to mix DNA and cells. The mixture was placed on ice for 10 min and then heat shocked for 1 min at 42^{o}C. The mixture was placed on ice again for 5 min, 950 µl of SOC medium was added to the mix and incubated at 37^{o}C with shaking for 1 h. 100 µl of the mix was spread onto LB agar Petri Dish with appropriate antibiotics and incubated at 37^{o}C overnight.

2.14. The analysis of flocculent layer production

The TB liquid medium containing appropriate antibiotics was inoculated with the required transformant, grown at 35^{o}C for 16 h. The flocculent layer heights were measured using a ruler after the centrifugation of glass capillary tubes containing 50 µl of *caf1* expression cultures sealed at the bottom using plasticine as in **Figure 2.2**. The centrifugation was performed at 2367 x g, 22^{o}C for 15 min.

2.15. Caf1R purification

E. coli C41 (DE3) cells were transformed with pTolT Caf1RCO shown in **Table 2.2**. The starter liquid culture was prepared by inoculating 50 ml of TB medium containing 100 µg/ml ampicillin and 0.2% D-glucose to reduce the

basal expression of the target gene. The starter culture was grown at 37°C, shaking at 180 rpm overnight and then used to inoculate the expression cultures. The expression cultures were prepared using 2 x 2 litre shake flasks each one containing 1 litre of TB medium with 100 µg/ml ampicillin and 0.2% D-glucose. The expression cultures were inoculated, grown at 37°C, 180 rpm for ≈ 3 h until OD_{600} ≈ 0.8. The cultures were induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown for additional 3 h.

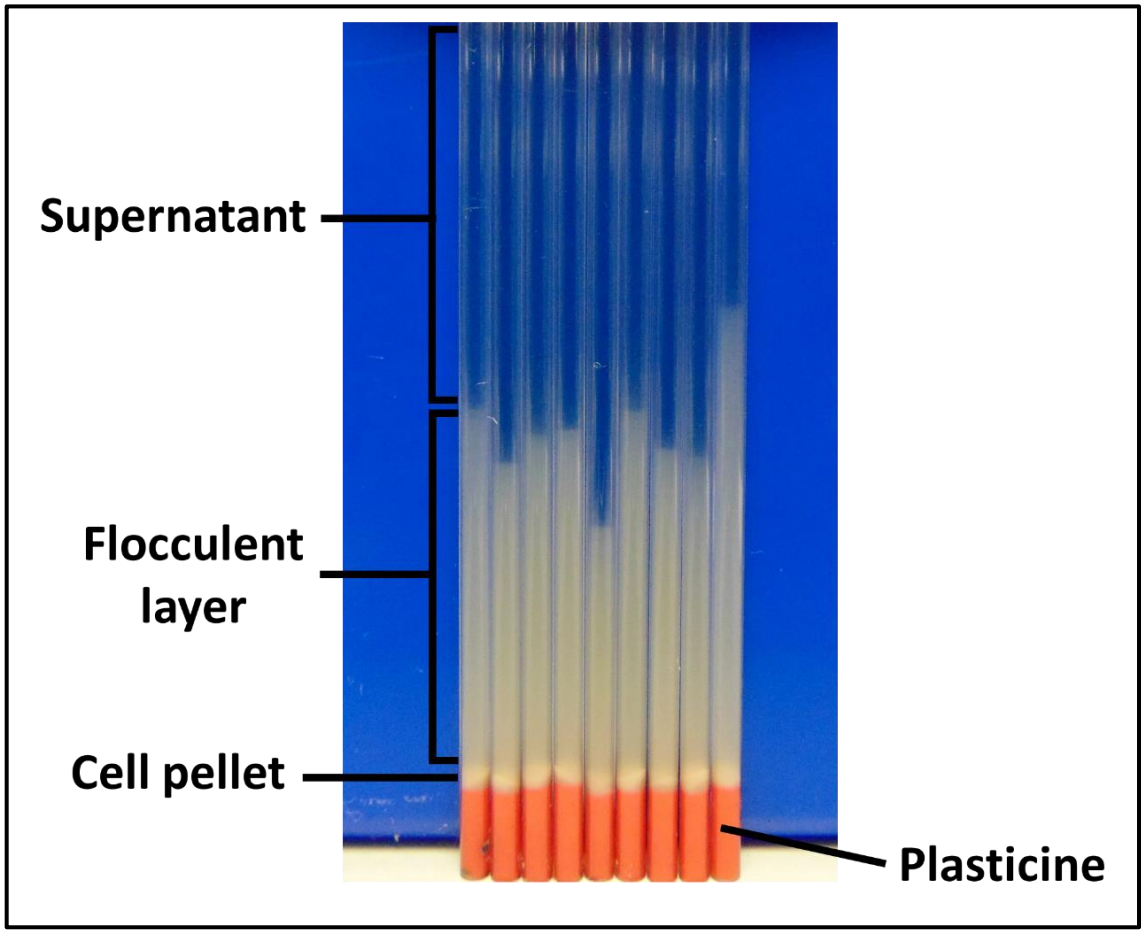

Figure 2-2. The separation of the flocculent layer in glass capillary tubes. Showing 9 glass capillary tubes sealed at the bottom using plasticine (red color) containing 9 different cultures of *E. coli* BL21 (DE3) transformed with pT7-COP and centrifuged at 2367 x g, 22°C for 15 min. The cell pellet is shown at the bottom of capillary tubes above the plasticine, the flocculent layer at the middle and the supernatant at the top.

The expression cultures were centrifuged at 7000 x g for 30 min using a JA-10, fixed angle rotor from Beckman. The cell pellet was resuspended in 50 mM phosphate buffer containing 25 mM imidazole pH 8 (5 ml buffer per 1 g of cell pellet). The cell disruption was achieved using a high pressure homogenises (Constant Systems LTD) at 23 Kpsi. The cell lysate was centrifuged at 30000 x g for 30 min using a JA-25.50, fixed angle rotor from Beckman. The supernatant was applied to a 5 ml HisTrap HP column (His-tag purification column) on an AKTA^{TM} Pure system (the fully automated protein purification system from GE healthcare) as recommended by manufacturer. The HisTrap HP column was washed with 5 column volumes of Nanopure water (with 5 ml/min flow rate in all steps), then equilibrated with 5 column volumes 50 mM phosphate buffer containing 25 mM imidazole pH 8.

80 ml of filtered cell lysate was injected in the system to be loaded on the column automatically. The column was washed with 10 column volumes of 50 mM phosphate buffer containing 25 mM imidazole pH 8 and eluted with 5 column volumes of 50 mM phosphate buffer containing 350 mM imidazole pH 8. The fractions were collected automatically. The protein concentrations were estimated for the collected fractions from both wash and elution steps by the UV monitor in the system and then selected fractions were analysed by SDS-PAGE. The eluted fractions with high protein concentration and producing protein bands close to the expected size were mixed and concentrated using 30 kDa molecular weight cut-off Vivaspin centrifugal concentrators from Sartorius by centrifugation at 2360 x g for 1 h.

The concentrated sample 1 ml of 3.3.mg/ml protein was used to perform size exclusion chromatography. The sample was injected in AKTA pure system to be loaded automatically on the Proteo SEC 3 - 70 kDa HR column after

equilibration. The equilibration was performed with 2 column volumes of phosphate buffered saline pH 7.4 (PBS). The elution was achieved using 1 column volume PBS with flow rate 1 ml/min for all steps. The eluted fractions were collected automatically according to their absorbance at 280_{nm}. All collected fractions were analysed by SDS-PAGE.

2.16. Bacterial culture preparation for RT-PCR and western blot

Transformations were performed using *E. coli* BL21 (DE3) from New England Biolabs. Three different colonies were selected from each plate and grown in LB medium overnight at 37°C, 180 rpm. Glycerol stocks were prepared for all transformed bacteria using 500 µl of bacterial culture and 500 µl of 60% (v/v) glycerol solution. Expression cultures were prepared using 5 ml TB medium containing 100 µg/ml ampicillin and/or 20 µg/ml chloramphenicol and were inoculated using a stab from one of the glycerol stocks. To induce expression from pBad plasmids, L-arabinose was added at the concentrations described in the text at the same time as the cultures were inoculated.

For analysis of *caf1* operon transcript levels, bacterial cells were cultured in triplicate from glycerol stocks of *E. coli* BL21 (DE3) cells transformed with pCOPF stated in **Table 2.2** and grown at 25°C overnight. The OD_{600nm} was measured for all cultures and then 2 ml were taken from each culture to be analysed by RT-PCR and western blot. In order to analyse the thermal responsive regulation of *caf1* operon, the remaining cultures were diluted to an OD_{600nm} of 0.5 and then each culture was split into two cultures: one to be incubated at 25°C and the other simultaneously at 35°C. After one-hour

incubation the OD_{600nm} was measured and 2 ml samples taken from each culture for RT-PCR and western blot analysis.

2.17. SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Bacterial cultures (1 ml of each) were centrifuged at 20,000 x g for 5 min. Supernatants and flocculent layers were discarded and cell pellets resuspended in 100 µl per OD_{600} of 100 mM DTT, 2% w/v SDS. The samples were heated at 95^{o}C for 10 min and centrifuged at 20,000 x g for 10 min. 50 µl of this supernatant was mixed with 50 µl SDS loading buffer (2% w/v SDS, 0.1% w/v Bromophenol blue, 5 mM EDTA, 125 mM Tris pH 6.8, 15% v/v Glycerol and 1% v/v β-mercaptoethanol), heated at 95^{o}C for 5 min and centrifuged at 20,000 x g for 5 min. The samples (10 µl) were resolved on SDS-PAGE. For the flocculent layer samples analysis, 20 µl of the flocculent layer was mixed with 100 µl SDS loading buffer (2% w/v SDS, 0.1% w/v Bromophenol blue, 5 mM EDTA, 125 mM Tris pH 6.8, 15% v/v Glycerol and 1% v/v β-mercaptoethanol), boiled for 10 min and centrifuged at 20,000 x g for 5 min. The samples (10 µl) were resolved on SDS-PAGE at 200 V for 60 min and then the SDS-PAGE gel either stained with Coomassie Brilliant Blue to show the protein bands or used in western blot (Lambin *et al.*, 1976).

2.18. Western blot

The SDS-PAGE was performed as stated in (2.17) without staining. Nitrocellulose membranes and blotting papers were soaked in 10 mM N-cyclohexyl-3-aminopropanesulfonic acid buffer (10mM CAPS buffer pH 11) containing 20% methanol for 10 min. SDS-PAGE gels were soaked in the same buffer for 2 min. The blotting paper, nitrocellulose membrane and gel

were assembled in a semi-dry blotter (Trans-Blot SD semi-dry transfer cell, Biorad) and 18V applied for 30 min. Blots were stained with Ponceau S solution (Sigma) for 10 min to view the efficiency of protein transfer and then rinsed with PBS buffer pH 7.4. The blots were blocked with TBS buffer (2.7 mM KCl, 38 mM Tris-HCl and 140 mM NaCl pH 8) containing 5% w/v milk (from powder) at room temperature for 2 h and then rinsed with (1x) Tris Buffered Saline pH 8 (TBS buffer). The membranes were incubated for 4 h at room temperature or overnight at 4^{o}C with 2 ml 5% w/v milk solution, 1 µg/ml mouse anti-FLAG antibody (Sigma) in order to detect the FLAG-tagged proteins and 6.9 µg/ml mouse anti-Caf1 antibody (Abcam) for Caf1 protein detection. The blots were washed with (1x) TBS buffer pH 8 two times for 5 min at room temperature then incubated for 4 h at room temperature or overnight at 4^{o}C with 2 ml TBS buffer pH 8 containing 2.5-5 µg/mL rabbit anti-mouse antibody conjugated with horseradish peroxidase (Sigma). The blots were washed twice for 15 min with TBS buffer pH 8 at room temperature, covered with developing solution and incubated at room temperature for a few min with shaking. Developing solution was prepared by dissolving 50 mg 4-chloronaphthol in 10 mL methanol mixed with 50 ml developing buffer (20 mM Tris-HCl, 140 mM NaCl and 1 mM Na_2HPO_4 pH 7.2) containing 60 µl hydrogen peroxide. The blots were dried in air and the images were taken using a gel documentation system (Gel DocTM XR+, Biorad).

2.19. RNA extraction

RNA was isolated using an EZ-10 Total RNA Mini-Preps Kit (Bio Basic Inc.), according to the manufacturer's protocol, by transferring 1 ml from each

culture (≈ 2 x 10^9 cells) separately to be centrifuged at 10,000 x g for 1 min. Supernatants were discarded and 100 µl of lysozyme solution (400 µg ml-1 lysozyme in RNase-free water) added to each sample pellet. The mixtures were suspended thoroughly and incubated at 37^{o}C for 5 min. 300 µl buffer RLT lysis buffer supplied with the kit was added to each sample, mixed thoroughly by vortexing and incubated at room temperature for 5 min. 250 µl of ethanol was added to each sample and mixed by inverting the tubes. EZ-10 spin columns were placed in 2ml collection tubes and then the mixtures were transferred to the spin columns, centrifuged at 12,000 x g for 2 min at room temperature and the flow-through was discarded. The EZ-10 Columns were placed in the collection tubes, 350 µl of DW Solution supplied with the kit was added to each sample, centrifuged at 12,000 x g for 1 min at room temperature and the flow-through were discarded. Preparation of DNase I solution was performed by mixing 10 µl RNase-free water, 40 µl buffer RDD and 30 µl DNase I (30 U), this mix (80 µl) is enough for one sample. Thus, 80 µl of DNase I solution was added to each EZ-10 spin column to the centre of the membrane and incubated at 37^{o}C for 30 min. 350 µl of DW solution was added to each spin column, kept at room temperature for 3 min, centrifuged at 12,000 x g for 1 min at room temperature and the flow-through were discarded. 500 µl of RPE solution supplied with the kit was added to each column, centrifuged at 12,000 x g for 1 min at room temperature and the flow-through was discarded (addition of RPE solution was repeated two times). The spin columns were centrifuged at 12,000 x g for 2 min at room temperature and the open columns were incubated at room temperature for 5 min until the ethanol evaporated. The columns were placed in a new RNase-free 1.5ml centrifuge tubes, 50 µl of RNase-free water

was added in each column, incubated at room temperature for 5 min and centrifuged at 12,000 x g for 1 min at room temperature. The concentration of RNA was measured at OD_{260} using Nanodrop UV spectrophotometer from Labtech. RNA samples were placed on ice to be used in cDNA synthesis or stored at -80^{o}C.

2.20. cDNA synthesis by reverse transcription reaction

The cDNA synthesis was performed using a QuantiTect Reverse Transcription Kit from QIAGEN. According to the manufacturer's protocol, the genomic DNA elimination reaction was prepared by mixing 12 µl from each RNA sample with final concentration of 1 µg with 2 µl of gDNA Wipeout Buffer, incubated for 2 min at 42^{o}C and then placed on ice immediately. The reverse transcription master mix was prepared by mixing 1 µl Quantiscript Reverse Transcriptase, 4 µl Quantiscript RT Buffer and 1 µl RT Primer mix (per one sample). The genomic DNA elimination reaction product was added to the reverse transcription master mix, incubated at 42^{o}C for 30 min and then at 95^{o}C for 3 min to inactivate Quantiscript Reverse Transcriptase. The concentration of cDNA was measured at OD_{260} using Nanodrop UV spectrophotometer (Labtech). The cDNA samples were kept at -20^{o}C for later use in RT-PCR experiment.

2.21. Real-Time Polymerase Chain Reaction (RT-PCR)

The master reaction mix for RT-PCR experiments was prepared using GoTaq Flexi DNA Polymerase from Promega by adding 4 µl nuclease-free water, 4 µl 5X Colorless GoTaq Flexi Buffer, 3 µl of 2 mM dNTPs, 3.2 µl of 25 mM $MgCl_2$, 0.2 µl SYBR Green (Sigma, S9430) (diluted 200 times with 100%

DMSO), 0.2 µl GoTaq Flexi DNA Polymerase, 0.8 µl 10 µM primer mix and 5 µl of 50 ng/µl cDNA per reaction. Samples were loaded in a Rotor-Gene Q instrument (Qiagen) and critical threshold cycle values (Ct values) were collected to measure the fold change in gene expression for each target gene, relative to β-*lactamase* transcription from the plasmid, with a threshold level of 0.5. Primer sequences are shown in **Table 2.3**. Thermal cycling conditions are stated in **Table 2.6**. The expression fold change was calculated as in (Livak and Schmittgen, 2001; Dheda *et al.*, 2004), the fold change in expression = ($2^{-\Delta Ct}$) and ΔCt = (Ct value of target gene – Ct value of β-*lactamase* gene). The Ct value is the required cycles number for the fluorescence signal to cross the threshold which is 0.5.

Table 2-6. The cycling conditions of the RT-PCR

Step	Temperature	Time	Number of Cycles
Initial Denaturation	95°C	10 min	1 cycle
Denaturation	95°C	30 s	40 cycles
Annealing	60°C	30 s	
Extension	72°C	30 s	

2.22. Rapid amplification of 5' cDNA ends (5' RACE)

5' RACE experiments were performed as described previously (Frohman *et al.*, 1988; Schaefer, 1995; Zhang and Frohman, 1997) with minor modifications. Briefly, cDNA was prepared by performing the reverse

transcriptase reaction using a 9 bp long gene specific primer stated in **Table 2.3** complementary to a 3' region of the *caf1* gene. cDNA samples were diluted two times with nuclease-free water and purified using 10 kDa molecular weight cut-off Vivaspin centrifugal concentrators (Sartorius) spun at 1000 x g for 30 min. This step was repeated to prepare the samples for poly-adenine tailing. Poly-adenine tailing was performed by mixing 0.5 µl terminal transferase enzyme (20 Units/µl, New England Biolabs), 5 µl 10 x terminal transferase buffer, 5 µl of 2.5 mM $CoCl_2$, 0.5 µl of 10 mM dATP for tailing, 5 pmol of cDNA to be tailed and nuclease-free water up to 50 µl. The mixture was incubated at 37^{o}C for 30 min and then heated at 70^{o}C for 10 min to inactivate the reaction.

Tailed cDNA was then amplified by PCR. The PCR reaction was prepared by mixing 0.5 µl of 2 U/µl Phusion DNA polymerase (New England Biolabs), 10 µl 5 x GC Phusion buffer, 1 µl of 10 mM dNTP mix, 1.2 µl of 10 µM (dT)17-adapter primer **(Table 2.3)**, 2.5 µl of 10 µM adapter primer **(Table 2.3)**, 2.5 µl of 10 µM gene specific primer **(Table 2.3)** and 3% DMSO with 2 µl of tailed cDNA and nuclease-free water up to 50 µl. Cycling conditions are stated in **Table 2.5**. The PCR product was analysed by agarose gel electrophoresis. The target DNA bands were extracted from the gel using a Monarch gel extraction kit (New England Biolabs) according to the manufacturer's protocol. The concentration of pure DNA was measured using a Nanodrop UV spectrophotometer (Labtech) and sequenced by Eurofins Genomics as in (2.23).

2.23. The preparation of DNA samples for sequencing

The DNA samples were prepared by adding 5 µl of template DNA with either of the purified plasmid DNA of (80 - 100 ng/µl) or purified PCR product of (20 - 80 ng/µl) to 5 µl of the sequencing primer of 5 µM. The characteristics of primers used for sequencing are as follows:

- The melting temperature (T_M) of the primer should be between 52^{o}C and 58^{o}C and the length should be between 17 - 19 bp. Ideally, the GC content of a 17 mer should be 10 GC.
- G or C should be at the 3' end, but not more than 3 Gs or Cs.
- The primer sequence should be a good mix of all 4 nucleotides with no more than 4 identical bases in a row (AAAA or GGGG).

2.24. Bioinformatics

Promoter sequences were predicted using the BPROM (Softberry Inc., Mount Kisco, NY, USA, http://www.softberry.com) (Solovyev and Salamov, 2011) and Neural Network Promoter Prediction 2.2 (NNPP2.2) (Reese, 2001) webservers. Sequence alignments were generated using the Clustal Omega (Sievers *et al.*, 2011) webserver and visualised using the ESPript (Robert and Gouet, 2014) webserver (http://espript.ibcp.fr). All protein structures and structural models were generated using SWISS-MODEL Workspace webserver (Waterhouse *et al.*, 2018). The 3D structures were visualised using CCP4mg software (McNicholas *et al.*, 2011).

2.25. Plate reader assays

A Labtech plate reader model was used to determine both OD_{600nm} and fluorescence at 485 nm of the GFP producing cultures. The plate wells were filled with 200 µl TB medium containing appropriate antibiotics (100 µg/ml ampicillin for all pGEM-T derived plasmids) and (30 µg/ml kanamycin for all pET28a derived plasmids). The liquid cultures (n=9, biological and technical triplicates) were incubated at either 25°C or 35°C for 16 h. Both OD_{600} and fluorescence at 485 nm were measured using a FLUOstar OPTIMA plate reader (BMG Labtech), every 10 min in order to calculate the GFP synthesis rate per cell per two hours according to the following equation (Davis *et al.*, 2011). The mean values of the GFP synthesis rate per cell of all cultures were subtracted from the GFP synthesis rate per cell of the negative control (*E. coli* BL21 (DE3) cells transformed with pGEM-T vector).

$$\textbf{GFP synthesis rate } x = \textbf{GFP}(x)_{tp2} - \textbf{GFP}(x)_{tp1} \,/\, \textbf{OD}_{600}\,(x)_{\text{average}}$$

2.26. Statistics

All figures and statistical analysis were generated by Sigmaplot software. The type of statistics method was chosen depending on the number of the compared groups. T-test was used to compare between two groups. Whereas, the comparisons among three groups or more were performed using either One- or Two-Way ANOVA tests.

Chapter Three: Caf1R control of the Caf1 polymer secretion by recombinant *Escherichia coli*

3. Chapter Three: Caf1R control of the Caf1 polymer secretion by recombinant *Escherichia coli*

3.1. Introduction

The Gram-negative bacterium *Yersinia pestis* is the causative agent of plague and responsible for approximately 200 million deaths around the world throughout history (Perry and Fetherston, 1997). *Y. pestis* has many virulence factors including lipopolysaccharides (endotoxins) (Carroll, 2015), virulence proteins (YopB and YopD) which are injected into the cytoplasm of the host cells to disrupt the actin cytoskeleton (Slonczewski, 2011) and other effector proteins YopH, YopE, YopJ, YopM, YopO and YopP (Cornelis, 2002). In addition, the Capsular Antigen Fraction 1 (Caf1) is one of the important virulence factors, its apparent function is to avoid being phagocytosed by macrophages.

Caf1 subunits are produced in large quantities to form a polymer which creates a thick gel-like capsule, which is considered to be a distinctive feature of this bacterium (Du *et al.*, 2002), the Caf1 polymer is formed by a group of proteins encoded by the *caf1* operon, which contains four structural genes, *caf1R*, *caf1M* , *caf1A* and *caf1* (Runco *et al.*, 2008). These encode for the putative transcription regulator Caf1R (Karlyshev *et al.*, 1992a), the chaperone Caf1M, the usher Caf1A (Karlyshev *et al.*, 1992c) and the Caf1 monomers which is the building block of the capsule polymer, respectively. The F1 polymer is assembled by chaperone-usher pathway as a non-pilus (atypical) adhesin (Thanassi *et al.*, 1998; Sauer *et al.*, 2004).

In brief, nascent Caf1 subunits are secreted from cytoplasm to the periplasm by the general secretion pathway, Caf1M (chaperones) bind to the Caf1 subunits to partially stabilise the Ig-like fold of the Caf1. The folded Caf1

subunits are then transported to the Caf1A usher in the outer membrane by Caf1M and the chaperone-subunit interaction will then be replaced by subunit-subunit interactions. The exchange process is achieved by a donor strand complementation process, since the chaperone's A1 and G1 β-strands, which cover the subunit, need to be released, allowing the N-terminal β-strand of the next subunit to be interleaved into the polymerization cleft of the growing Caf1 polymer on the periplasmic side of the usher.

The Caf1A (usher) outer membrane protein forms a dedicated secretion channel to assemble the Caf1 subunits into a polymer and secrete the growing polymer on the cell surface (Zavialov *et al.*, 2002). The production of Caf1 is regulated by Caf1R, considered to be a positive transcription factor (Karlyshev *et al.*, 1992a) and the expression is induced in response to temperature at ≈ 37^{o}C (Perry and Fetherston, 1997; Du *et al.*, 2002; Han *et al.*, 2004). However, the evidence for Caf1R's role as a positive transcriptional regulator is limited (Karlyshev *et al.*, 1992a).

In this chapter, a new method was developed using capillary tubes filled with *caf1* expressing culture to measure the flocculant layer amount which is the accumulated capsule material after centrifugation. The amount of flocculent layer is considered as an indication for Caf1 secretion, the higher the amount of flocculent layer obtained, the more Caf1 polymer is produced. Two advantages were attained from the new developed method (using capillary tubes) as compared to the old method (using 5 ml Eppendorf tube), the high accuracy and low culture consumption. In addition, the effect of *caf1* operon genes overexpression on the Caf1 polymer production was investigated for each gene separately.

The *caf1R* overexpression has a negative effect on the Caf1 polymer production, while *caf1M*, *caf1A* and *caf1* overexpression has no significant effect on its secretion. Moreover, the effect of *caf1R* knock-out on the *caf1* expression was examined in the presence and absence of T7 promoter in *E. coli* BL21 (DE3). The *caf1R* gene deletion in the absence of T7 promoter leads to stop the secretion of Caf1 polymer. Whereas, the amount of Caf1 polymer significantly increased when the *caf1R* was deleted in the presence of T7 promoter as an artificial promoter. Caf1R affects the T7 RNA polymerase performance negatively at both low (30^{o}C) and high (35^{o}C) temperature.

The purification of Caf1R was achieved using different host cells including BL21 (DE3), BL21 (DE3) pLYSE, BL21 AI, Tuner (DE3) and C41 (DE3) and different vectors including pBAD Caf1R, pET28A Caf1R^{His}, pTolT Caf1R (the derivative of pET8c which contains TolA protein with N-terminal 6 Histidine tag as fusion partner attached to Caf1R N-terminus) (Anderluh *et al.*, 2003) and pTolT Caf1RCO (contains TolA protein as fusion partner and codon optimised Caf1R) in order to produce a pure Caf1R for biophysical examinations, but at the end no pure Caf1R was obtained.

3.2. Results

3.2.1. Developing a new method for flocculent layer measurement

The main indicator of Caf1 polymer production is the presence of a flocculent layer after centrifugation of the bacterial culture. The centrifugation process leads to the accumulation of capsular materials from the cells to form the fluffy flocculent layer above the pellet of intact cells. The old method of flocculent layer measurement in our laboratory used 5 ml Eppendorf tube filled with 2 ml of bacterial culture to be centrifuged at 2367 x g for 15 minutes. The flocculent layer will form and could be measured in mm using a ruler (Roque *et al.*, 2014). In order to improve the method of flocculent layer separation and measurement 50 µl calibrated glass capillary tubes were used instead of 5 ml Eppendorf tubes, filled with 50 µl bacterial culture, closed with Plasticine from one end and centrifuged at 2367 x g for 15 minutes as stated in methodology to again measure the flocculent layer in mm using a ruler.

The differentiation between samples in the old method was not as accurate as in the new one **(Figure 3-1)**, whilst the flocculent layer height measurements in the capillary tubes method were as follows: 32 mm, 31 mm, 30 mm for samples 1, 2 and 3 and 41 mm, 39 mm and 40 mm for samples 4, 5 and 6, respectively **(Figure 3-1 A)**, the samples 1, 2 and 3 are three cultures of *E. coli* BL21 (DE3) transformed with pT7-COP and the samples 4, 5 and 6 are three cultures of BL21 (DE3) transformed with pT7-COPΔR which produce more flocculent layer than pT7-COP cultures as a result of *caf1R* deletion (the reason will be explained later).

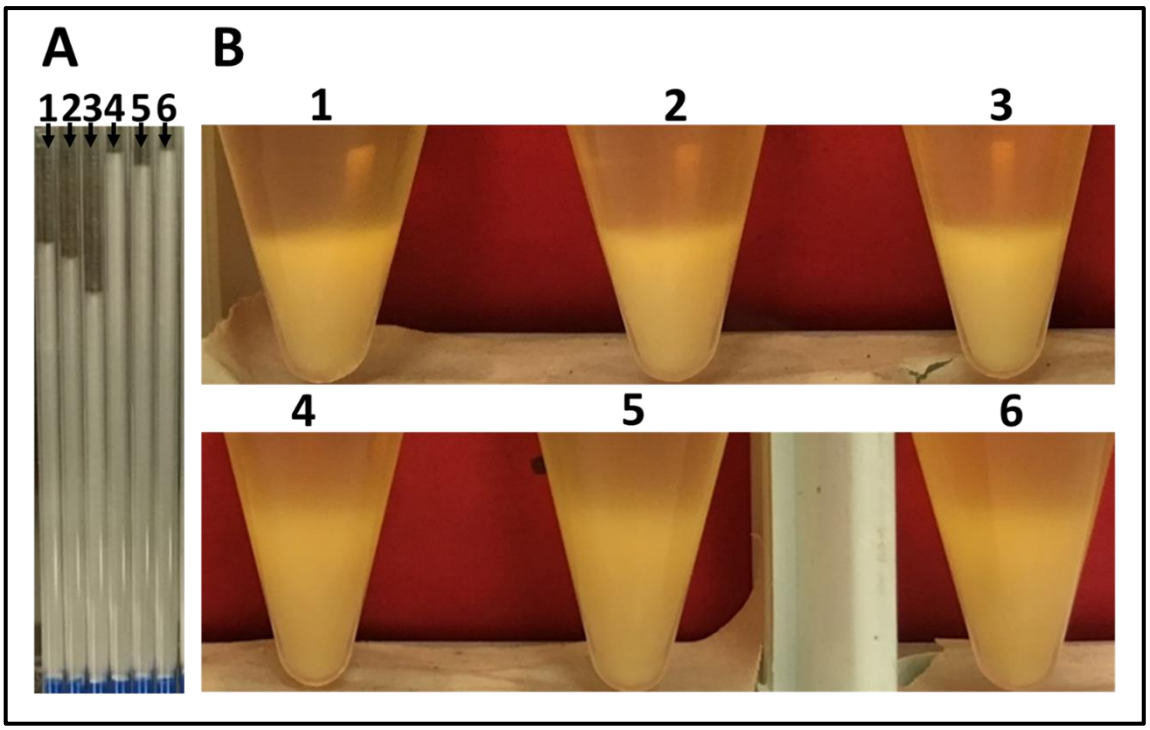

Figure 3-1. The flocculent layer height measurement methods. (A) The new method using capillary tubes filled with F1 capsule producing bacterial culture which are *E. coli* BL21 (DE3) transformed either with pT7-COP in sample 1, 2 and 3 (biological triplicates) or with pT7-COPΔR in sample 4, 5 and 6 (biological triplicate), grown at 35°C for 16 hours, centrifuged at 2367 x g, 22°C for 15 minutes showing the flocculent layer separation. (B) The old method using 5ml Eppendorff tubes filled with the same bacterial cultures as in (A) in the same order, centrifuged at 2367 x g, 22°C for 15 min showing the separation of flocculent layer

The measurement of flocculent layer height in the old method were the same, 11 mm for sample 1,2 and 3 and 14 mm for sample 4, 5 and 6 **(Figure 3-1 B)**. There was no clear difference between samples 1, 2 and 3 and samples 4, 5 and 6 when the Eppendorf tubes were used in the separation. In many cases the comparison between 2 samples containing a similar amount of flocculent layer is difficult in the old method especially with high production of Caf1 as in **Figure 3-1**. The main advantages of this developed method are to minimize the bacterial culture volume used and enhance the accuracy of separation and measurement. Therefore, the capillary tubes

method was chosen to be used in further measurement of the flocculent layer heights due to its higher senitivity and ease of use as compared with the old method.

3.2.2. Caf1 content in the flocculent layer

The flocculent layer appearance after the centrifugation of Caf1 producing bacterial culture is thus an indicator of Caf1 polymer production in the culture (Machado Roque, 2013). The Caf1 polymer is the major protein component of this flocculated layer (Miller *et al.*, 1998). This experiment was done to assess the ratio of Caf1 in the flocculated material after bacterial culture centrifugation. *E. coli* BL21 (DE3) cells were transformed with pT7-COP which contains the four structural genes encoding for Caf1, grown at 35^{o}C for 24 h, the flocculent layer heights were measured every 2 h for 6 time points (**Figure 3-2**). Each sample of same volume flocculent layer was taken from each time point, analysed by SDS-PAGE as stated in methodolgy section and then, the intensities of Caf1 bands were evaluated using a gel documentation system (Gel DocTM XR+, Biorad).

The mean values of flocculent layer heights were rising gradually (**Figures 3-2 and 3-3 A)** as follows: 11.6 mm at 14 h, 16.3 mm at 16 h, 20.3 mm at 18 h, 25.6 mm at 20 h, 28 mm at 22 h and 30.6 mm at 24 h, the statistical analysis shows significant differences in the layer heights between either 14 and 16 h, 16 and 18 h, 18 and 20 h and 20 and 22 h, the difference between 22 and 24 h is not significant. On the other hand, the estimation of Caf1 band intensities on SDS-PAGE was done (**Figures 3-3 B and 3-3 C**) using flocculent layer samples mentioned above and the averages were measured in arbitrary units 13.2, 14.9, 15.9, 13.9, 15.8 and 14 x 10^{6}. The statistical analysis

using One-Way ANOVA test shows no significant differences between the intensities of Caf1 bands of these different time points. These results suggest that Caf1 amount is constant per unit volume of flocculent layer even if these flocculated materials are obtained from different time points and have different total volumes.

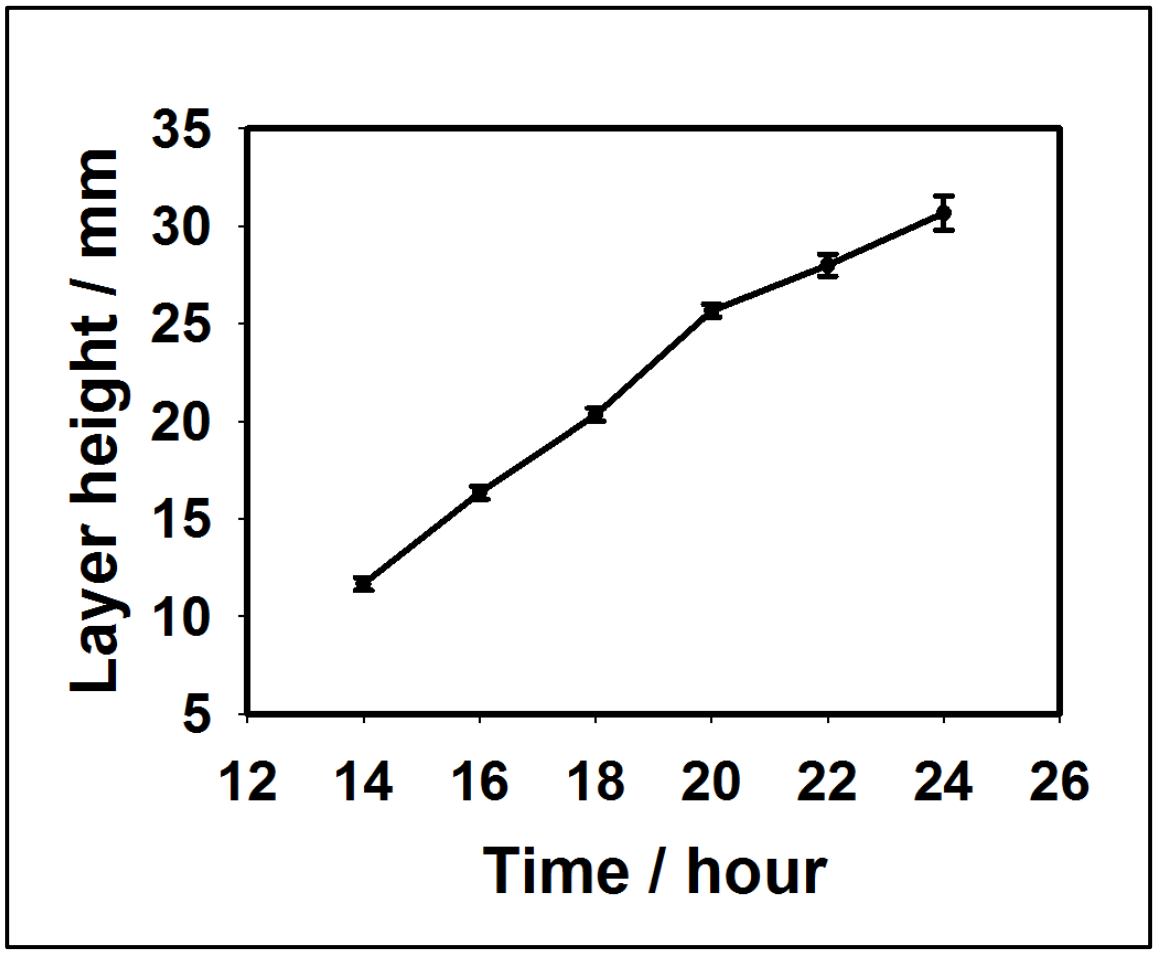

Figure 3-2. Flocculant layer production curve of the cultures of *E. coli* BL21 (DE3) transformed with pT7-COP. Grown at 35°C from different amount of times as stated and centrifuged in capillary tubes to visualise the flocculent layer height in each sample for comparison. Biological replicates = 3. Statistical analysis was performed using One-Way ANOVA test.

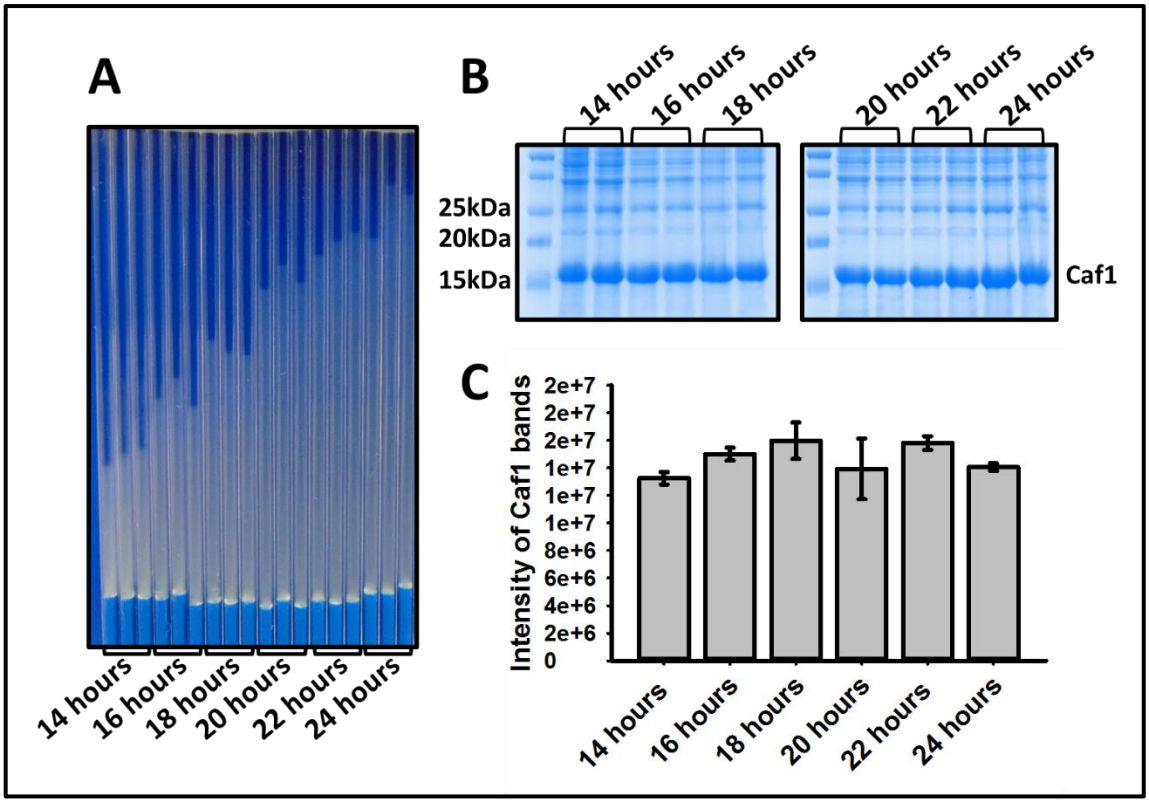

Figure 3-3. Analysis of Caf1 content in the flocculent layer. (A) Image of *E. coli* BL21 (DE3) cultures containing the pT7-COP plasmid grown at 35°C for different amounts of time as stated, and centrifuged in capillary tubes to visualise the flocculent layer height. (B) SDS-PAGE analysis of the flocculent layers of the cultures from (A). (C) Graph of Caf1 band intensities in arbitrary units, obtained by densitometry of the gel shown in (B). Biological replicates = 3. Statistical analysis was performed using One-Way ANOVA test.

3.2.3. The overexpression of *caf1* operon structural genes

As preliminary experiment and to clarify the effect of each *caf1* operon gene on the Caf1 polymer production, the overexpression of genes was achieved using *E. coli* BL21 (DE3) co-transformed with pT7-COP and pBAD Caf1R for *caf1R* overexpression, pT7-COP and pBAD Caf1M for *caf1M* overexpression, pT7-COP and pBAD Caf1A for *caf1A* overexpression and in the case of *caf1* overexpression, pT7-COP and pBAD Caf1 were used. The expression of *caf1* was derived by the leaky expression of T7 RNA polymerase in pT7-COP which is composed of pGEM-T easy vector and *caf1* operon (Machado Roque, 2013), whereas, the overexpression of either *caf1R, caf1M, caf1A or caf1*

was induced by adding L-arabinose when using pBAD vector with L-arabinose inducible promoter. The expression cultures were grown at 35°C for 16 h with different L-arabinose concentrations (0%, 0.05%, 0.1%, 0.2%, 0.4% and 1%) and the flocculent layer production was assayed in each case. Different L-arabinose concentrations were used in the cultures of *E. coli* BL21 (DE3) transformed with pCOP only to evaluate the effect L-arabinose on Caf1 production by the natural *caf1* operon and the mean values of flocculent layer thickness were shown in **Table 3-1**.

Table 3-1. The overexpression of *caf1* operon genes (*caf1R*, *caf1M*, *caf1A* and *caf1*). Using different amounts of L-arabinose for induction. Showing the mean values of the flocculent layer thickness for each culture. ± value is the standard error of the mean.

L-arabinose concentration	The mean values of flocculent layer thickness of the cultures of				
	pCOP only	pCOP + pBADCaf1R	pCOP + pBADCaf1M	pCOP + pBADCaf1A	pCOP + pBADCaf1
0%	17 mm ± 6	12 mm ± 2	13 mm ± 2	11.6 mm ± 1	18 mm ± 1.5
0.05%	18.6 mm ± 7	8.3 mm ± 0.3	10.6 mm ± 1	9.6 mm ± 0.8	20 mm ± 1
0.1%	17.3 mm ± 3	6 mm ± 0.5	8 mm ± 2	10 mm ± 0	18.3 mm ± 0.3
0.2%	21.6 mm ± 3	2.6 mm ± 0.8	11.3 mm ± 4	9.3 mm ± 4	20 mm ± 0.5
0.4%	22.3 mm ± 2	1.3 mm ± 0.3	15.3 mm ± 4.2	14 mm ± 1	24 mm ± 1.5

1%	20 mm ± 4	0 mm ± 0	6.6 mm ± 0.3	7 mm ± 0.5	16.6 mm ± 3.8

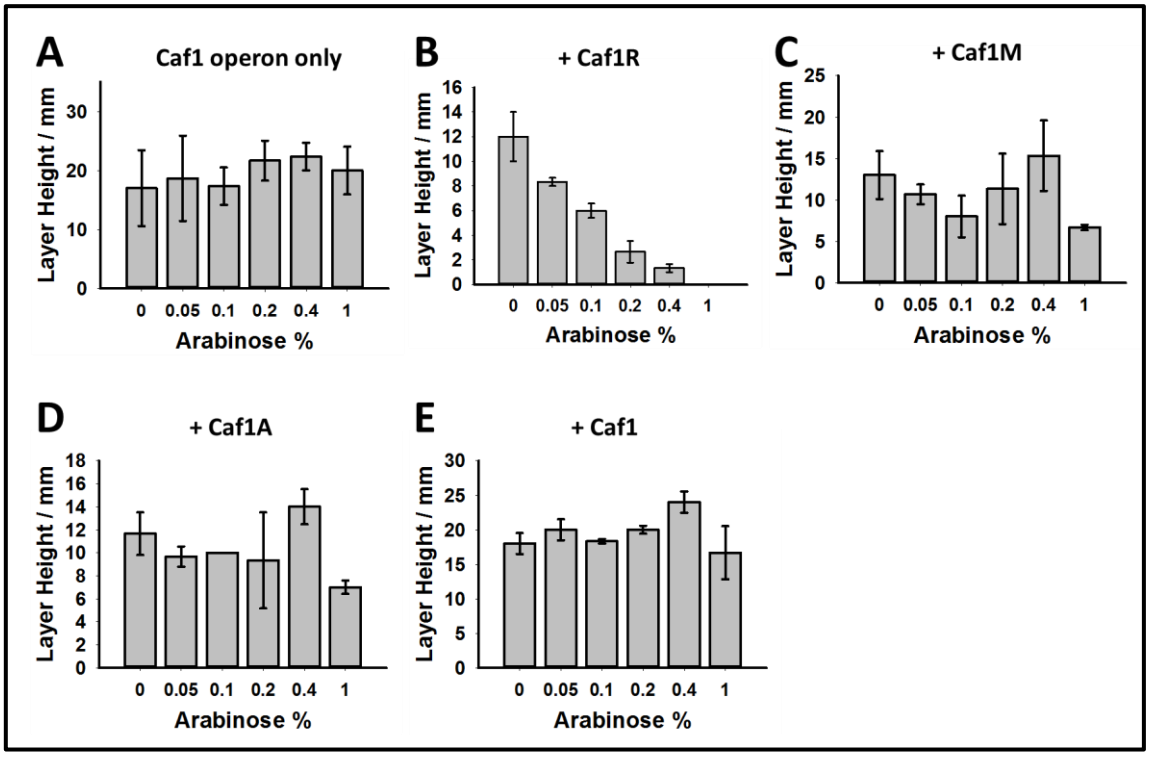

Figure 3-4. *caf1* operon genes overexpression effect on flocculent layer thickness. Graphs of flocculent height obtained with different concentrations of arabinose are shown for cultures of *E. coli* transformed with pCOP (*caf1* operon) only (A), and when co-transformed with arabinose inducible pBad Caf1R (B) Caf1M (C), Caf1A (D), and Caf1 (E). Cultures were grown for 16 hrs at 35°C. Bar heights correspond to mean flocculent layer height obtained from three separate cultures, where error bars represent the standard error of the mean (S.E.M). Biological replicates = 3. Statistical analysis was performed using One-Way ANOVA test.

The statistical analysis using One-Way ANOVA test revealed no significant differences **(Table 3-2)** among the treatments in either negative control which contains the pCOP only, *caf1M*, *caf1A* or *caf1* overexpression cultures (**Figure 3-4 A, C, D and E**). Alternatively, *caf1R* overexpression resulted in a distinct stepwise decline in the Caf1 amount produced (**Figure 3-4 B**), there is significant differences between the different L-arabinose concentrations.

The SDS-PAGE results (**Figure 3-5**) displayed apparent Caf1 bands at ≈ 15.5 kDa in all heated flocculent layer samples due to dissociation of Caf1 polymer to monomers, at the same time, the unheated flocculent layer samples do not show such bands in SDS-PAGE results. Therefore, the overexpression of *caf1R* affected the production Caf1 polymer negatively. Whereas, the overexpression of other *caf1* operon genes (*caf1M*, *caf1A* and *caf1*) did not have a significant effect on the *caf1* expression. For further investigation of *caf1R* overexpression effect, we constructed pBAD Caf1R with a C-terminal FLAG tag to be tested in additional experiments.

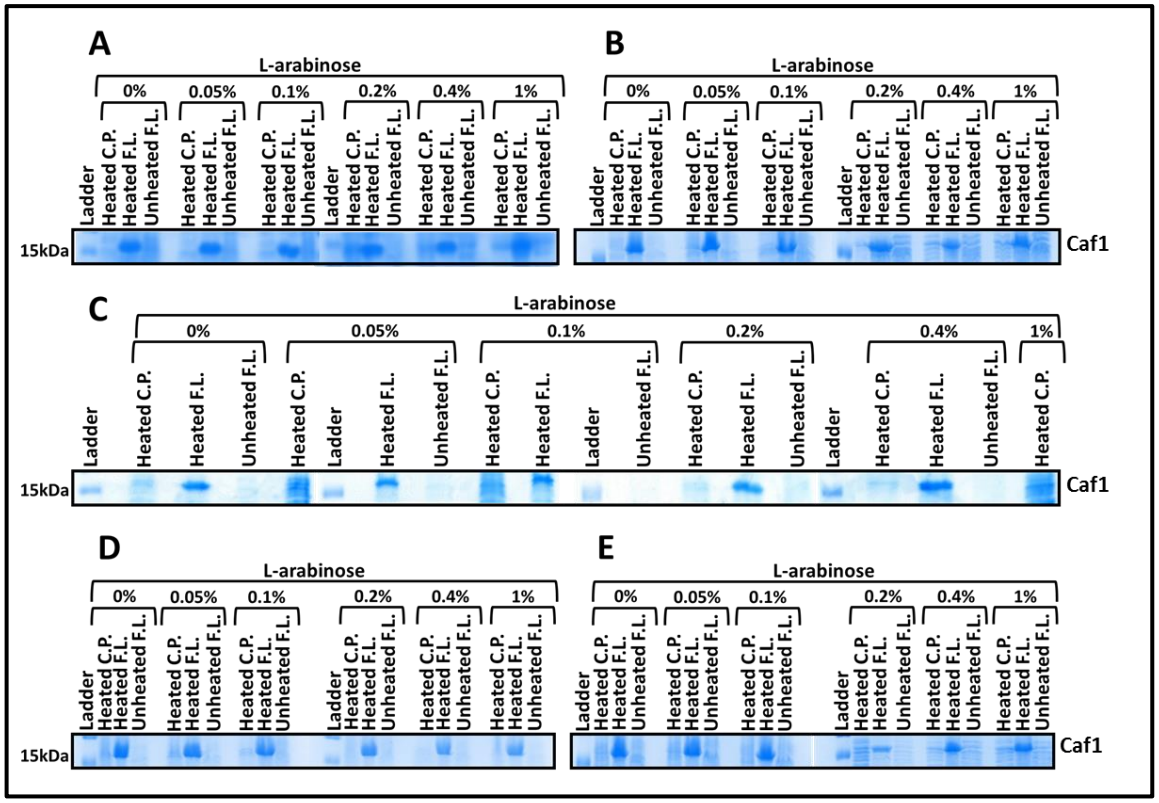

Figure 3-5. SDS-PAGE analysis of cultures of *E. coli* transformed with pCOP (*caf1* operon) only (A), and when co-transformed with arabinose inducible pBad Caf1M (B) Caf1R (C), Caf1A (D), and Caf1 (E). Cultures were grown for 16 hrs at 35°C, the heated C.P. is cell pellet heated at 100°C for 10 minutes, the heated F.L. is the flocculant layer heated at 100°C for 10 minutes and the unheated F.L. is unheated flocculant layer. Caf1 bands (15.5 kDa) are clear in all heated flocculant layer samples. Biological replicates = 3.

Table 3-2. The statistical analysis of the *caf1* operon gene overexpression. Showing the significance of differences represented by P-values among the groups of cultures induced with different concentrations of L-arabinose using One-Way ANOVA test.

L-arabinose concentrations	The overexpressed gene				
	No	*caf1R*	*caf1M*	*caf1A*	*caf1*
	P-values				
0% vs. 0.05%	0.87	0.1	0.49	0.38	0.40
0.05% vs. 0.1%	0.87	0.2	0.39	0.72	0.34
0.1% vs. 0.2%	0.40	0.1	0.53	0.88	0.06
0.2% vs. 0.4%	0.87	0.5	0.54	0.35	0.07
0.4% vs. 1%	0.64	0.3	0.11	0.01	0.15

3.2.4. The effect of *caf1R* overexpression on the Caf1 production

In order to elucidate the effect of *caf1R* overexpression in more details, *E. coli* BL21 (DE3) were either transformed with pT7-COPFΔR (with $Caf1M^{FLAG}$, $Caf1A^{FLAG}$) only, pBAD Caf1RF ($Caf1R^{FLAG}$) only or co-transformed with pT7-COPFΔR and pBAD Caf1RF to be induced with different concentrations of L-arabinose (**Figure 3-6 A**) using 0%, 0.05%, 0.1%, 0.2% and 0.4%. The measurements of the flocculent layer thickness of these samples demonstrated that there is a graduated loss in *caf1* expression from 0% to 0.4% L-arabinose (**Figure 3-6 A**) as shown in the previous experiment. the height averages were: 15.6 mm, 0 mm, 4 mm, 3 mm, 0.83 mm, 0 mm and 0 mm for cultures of *E. coli* BL21 (DE3) transformed with pT7-COPFΔR only, transformed with pBAD Caf1RF only, co-transformed with pT7-COPFΔR and pBAD Caf1RF with 0%, 0.05%, 0.1%, 0.2% and 0.4% L-arabinose in that order.

The statistical analysis revealed that the differences among these treatments are significant. The western blotting results of the cell pellets (**Figure 3-6 B**)

prove the overexpression of *caf1R* in culture of *E. coli* BL21 (DE3) co-transformed with pT7-COPFΔR and pBAD Caf1RF induced with 0.1%, 0.2% and 0.4% L-arabinose, these mentioned samples produced clear bands for Caf1R at ≈ 36 kDa with stepwise increase in their intensity. In contrast, there were no Caf1R bands in the cultures of either *E. coli* BL21 (DE3) transformed with pT7-COPFΔR only, transformed with pBAD Caf1RF only or co-transformed with the both plasmids without L-arabinose induction.

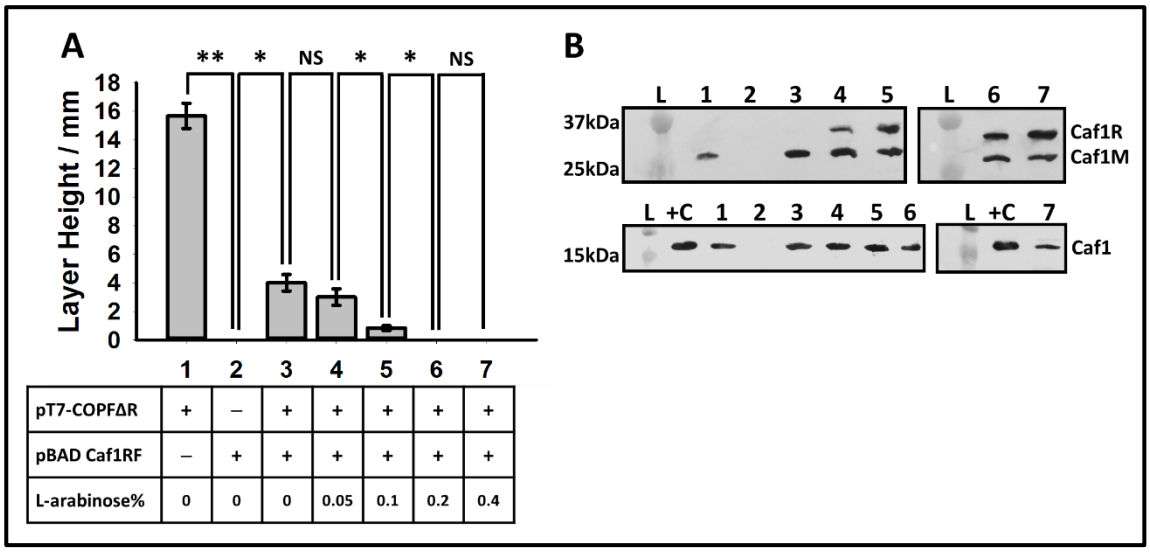

Figure 3-6. The *caf1R* overexpression inhibits flocculent formation. (A) Graphs of flocculent height obtained for cultures of *E. coli* BL21 (DE3) etheir transformed with pCOPFΔR (*caf1* operon, *caf1M*-FLAG and *caf1A*-FLAG) only (1), transformed with arabinose inducible pBad Caf1RF only as a negative control (2) or co-transformed with pCOPFΔR and pBAD Caf1RF. Grown at 35°C for 16 hours with the addition of different amounts of L-arabinose in (3), (4), (5), (6) and (7). Bar heights correspond to mean flocculent layer height obtained from three separate cultures, where error bars represent the standard error of the mean (S.E.M). (B) Western blot of the same cultures mentioned in (A) using anti-FLAG and anti-Caf1 antibodies, shows stepwise increase in Caf1R bands intensity at ≈ 36 kDa in samples (4), (5), (6) and (7), Caf1M bands at ≈ 28 kDa in samples (1), (3), (4), (5), (6) and (7), and Caf1 bands in +C (pure Caf1), (1), (3), (4), (5), (6) and (7). Images formed from *4* separate blots. NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). Biological replicates = 3. Statistical analysis was performed using One-Way ANOVA test.

It seems that the required Caf1R amount for *caf1* operon expression is not detectable by western blotting. On the other hand, the production of Caf1M (chaperone) and Caf1 (subunit) was obvious for all samples **(Figure 3-6 B)**, with the exception of the negative control mentioned above which is the culture of *E. coli* BL21 (DE3) transformed with pBAD Caf1RF only. Therefore, the negative effect of the *caf1R* overexpression was confirmed by this experiment and the bands of Caf1R was shown by western blot results from *caf1R* overexpressing cultures. The inhibitory effect of *caf1R* overexpression on the Caf1 production driven by T7 RNA polymerase was an encouraging finding to examine the role of Caf1R in the Caf1 polymer secretion in the natural system by deletion of *caf1R* gene from *caf1* operon.

3.2.5. *caf1R* is required for *caf1* operon expression in the absence of a T7 promoter

Caf1R has been referred to as a positive transcriptional regulator for *caf1* operon gene expression (Karlyshev *et al.*, 1992b), but there was no clear evidence shown for this function. Therefore, *caf1R* gene was knocked- out from pCOP to create pCOPΔR and examine its effect on the Caf1 polymer production. *E. coli* BL21 (DE3) were transformed with either pCOP, pCOPΔR, co-transformed with both pCOPΔR and pBAD Caf1R with and without L-arabinose induction or transformed with pBAD Caf1R only as a negative control. These were grown at 35^{o}C for 16 h, the flocculent layer thickness was measured and all samples were analysed by western blot using anti-DnaK antibody (loading control) and anti-Caf1 antibody.

The average values of flocculent layer heights for cultures of *E. coli* BL21 (DE3) transformant for either pCOP, pCOPΔR, co-transformant for both

pCOPΔR and pBAD Caf1R with and without L-arabinose induction or transformed with pBAD Caf1R only **(Figure 3-7 A)** were: 3.5 mm, 0 mm, 1.8 mm, 0.8 mm and 0 mm, respectively. Western blotting of the same samples mentioned above **(Figure 3-7 B)** confirmed the presence of Caf1 bands in both pCOP transformants, pCOPΔR and pBAD Caf1R co-transformants with 0% L-arabinose and with 0.005% L-arabinose. On the other hand, transformant for either pCOPΔR or pBAD Caf1R only (negative control) did not produce Caf1 bands, demonstrating that Caf1R is required for Caf1 secretion in the natural operon and the co-transformation of pCOPΔR and pBAD Caf1R rescued the production of Caf1 without induction and based on the leaky expression of *caf1R* from pBAD Caf1R.

The induction using 0.005% produced a decreased amount of flocculent layer related to non-induced culture as a consequence for *caf1R* overexpression which is fitting with previous results of *caf1R* overexpression. In addition, the fold change in expression were measured for all *caf1* operon genes using RT-PCR analysis **(Figure 3-7 C)** for cultures of *E. coli* BL21 (DE3) transformed with either pCOPF or pCOPFΔR relative to β-*lactamase* (as a house keeping gene), the fold-change in expression is equal to (2^-ΔCt), where ΔCt = (Ct value of target gene - Ct value of β-*lactamase* gene) (Livak and Schmittgen, 2001; Dheda *et al.*, 2004).

The means of expression fold change of pCOPF transformants are shown in **Table 3-3**. The statistics using T-test indicated that there are significant differences among the transcriptional levels of all *caf1* operon genes *caf1R*, *caf1M*, *caf1A* and *caf1* between the cultures of pCOPF and pCOPFΔR transformants (**Table 3-4**).

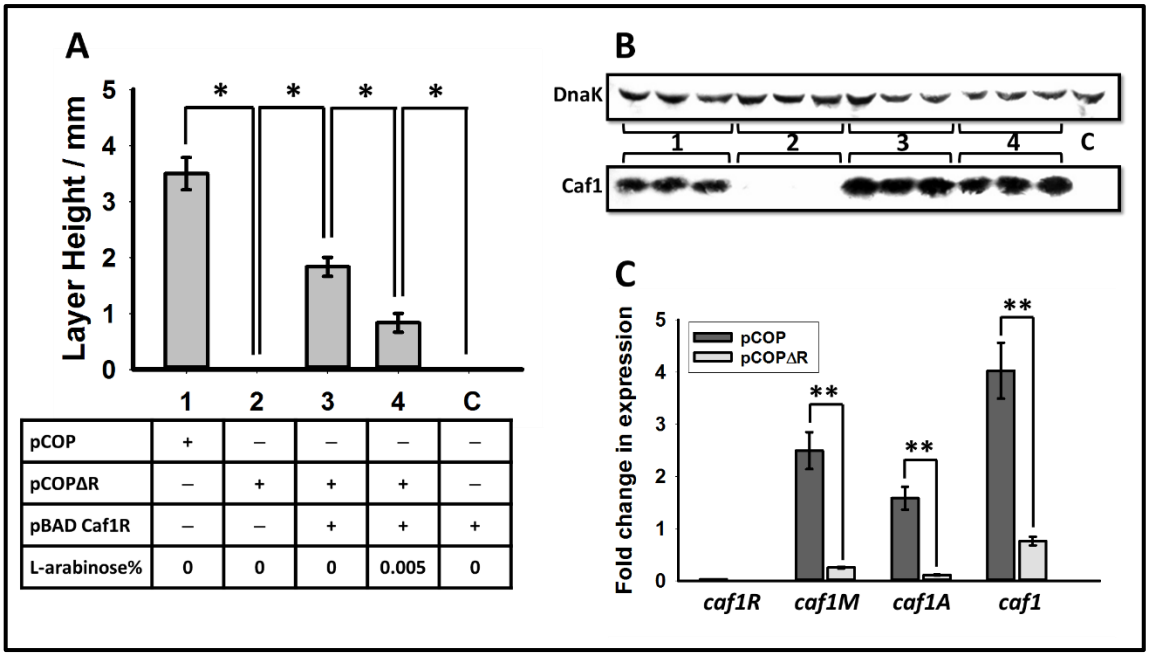

Figure 3-7. Rescue of Caf1 protein production in ΔCaf1R cells. (A) Graph of the flocculent layer height obtained from *E. coli* cultures containing either pCOP (full *caf1* operon, 1), pCOPΔR (*caf1* operon lacking *caf1R*, 2), a co-transformation of pCOPΔR and pBad Caf1R (arabinose inducible *caf1R*) with (3) and without (4) 0.005% w/v arabinose added to the culture, and pBad Caf1R only control (-C). Cultures were grown for 16 h at 35°C. Bar heights correspond to the mean flocculent layer height obtained from three separate cultures where error bars represent the standard error of the mean (S.E.M). (B) Western blot of three separate cell pellets from the above cultures, probed with an anti-Caf1 antibody. DnaK was used as a loading control, probed with an anti-DnaK antibody. (C) Transcript levels, determined by RT-PCR, of each gene are shown for cultures of *E. coli* transformed with pCOPF and pCOPFΔR, grown at 35°C for 16 hours. Three cultures of each transformant were grown, with RT-PCR reactions run in duplicate for each culture. Bar heights correspond to mean fold-change in expression relative to β-*lactamase*. Error bars represent standard error of the mean (S.E.M). NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). Biological replicates = 3. Statistical analysis was performed using T-test.

Therefore, Caf1R is required for *caf1* transcription and consequently its expression and Caf1 polymer secretion outside the bacterial cells. However, the overexpression of *caf1R* resulted in a clear decrease in Caf1 production which confirmed the negative effect of *caf1R* overexpression again. These interesting effects of Caf1R on the production of Caf1 polymer opened the door for further investigation of the *caf1R* deletion in the presence of T7 as an artificial promoter to produce more Caf1 polymer by eliminating such potential inhibition for production optimisation.

Table 3-3. The effect of *caf1R* deletion on the gene transcripts by RT-PCR. Showing the mean values of the expression fold-change of the *caf1* operon gene transcripts. ± value is the standard error of the mean.

Transformant cultures	The mean values of the expression fold-change of the gene transcripts			
	caf1R	***caf1M***	***caf1A***	***caf1***
pCOPF	0.02 ± 0.002	2.49 ± 0.3	1.58 ± 0.2	4.02 ± 0.5
pCOPFΔR	0	0.25 ± 0.01	0.11 ± 0.007	0.76 ± 0.08

Table 3-4. The statistical analysis of the *caf1R* deletion effect on the *caf1* operon gene transcripts by RT-PCR. Showing the P-values calculated for comparisons made between pCOPF and pCOPFΔR cultures using T-test.

Gene transcript	The P-values of the comparison between pCOPF vs. pCOPFΔR cultures
caf1R	<0.001
caf1M	<0.001
caf1A	<0.001
caf1	<0.001

3.2.6. A *caf1R* knock-out increases the *caf1* operon expression driven by T7 RNA polymerase

Due to the inhibition effect of Caf1R on T7 transcription system which was seen in the *caf1R* overexpression experiment, an additional investigation was needed to examine the effect of Caf1R on the Caf1 polymer production driven by T7 RNA polymerase. The plasmid pT7-COPΔR was constructed using In-Fusion HD cloning technique as stated in methods by deleting *caf1R* from *caf1* operon from pT7-COP plasmid. *E. coli* BL21 (DE3) were transformed with either pT7-COP or pT7-COPΔR, 3 single colonies from each culture were grown separately for 16 h at 35^{o}C, centrifuged in capillary tubes to measure the flocculent layer height for each sample and these samples were analysed in western blotting using the anti-Caf1 antibody.

The levels of *caf1R*, *caf1M*, *caf1A* and *caf1* transcripts were assessed by RT-PCR relative to β-*lactamase* (as a house keeping gene) using specific primers for each gene as stated. The averages of thicknesses of flocculent layer produced from pT7-COP and pT7-COPΔR were: 8.3 mm and 18 mm, respectively **(Figure 3-8 A)** and the statistical analysis proved there is a significant difference between them with P-value (0.0001). The western blotting results **(figure 3-8 C)** using anti-Caf1 antibody showed the presence of Caf1 bands in both cultures. The mean values of transcriptional levels (fold change in expression) of *caf1* operon genes obtained by RT-PCR are shown in **Table 3-5.**

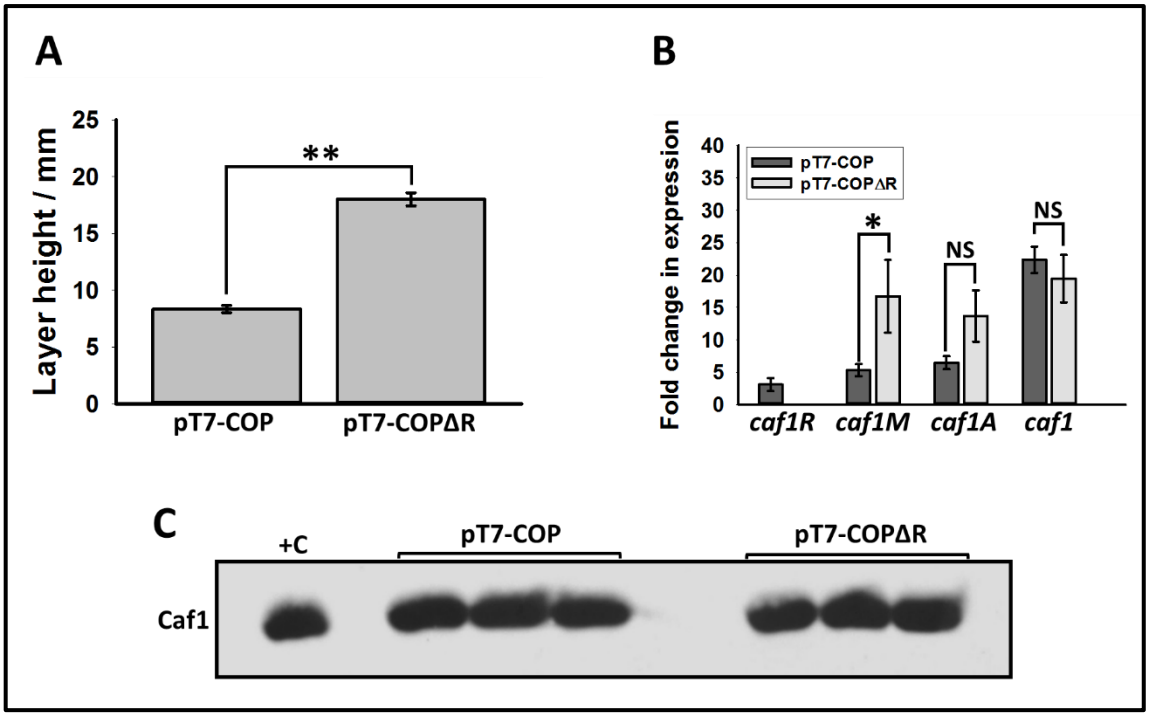

Figure 3-8. The effect of *caf1R* knock-out on the *caf1* operon expression. (A) Graph of flocculent height obtained for cultures of *E. coli* BL21 (DE3) transformed with either pT7-COP or pT7-COPΔR. Bar heights correspond to mean flocculent layer height obtained from three separate cultures, where error bars represent the standard error of the mean (S.E.M). (B) Transcript levels, determined by RT-PCR, of each *caf1* operon genes are shown for cultures of *E. coli* transformed with pT7-COPF (full *caf1* operon) and pT7-COPFΔR (*caf1* operon, *caf1R* deleted), grown at 35°C for 16 hours. Three cultures of each transformant were grown, with RT-PCR reactions run in duplicate for each culture. Bar heights correspond to mean fold-change in expression relative to β-*lactamase* gene. Error bars represent standard error of the mean (S.E.M). (C) Western blot showing Caf1 bands in both cultures (detected using anti-Caf1 antibodies). +C is pure Caf1. NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). Biological replicates = 3. Statistical analysis was performed using T-test.

Table 3-5. The effect of *caf1R* deletion on the levels of the *caf1* operon gene transcripts in the presence of T7 promoter by RT-PCR. Showing the mean values of the expression fold-change of the *caf1* operon gene transcripts. ± value is the standard error of the mean.

Transformant cultures	The mean values of the expression fold-change of the gene transcripts			
	caf1R	***caf1M***	***caf1A***	***caf1***
pT7-COPF	3.1 ± 1	5.3 ± 0.9	6.5 ± 0.9	22.3 ± 2
pT7-COPFΔR	0	16.7 ± 5	13.6 ± 3	19.4 ± 3

The statistics using T-test revealed a significant difference between *caf1M* transcript for pT7-COP and pT7-COPΔR (p-value 0.009), despite *caf1A* being much lower, the difference in *caf1A* and *caf1* levels were not significant. This means that the presence of Caf1R might inhibit the transcription of *caf1M* driven by T7 RNA polymerase rather than *caf1A* and *caf1* genes. Consequently, *caf1R* deletion led to a significant increase in the production of Caf1 polymer driven by T7 RNA polymerase which indicates that Caf1R binding to DNA affects the performance of T7 RNA polymerase by hindering it from moving forward during transcription process. One potential explanation of increased Caf1 production by *caf1R* knock-out, is the length of DNA between T7 promoter and *caf1* operon, in other words the deletion of 900 nt (*caf1R* gene length) could ease the transcription of other *caf1* operon genes by T7 RNA polymerase. For that reason, the start codon was substituted by a random sequence and three stop codons were inserted instead of complete knock-out of *caf1R*. This construct was further examined for Caf1 production.

3.2.7. The lack of Caf1R is responsible for increased Caf1 expression by T7 system

In order to investigate the actual reason for such expansion in the flocculent layer production, pT7-COPΔRSC was constructed by In-Fusion HD cloning. This contains the *caf1* operon with the *caf1R* start codon substituted by a random 3 nt and three stop codons inserted at the beginning of Caf1R N-terminus rather than complete deletion of *caf1R*. *E. coli* BL21 (DE3) were transformed either with pT7-COP, pT7-COPΔR or pT7-COPΔRSC, grown for 16 h at 35^{o}C, the cultures were centrifuged in capillary tubes to measure the thickness of flocculent layer for each culture and then all samples were analysed by SDS-PAGE.

The *caf1R* start codon substitution with a random three nt in pT7-COPΔRSC results in increase in Caf1 production as in pT7-COPΔR construct, since the means of flocculent layer thickness from cultures of *E. coli* BL21 (DE3) transformed with either pT7-COP, pT7-COPΔR or pT7-COPΔRSC were: 31.3 mm, 41.3 mm and 42.3 mm, respectively **(Figure 3-9 A)**. The statistics using One-Way ANOVA test revealed significant differences between pT7-COP and pT7-COPΔR cultures as expected with p-value (0.002) and between pT7-COP pT7-COPΔRSC cultures with p-value (0.0005), while, there was no significant difference between pT7-COPΔR and pT7-COPΔRSC cultures, the p-value (0.5).

The SDS-PAGE results **(Figure 3-9 B)** presented that there are obvious Caf1 bands at 15.5 kDa in heated flocculent layer samples (at 100^{o}C for 10 minutes) taken from all three cultures and no Caf1 band in unheated samples, by reason of the monomerization of Caf1 polymer after heating.

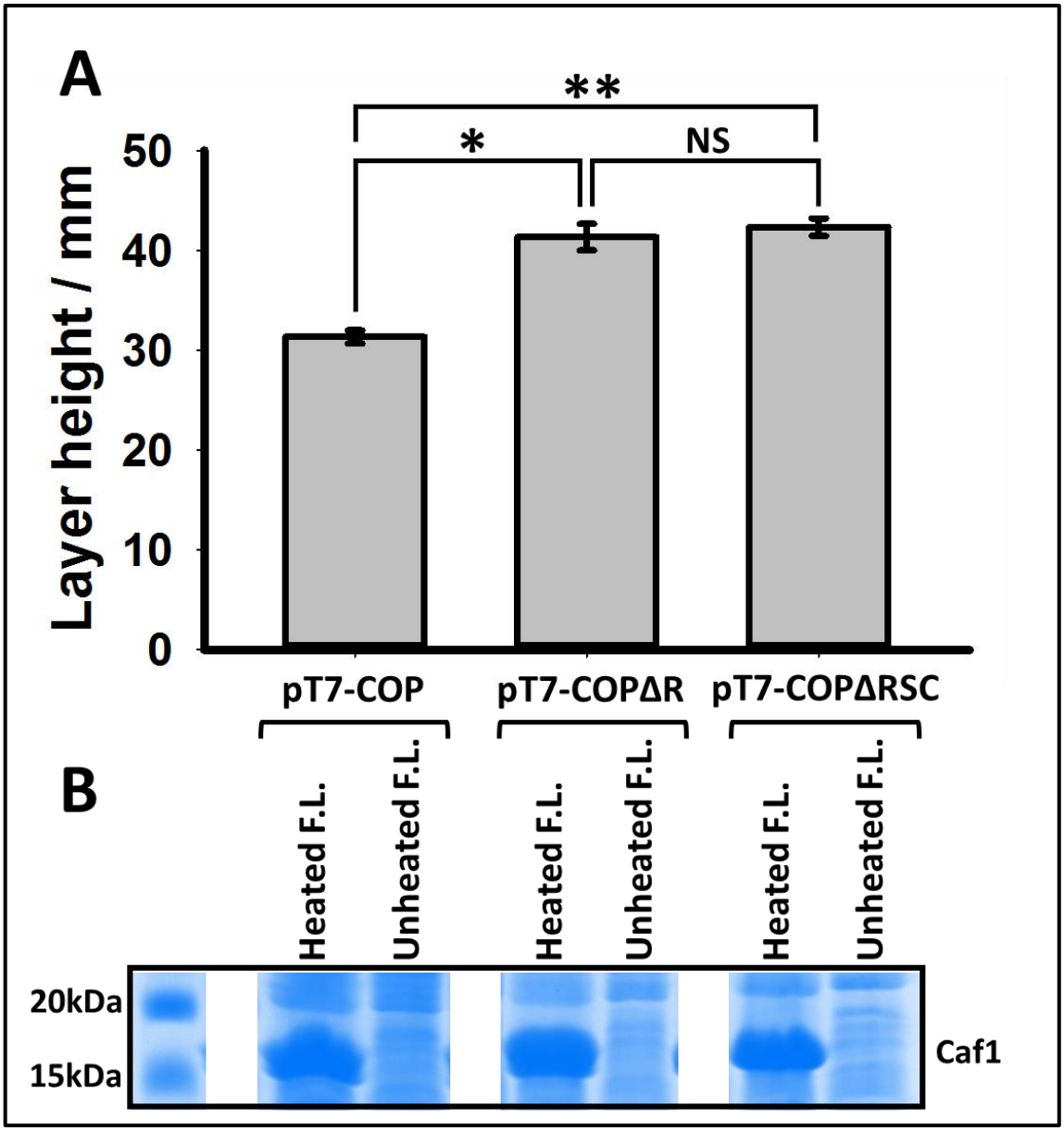

Figure 3-9. The effect of *caf1R* start codon substitution on Caf1 production. (A) Graph of flocculent height obtained for cultures of *E. coli* BL21 (DE3) transformed with either pT7-COP, pT7-COPΔR or pT7-COPΔRSC, all cultures grown at 35°C for 16 hours. Bar heights correspond to mean flocculent layer height obtained from three separate cultures from each transformant, where error bars represent the standard error of the mean (S.E.M). (B) SDS-PAGE of the same cultures in (A) showing two samples from each culture, heated flocculent (at 100 °C for 10 minutes) and unheated flocculant layer, the Caf1 bands are obvious at 15.5 kDa in all heated samples as a result of Caf1 monomerization, while all unheated samples had no monomeric Caf1 bands. NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). Biological replicates = 3. Statistical analysis was performed using One-Way ANOVA test.

Therefore, the rise in Caf1 production driven by T7 RNA polymerase in pT7-COPΔR was not due to the curtailing in the DNA length between the T7 promoter and *caf1* operon genes, but more likely the *caf1R* gene knock-out. Consequently, we propose that the presence of *caf1R* inhibits the function of the T7 system and this inhibition effect could be owing to Caf1R binding to the intergenic region between *caf1R* and *caf1M* genes obstructing the T7 RNA polymerase.

This phenomeneon was seen at 35^{o}C at which the Caf1R protein is in an active form to trigger the transcription of *caf1* operon genes in response to such high temperature (Perry and Fetherston, 1997; Han *et al.*, 2004; Motin *et al.*, 2004). The low temperature at which the Caf1R induced expression is assumed to be inoperative was needed to be tested to examine the effect of Caf1R on the T7 RNA polymerase function at low temperature.

3.2.8. Caf1R reduces the T7 RNA polymerase performance at low and high temperatures

Although, Caf1R, the positive transcriptional regulator is responsible for upregulation of *caf1* operon genes at high temperature close to 37^{o}C (Motin *et al.*, 2004), the levels of *caf1* operon gene products (Caf1M, Caf1A and Caf1) are much less at low tempreture (26^{o}C) (Han *et al.*, 2004), in other words, Caf1R is likely to be a thermoresponsive transcription factor. So the main aim of this experiment is to assess the inhibition effect of Caf1R on T7 system at two different temperatures low 30^{o}C and high 35^{o}C, 30^{o}C was chosen as a low temperature to produce a considerable amount of Caf1 from the leaky expression of T7 RNA polymerase.

E. coli BL21 (DE3) were transformed with either pCOP (contains *caf1* operon genes without T7 promoter), pT7-COP (contains T7 promoter and *caf1* operon genes) or pT7-COPΔR (contains T7 promoter and *caf1* operon genes lacking *caf1R*), each culture was grown at two temperatures 30^{o}C and 35^{o}C for 16 h, the height of flocculent layer was meassured after the centrifugation of mentioned cultures in capillary tubes **(Figure 3-10 A)**. The mean values of flocculent layer thickness are shown in **Table 3-6**. There was no Caf1 production by pCOP transformant cultures at 30^{o}C, small amounts of flocculent layer produced by pCOP cultures at 35^{o}C based on the expression by natural *caf1* operon. Although both pT7-COP and pT7-COPΔR transformants cultures produced considerable amounts of flocculent layer at 30^{o}C, the amounts of flocculent layer were higher at 35^{o}C. The statistics were performed using One-Way ANOVA test as shown in **Table 3-7**.

Additionally, all stated cultures were analysed by SDS-PAGE **(Figure 3-10 B)**, sample (1) which is *E. coli* BL21 (DE3) transformed with pCOP grown at 30^{o}C did not show a Caf1 band from the cell pellet, whereas, all other samples produced obvious Caf1 bands at 15.5 kDa. These results indicates that Caf1 production is temperature-sensitive in pCOP transformant culture, Caf1 is secreted at 35^{o}C but not at 30^{o}C based on natural promoter.

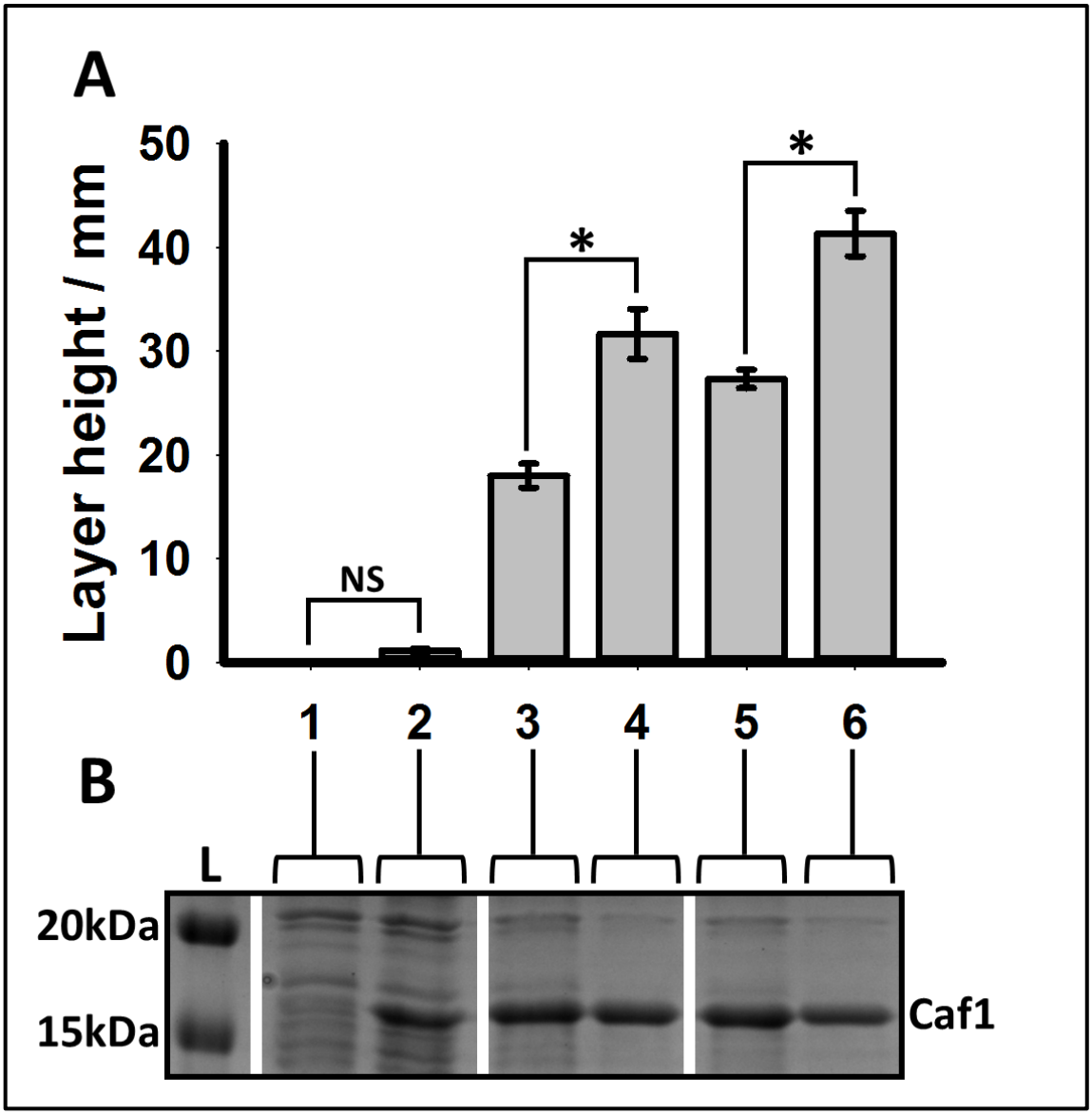

Figure 3-10. *caf1* expression at low and high temperatures. (A) Graph of flocculent height obtained for cultures of *E. coli* BL21 (DE3) transformed with either pCOP (1) grown at 30°C and (2) at 35°C, pT7-COP (3) grown at 30°C and (4) at 35°C or pT7-COPΔR (5) grown at 30°C and (6) at 35°C all for 16 hours. Bar heights correspond to mean flocculent layer height obtained from three separate cultures from each individual sample, where error bars represent the standard error of the mean (S.E.M). (B) SDS-PAGE analysis of heated flocculant layer (at 100 °C for 10 minutes) of the same samples mentioned in (A), showing Caf1 bands in all cultures except sample (1) which is the natural operon at low temperature did not produce Caf1 band. NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). Biological replicates = 3. Statistical analysis was performed using One-Way ANOVA test.

Table 3-6. Reduction of T7 RNA polymerase performance by Caf1R at low and high temperature. Showing the mean values of flocculent layer thickness of pCOP, pT7-COP and pT7-COPΔR cultures. ± value is the standard error of the mean.

Incubation temperature	The mean values of flocculent layer thickness produced by the cultures of		
	pCOP	pT7-COP	pT7-COPΔR
30°C	0 mm	18 mm ± 1	27 mm ± 0.8
35°C	1 mm ± 0.1	31 mm ± 2	41 mm ± 2

Table 3-7. The statistical analysis of the T7 RNA polymerase performance reduction by Caf1R at high and low temperatures. Showing the P-values calculated for the comparisons among the mean values of flocculent layer thickness produced by different cultures as in the table using One-Way ANOVA test.

The P-values of comparisons among different culture groups		
pCOP at 30°C vs. pCOP at 35°C	**pT7-COP at 30°C vs. pT7-COP at 35°C**	**pT7-COPΔR at 30°C vs. pT7-COPΔR at 35°C**
0.5	<0.001	<0.001

However, the quantity of flocculent layer at both low and high temperature increased when *caf1R* was deleted in the presence of the T7 promoter, implying that Caf1R binds DNA at both temperatures to reduce the transcription achieved by T7 RNA polymerase. All these interesting results were an encouraging to do further studies including molecular biology and biophysics technique to examine Caf1R function in more details, but most of these techniques need pure Caf1R, hence, Caf1R purification was the next step for our study.

3.2.9. Caf1R purification

In order to further study the behaviour and structure of Caf1R *in vitro*, the purification of Caf1R was a crucial achievement. pBAD Caf1R, pET28A Caf1R^{His}, pTolT Caf1R (the derivative of pET8c which contains the TolAIII protein with N-terminal 6 Histidine tag as fusion partner attached to Caf1R N-terminus) (Anderluh *et al.*, 2003) and pTolT Caf1RCO (contains the TolAIII protein as fusion partner and codon optimised Caf1R) were cloned using the In-Fusion HD cloning technique. Different strains of *E. coli* were used for Caf1R purification BL21 (DE3), BL21 (DE3) pLYSE, BL21 AI, Tuner (DE3) and C41 (DE3) cells. The results showed that there was no obvious overexpression of Caf1R, even when the TolAIII protein was used as fusion partner after codon optimisation using different amounts of inducer in each experiment, different temperatures, different amounts of time after induction as shown in **Table 3-8**.

Table 3-8. The conditions of *caf1R* overexpression performed for the purification process.

***E. coli* strain**	**Plasmid**	**Inducer and its concentration**	**Incubation temperatures**	**Incubation time after induction**
BL21 (DE3)	pET28A Caf1R^{His}	1 mM IPTG	16^{o}C	Overnight
BL21 (DE3) pLYSE	pET28A Caf1R^{His}	1 mM IPTG	16^{o}C	Overnight
BL21 AI	pET28A Caf1R^{His}	2% L-arabinose	37^{o}C	1 h, 2 h, 3 h, 4 h

BL21 AI	pBAD Caf1R	2% L-arabinose	37°C	1 h, 2 h, 3 h, 4 h
BL21 AI	pTolT Caf1R	2% L-arabinose + 1 mM IPTG	25°C, 37°C	1 h, 2 h, 3 h, 4 h, overnight
BL21 AI	pET28A $Caf1R^{His}$	0.2 L-arabinose + 1 mM IPTG	25°C, 30°C, 37°C	3 h
BL21 AI	pTolT Caf1R	0.2 L-arabinose + 1 mM IPTG	22°C	overnight
BL21 AI	pTolT Caf1RCO	0.2 L-arabinose + 1 mM IPTG	37°C	3 h
Tuner (DE3)	pET28A $Caf1R^{His}$	1 µM IPTG	16°C , 20°C , 25°C	Overnight
Tuner (DE3)	pET28A $Caf1R^{His}$	10 µM IPTG	16°C , 20°C , 25°C	Overnight
Tuner (DE3)	pET28A $Caf1R^{His}$	50 µM IPTG	16°C , 20°C , 25°C	Overnight
Tuner (DE3)	pET28A $Caf1R^{His}$	100 1 µM IPTG	16°C , 20°C , 25°C	Overnight
Tuner (DE3)	pET28A $Caf1R^{His}$	1 mM IPTG	16°C , 20°C , 25°C	Overnight
Tuner (DE3)	pET28A $Caf1R^{His}$	2 mM IPTG	16°C , 20°C , 25°C	Overnight
Tuner (DE3)	pTolT Caf1R	1 µM IPTG	16°C , 20°C , 25°C	Overnight
Tuner (DE3)	pTolT Caf1R	10 µM IPTG	16°C , 20°C , 25°C	Overnight
Tuner (DE3)	pTolT Caf1R	100 µM IPTG	16°C , 20°C , 25°C	Overnight
C41 (DE3)	pET28A $Caf1R^{His}$	1 mM IPTG	37°C	Overnight

C41 (DE3)	pTolT Caf1R	1 mM IPTG	37°C	Overnight
C41 (DE3)	pTolT Caf1RCO	1 mM IPTG	37°C	3 h

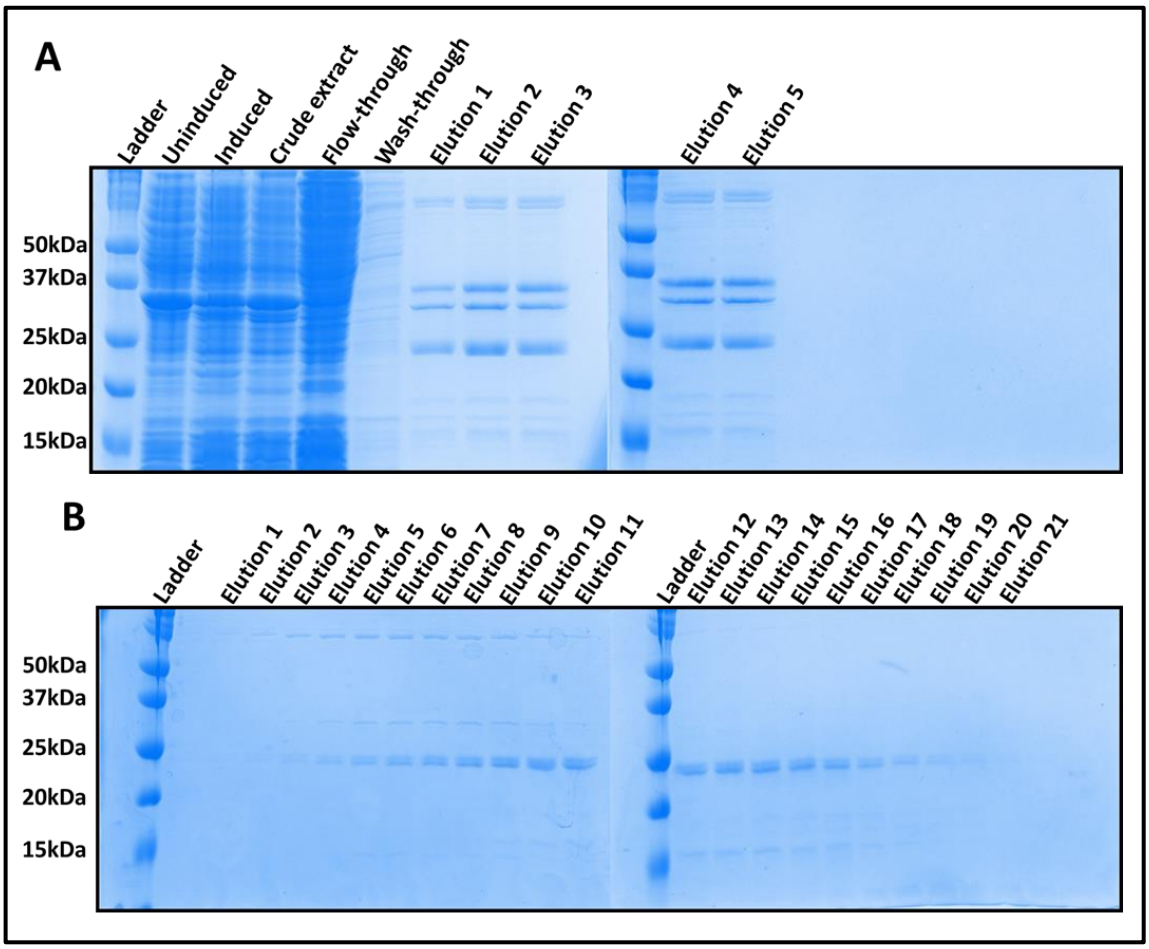

Figure 3-11. SDS-PAGE analysis of Caf1R purification process. Extracted from culture of *E. coli* Tuner (DE3) transformed with pET28A Caf1R^{His}, the starter culture was grown at 37°C overnight, the expression culture was inoculated and grown at 20°C overnight. (A) Lanes 1: protein ladder, 2: the culture before induction, 3: induced culture, 4: crude protein extract before loading on Ni Separose column (1ml HisTrap HP from GE Healthcare), 5: flow-through, 6: wash-through and 7, 8, 9, 10 and 11 are samples eluted from Ni Separose column, respectively. (B) Elution fractions collected after size exclusion chromatography using (Proteo SEC 3-70 kDa HR), showing no bands at 38 kDa (Caf1R CT-His molecular weight).

Nevertheless, several transformants were elected to be used in Caf1R purification based on SDS-PAGE results **(Figure 3-11)**. Many attempts of Caf1R purification experiments were performed, three examples will be mentioned here, *E. coli* Tuner (DE3) cells transformed with pET28A Caf1R^{His} were used to overexpress *caf1R*, the starter culture were grown at 37^{o}C overnight using 50 ml TB medium (30 µg/ml) kanamycin. Four flasks of 750 ml TB (30 µg/ml) kanamycin were innoculated from starter culture as expression culture, grown to OD_{600} (0.6-0.8), then induced with 10 µM IPTG and then incubated at 20^{o}C overnight. Crude proteins were extracted as stated in methodology section, the protein solution loaded on Ni Sepharose High Performance column (1 ml HisTrap HP from GE Healthcare), the column was washed with 10 column volumes wash buffer and then the bound proteins eluted with 5 column volumes elution buffer containing 250 mM imidazol. All fractions collected from flow-through, wash-through and elution were analysed by SDS-PAGE **(Figure 3-11 A)**.

All eluted fractions which produced bands at ≈ 38 kDa (the molecular weight of Caf1R CT-His as measured in ProtParam software (Gasteiger *et al.*, 2005)) were collected and mixed to be dialysed at 4^{o}C overnight, the concentration of protein was (4.29 mg/ml) after dialysis and concentration. Size exclusion chromatography was performed as a second purification step using Proteo SEC 3-70 kDa HR (high performance rigid resin derived from a copolymer of Dextran and Agarose), the eluted samples were analysed in SDS-PAGE **(Figure 3-11 B)**, SDS-PAGE results showed that the potential Caf1R band at ≈ 38 kDa disappeared after dialysis and size exclusion chromatography and no Caf1R was obtained.

Furthermore, *E. coli* BL21 AI cells were transformed with pTolT Caf1R (the derivative of pET8c which contains TolAIII protein with N-terminal 6 Histidine tag as fusion partner attached to Caf1R N-terminus), the expression culture was grown at 37^{o}C until OD600 (0.6-0.8) to be induced then by adding 0.2% L-arabinose to induce the expression of T7 RNA polymerase in BL21 AI cells and 1mM IPTG to induce the expression from pTolT Caf1R (T7-based expression vector), the crude protein extract was loaded on Ni Sepharose column (1 ml HisTrap HP from GE Healthcare) and both wash-through and elution-through fractions were collected to be analysed in SDS-PAGE **(Figure 3-12)**. The N-terminal His-tagged TolA-Caf1R protein parameters were calculated by ProtParam software (Gasteiger *et al.*, 2005), the expected molecular weight is ≈ 47.8 kDa.

However, the results from SDS-PAGE indicate that there is an obvious band at ≈ 48 kDa in the sample taken from induced culture which is not in found in the uninduced sample. This band is not visible in elution fractions after His-tag purification step using Ni Sepharose cloumn (1 ml HisTrap HP from GE Healthcare), indicating that the protein has degraded or did not bind to the column. The analysis of these bands by mass spectrometry PMF (peptide mass fingerprinting) technique shows that the massive band at the top of elution fractions lanes is DnaK protein (70 kDa molecular chaperone), the second band is groEL protein (60 kDa chaperonine) from *E. coli* and the third one is a combination of TolAIII and Caf1R (the target band) which was faint in elution fractions at ≈ 48 kDa **(Figure 3-12)**, thus the TolAIII-Caf1R protein was a tiny portion of elution fractions as compared with non-specific proteins bound to Ni Sepharose column (1 ml HisTrap HP from GE Healthcare).

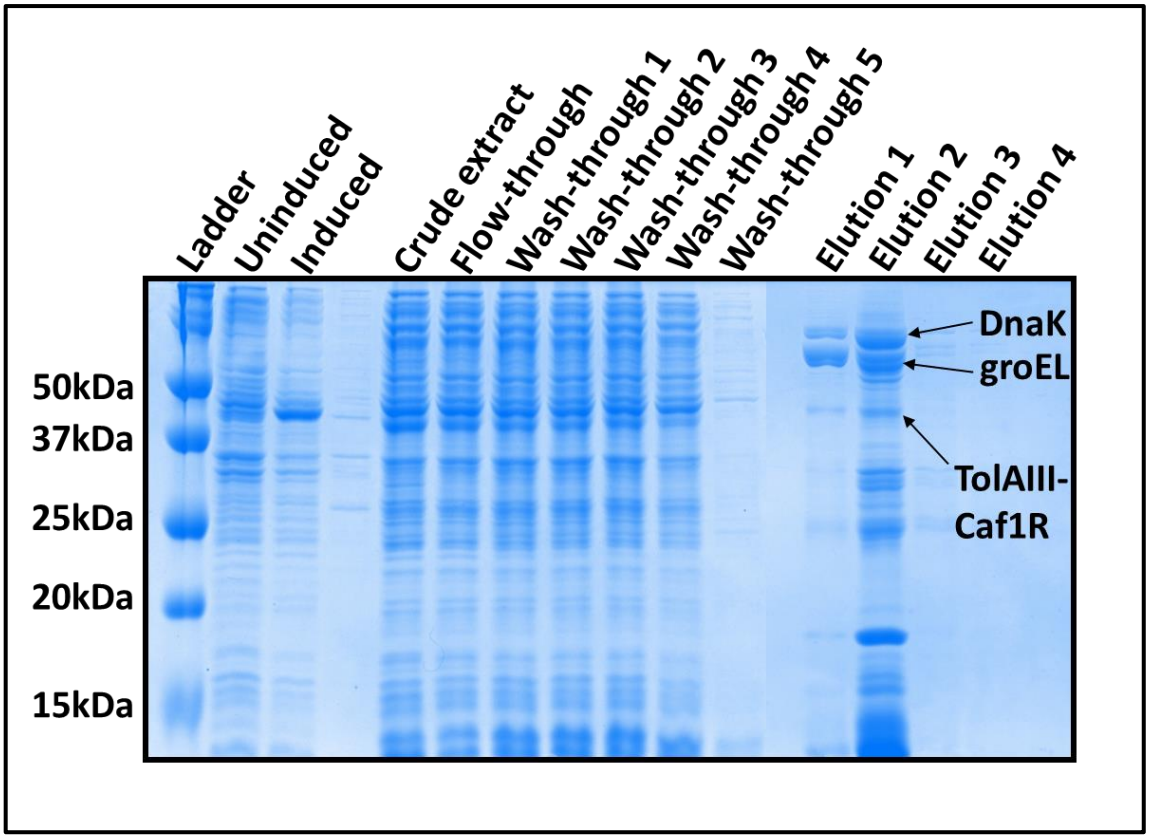

Figure 3-12. SDS-PAGE analysis of Caf1R purification process. Extracted from culture of *E. coli* BL21 AI transformed with pTolT Caf1R, the starter culture was grown at 37°C overnight, the expression culture was inoculated and grown at 25°C overnight. (A) Lanes 1: protein ladder, 2: the culture before induction, 3: induced culture with 0.2% L-arabinose showing a potential TolA-Caf1R band at ≈ 48 kDa, 4: crude protein extract before loading on Ni Sepharose column, 5: flow-through, 6, 7, 8, 9 and 10: wash-through samples 1, 2, 3, 4 and 5 respectively. Lanes 11, 12, 13 and 14 are samples eluted from Ni Sepharose column. As detected by PMF (peptide mass fingerprinting), the DnaK band at ≈ 70 kDa, groEL at ≈ 60 kDa and TolAIII-Caf1R at ≈ 48 kDa.

This low level of expression may be a result of Caf1R toxicity to the *E. coli* cells, hence *E. coli* C41 (DE3) cells were selected due to their ability to provide a reliable overexpression of toxic proteins (Miroux and Walker, 1996; Dumon-Seignovert *et al.*, 2004), the C41 (DE3) cells were transformed with pTolT Caf1RCO (the derivative of pET8c which contains TolA protein with N-terminal 6 Histidine tag as fusion partner attached to codon-optimised

Caf1R), grown as starter culture at 37^{o}C overnight, the expression cultures were grown until OD_{600} (0.6-0.8) and then induced with 1 mM IPTG, the induced cultures were re-incubated at 37^{o}C for 3 h. The bacterial cells were broken with high pressure (23Kpsi) using a Constant Systems cell disrupter, loaded on Ni Sepharose column (1 ml HisTrap HP from GE Healthcare), the wash and elution was performed using phosphate buffer with 20 mM imidazole for wash and 350 mM imidazole for elution.

The size exclusion chromatography was achieved after His-tag purification step using Proteo SEC 3-70 kDa HR (high performance rigid resin derived from a copolymer of Dextran and Agarose) following the recommended protocol from the supplier company. Three peaks were obtained after these two step purification process **(Figure 3-13 A)**, according to SDS-PAGE results **(Figure 3-13 B)** the first peak produced two faint bands at ≈ 90 kDa and 75 kDa, the second one comprises a large band at ≈ 65 kDa and the third peak covers several bands ranging from 30 kDa and 20 kDa. The 65 kDa band was cut and identified by mass spectrometry PMF (peptide mass finger printing) to be glmS (glutamine-fructose-6-phosphate aminotransferase), once again no Caf1R was gained from the last purification attempt. The purification process of Caf1R was not successful after these attempts. Therefore, the unproductive purification of Caf1R protein was a turning point to move toward molecular biology techniques to investigate the function of Caf1R in more details.

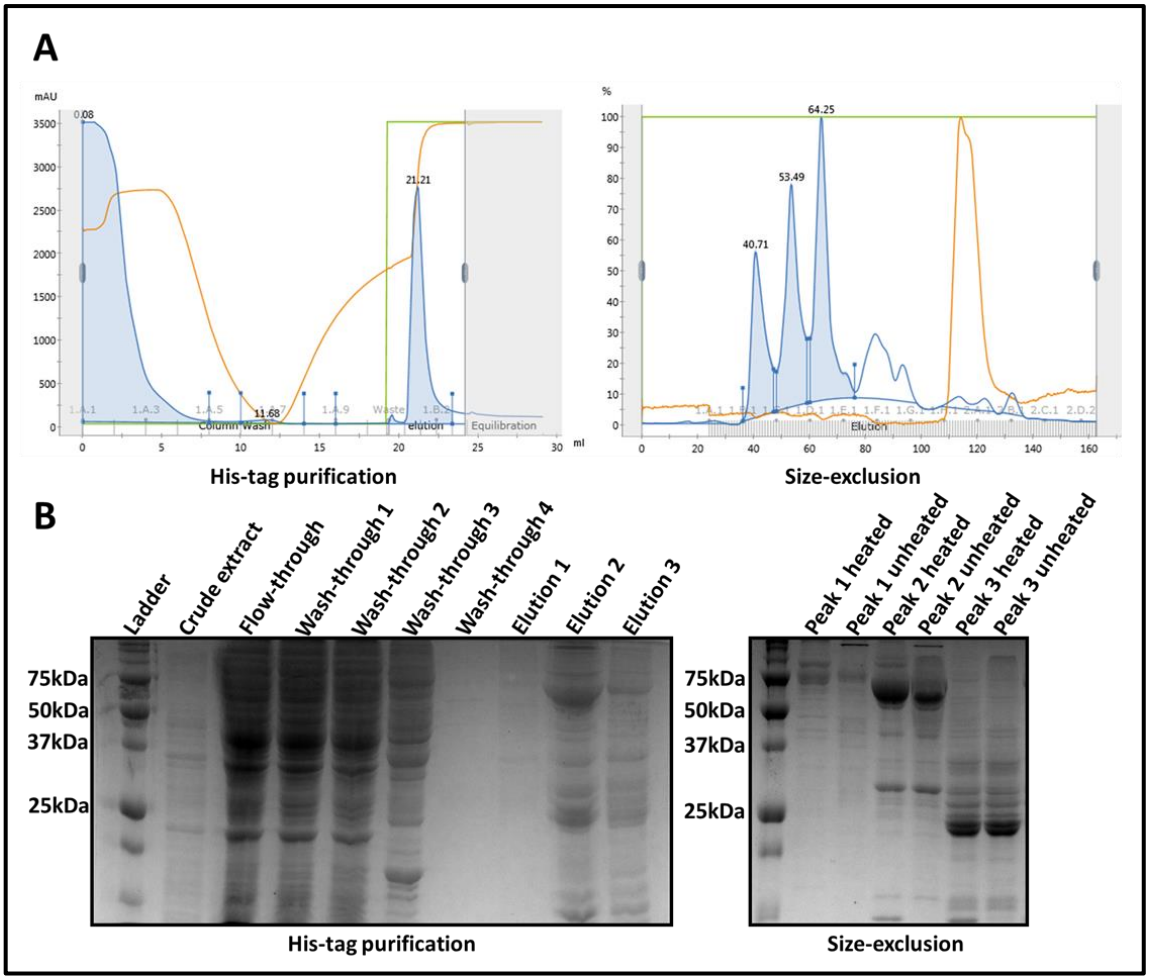

Figure 3-13. Caf1R purification process. Using *E. coli* C41 (DE3) cells transformed with pTolT Caf1RCO (the derivative of pET8c which contains TolA protein with N-terminal 6 Histidine tag as fusion partner attached to codon-optimised *caf1R*). (A) Graphs of two purification steps, His-tag purification using Ni Separose column (1ml HisTrap HP from GE Healthcare) and size exclusion using Proteo SEC 3-70 kDa HR column showing the peaks of proteins after each step. (B) SDS-PAGE analysis of the samples obtained after each purification steps.

3.3. Discussion

It was observed in this study that the presence of flocculated materials which were separated from bacterial cells by centrifugation indicates the production of Caf1 polymer. In addition, it was proved that the amounts of Caf1 polymer produced corresponds to the flocculent layer volume in the Caf1 producing cultures as in (3.2.2). Based on this phenomenon, the levels of Caf1 polymer was estimated by measuring the flocculent layer in each culture. The recombinant Caf1 polymer was the main protein component of the flocculent layer produced by *E. coli* JM101 transformed with pAH34L after the centrifugation process (Miller *et al.*, 1998). Miller *et al.* (1998) also identified that the centrifugation of Caf1 producing cultures led to separation of the capsule material to form the flocculent layer above the layer of the sedimented bacterial cells. Friedman *et al.* (1969) stated that several capsule producing Gram-negative bacteria produce a flocculated material which is composed of extracellular polymer along with polysaccharide components.

Although the Caf1R is the positive transcriptional regulator of the *caf1* operon as shown in this study and also reported by Karlyshev *et al.* (1992b), the overexpression of *caf1R* as shown in (3.2.3) and (3.2.4) has a negative effect on the Caf1 production. Whereas, the overexpression of the positive transcription factors from AraC family led to an increase in the transcription of the operons which they are responsible for. Gambino *et al.* (1993) described that the ability of resistance to multiple antibiotics was shown by *E. coli* when the MarA transcription activator was overexpressed. Asako *et al.* (1997) indicated that the overexpression of MarA results in an increase in antibiotic resistance and organic solvents tolerance in *E. coli*. Similarly, the

overexpression of the Rob transcription activator led to an increase in the resistance to multiple antibiotics in *E. coli* (Ariza *et al.*, 1995).

The overexpression of other AraC positive transcription regulator SoxS also conferred multiple antibiotics resistance in *E. coli* (Nakajima *et al.*, 1995). SoxS is responsible for activation of 17 genes in response to superoxide stress conditions in *E. coli* (Liochev and Fridovich, 1992; Fawcett and Wolf, 1995; Liochev *et al.*, 1999). Therefore, it was expected to observe an increase in Caf1 production as a result of *caf1R* overexpression which was not the case. The negative impact of *caf1R* overexpression on the *caf1* expression implied that the Caf1R transcription activator could have different mechanism of regulation as compared with its homologs, MarA, Rob and SoxS.

The presence of *caf1R* gene is required for *caf1* operon expression as shown in this study (3.2.5) which agrees with what was reported by Karlyshev *et al.* (1992b), since, Caf1R acts to activate the transcription of *caf1* operon in response to temperature and without Caf1R no expression will be obtained. Hächler *et al.* (1991) stated that the mutation of *marA* locus led to revert *E. coli* cells to being susceptible after being resistant to low concentrations of tetracycline and chloramphenicol.

The resistance to these antibiotics was rescued again after being transformed with a plasmid containing an intact *marA* locus (Hächler *et al.*, 1991). Kabir and Shimizu (2006) showed that the deletion of the *soxS* gene caused glucose uptake and growth reduction in *E. coli*. At the same time, the production of acetate was increased considerably as compared to the wild type strain as a result of tricarboxylic acid cycle downregulation (Kabir and Shimizu, 2006). This effect of *soxS* gene deletion was attributed to the loss

of upregulation of *cyoA*, *nuoE* and *ndh* genes which are involved in the chain of electron transport and respiratory system (Kabir and Shimizu, 2006).

It was confirmed by this study (3.2.8) that the presence of *caf1R* gene affects the performance of gene expression driven by T7 RNA polymerase negatively. The deletion of the *caf1R* gene in pT7-COP caused a large increase in Caf1 production as in (3.2.6). The silencing of the *caf1R* gene by substitution of start codon with a random three nitrogen bases and insertion of three stop codons at the beginning of the gene led to the production of higher amounts of Caf1 polymer (3.2.7) which is another evidence of the negative impact of Caf1R on the T7 RNA polymerase function. The reason for such effect was thought due to the binding of Caf1R in the way of T7 RNA polymerase during the transcription process.

Guajardo *et al.* (1998) showed that the binding of *lac* repressor downstream to the T7 promoter leads to create a roadblock to hinder the T7 RNA polymerase from moving forward during transcription process. The obstruction of T7 RNA polymerase from performing the transcription by the bound *lac* repressor is a consequence of either the physical conflicts between T7 RNA polymerase and the DNA-bound protein or the structural changes in DNA as a result of the transcription bubble formation which compete with the retaining of bound protein interaction with the DNA (Guajardo *et al.*, 1998). Lopez *et al.* (1998) also stated that the transcription extension by T7 RNA polymerase was thwarted by binding of *lac* repressor downstream to the T7 promoter. Madan Babu and Teichmann (2003) reported that the majority of the transcription repressors bind downstream to the promoters to hinder the processing of the RNA polymerases.

The purification of Caf1R was performed using different strains of *E. coli* and different expression vectors. At the end of this study, no pure Caf1R was obtained. The analysis of protein parameters by ProtParam webserver (Gasteiger *et al.*, 2005) predicted that Caf1R is unstable protein. It was reported by Kolodrubetz and Schleif (1981) that both activators and repressors of AraC family transcription factors are unstable *in vivo* with half-lives around 60 minutes. The instability and insolubility were clearly verified for AraC family members and their homologs which contain helix-turn-helix DNA-binding domains (Schleif, 2000). Therefore, the biochemical examinations of this family were affected negatively (Schleif, 2000). It was reported that the MelR transcription activator from the AraC family is as unstable and insoluble as other members in the family (Michan *et al.*, 1995).

Even though, several other transcription factors from the same family were purified and characterised successfully such as XylR (Ni *et al.*, 2013), MarA (Jair *et al.*, 1995) and Rob (Hyock Joo *et al.*, 2000) which are Caf1R homologs. Kumar (2016) reported that seven different plasmid constructs were used to optimise the purification of Caf1R including both codons optimised and native *caf1R* gene. After a long optimisation process, the MBP protein expression tag was used as a fusion partner with Caf1R to enhance its solubility (Kumar, 2016). The total amount of Caf1R obtained from this method was 300-350 µg and the protein was obviously suboptimal (Kumar, 2016). It was recommended that further optimisation studies are required to obtain a soluble, functional and sufficiently abundant sample of Caf1R to be used for characterisation (Kumar, 2016).

Chapter Four: Caf1R regulates transcription of the plague Capsular Antigen F1 (*caf1*) operon in response to temperature

4. Chapter Four: Caf1R regulates transcription of the plague Capsular Antigen F1 (*caf1*) operon in response to temperature

4.1. Introduction

Pathogenic bacteria can sense the temperature of the host body to provoke the production of required virulence factors to defend themselves from the host immune system. The molecular mechanism of such temperature-sensitive regulation might be at the transcriptional or translational level involving DNA, RNA or protein (Shapiro and Cowen, 2012). For example, a DNA thermosensor can be an efficient way to regulate transcription in response to temperature alteration by making changes in DNA supercoiling (López-García and Forterre, 2000).

In *E. coli*, negatively supercoiled DNA is loosened transiently as a response to heat stress which can result in plasmid relaxation and transcription initiation. Whereas, the cold shock can induce the transient negative supercoiling of DNA to obstruct the transcription (Kataoka *et al.*, 1996; Mizushima *et al.*, 1997). RNA thermometers are also involved in the temperature responsive regulation in some Gram-negative bacteria such as *E. coli* and *Salmonella* species (Waldminghaus *et al.*, 2005).

The mechanism by which the RNA thermometers regulate the translation process is by forming hairpin structure when the ribosome-binding sequence (Shine-Dalgarno sequence) base pairs with the AUG start codon in the 5' untranslated region of mRNA, this hairpin structure can prevent the mRNA from binding to the ribosome and halt the translation process. The hairpin structure is destabilised in response to raised temperature to facilitate the ribosome binding and translation initiation (Jens and Franz, 2012). In addition, the regulation of gene expression could be achieved using protein

temperature sensors including transcriptional regulators, kinases and chaperones (Klinkert and Narberhaus, 2009).

The Caf1R protein is a positive transcriptional regulator which shows high homology with the non-canonical transcriptional activators, Rob, SoxS and MarA of the AraC family which are involved in response to the superoxide and resistance to antibiotics and heavy metals (Gallegos *et al.*, 1997). When *Yersinia pestis* infects the flea where the temperature ≈ 25^{o}C the bacteria do not produce the Capsular Antigen F1 (Caf1), while, the bacteria can sense the change in temperature when they invade the human body (Perry and Fetherston, 1997) to produce Caf1 polymer as a coat to avoid the phagocytosis by macrophages (Du *et al.*, 2002). The gene transcription of the *caf1* operon is increased in response to high temperature (37^{o}C) (Han *et al.*, 2004; Motin *et al.*, 2004). Although Caf1R is thought to be the responsible factor for such upregulation (Karlyshev *et al.*, 1992a), there are no clear results showing the role of Caf1R in this thermal responsive expression of the *caf1* operon genes.

In this chapter, it was found that *caf1* operon genes *caf1M*, *caf1A* and *caf1* are transcribed as a single long mRNA. Both intergenic sequence 2 (between *caf1M* and *caf1A*) and 3 (between *caf1A* and *caf1*) were detected in the cDNA of *E. coli* BL21 (DE3) culture transformed with the plasmid pCOP (contains *caf1* operon). Furthermore, the results obtained from a rapid amplification of 5' cDNA ends experiment (RACE) showed that part of *caf1*, intergenic region 3 and part of *caf1A* are present together in a single cDNA which was synthesized using the total RNA extracted from pCOP transformant cultures.

The deletion of one predicted promoter region in the intergenic sequence 1 (between *caf1R* and *caf1M*) was seen to have a significant effect on the transcription of *caf1* operon genes and protein production when analysed by RT-PCR and western blot. At the same time, the deletion of the other two predicted promoters in the same intergenic region (1) did not have a substantial effect on the both *caf1* operon transcripts and proteins levels. Moreover, the analysis of ribosome-binding site function (RBS) using RT-PCR and western blot revealed that the substitution of these predicted RBS led to change the transcript levels of *caf1* operon genes which is not supposed to happen theoretically.

Nonetheless, the predicted RBS of *caf1* transcript (TAGAGGT) could be confirmed, because though the levels all *caf1* transcripts were similar in the cultures of either pCOPFΔM-RBS (where predicted *caf1M*-RBS is substituted with a random sequence), pCOPFΔA-RBS (where predicted *caf1A*-RBS is substituted with a random sequence) or pCOPFΔF-RBS (where predicted *caf1*-RBS is substituted with a random sequence), the Caf1 protein bands were shown by western blot in both cultures of pCOPFΔM-RBS and pCOPFΔA-RBS but not with pCOPFΔF-RBS.

The investigation of *caf1* operon gene expression by RT-PCR and western blot in the presence and absence of *caf1R* gene at two different temperatures illustrated that Caf1R is responsible for the immediate upregulation of the *caf1* operon gene expression in response to the thermal change from 25^{o}C to 35^{o}C. The sequence alignment and structural modelling demonstrated that there is a high similarity among Caf1R, Rob and MarA in the two helix-turn-helix motifs in the N-terminus which are involved in DNA binding. The DNA-binding residues are partially conserved in these three

proteins with several mutations in Caf1R suggesting that Caf1R-binding site could be slightly different. On the other hand, the C-terminus of Caf1R is quite different from Rob's and MarA's one and it has been identified as a member of GyrI-like small molecule binding domain family which contains two strand-helix-strand-strand (SHS2) motifs (Anantharaman and Aravind, 2004) opposite to each other and that could indicate that small molecule effectors might be involve in the function of Caf1R's C-terminus. Finally, the deletion of Caf1R's C-terminus stopped the production of Caf1 protein implying that the C-terminus of Caf1R is important for its function as a transcriptional activator.

4.2. Results

4.2.1. The *caf1* operon genes, *caf1M*, *caf1A* and *caf1* are transcribed as a single polycistronic transcript

caf1 operon consists of four structural genes which are *caf1R*, *caf1M*, *caf1A* and *caf1* in that order (**Figure 4-1 A**), they are separated by three intergenic regions, the first intergenic region (*intergenic region 1*) is 328 nucleotides (nt) long between *caf1R* and *caf1M*, *intergenic region 2* is located between *caf1M* and *caf1A* and is only 24 nt long, finally the third intergenic region which separates *caf1A* from *caf1* that is 80 nt long. The *caf1R* gene, which encodes for the putative transcriptional regulator, is transcribed in the opposite direction to the rest of the operon. Here, I wished to characterise the expression of the *caf1* operon genes using PCR analysis of RNA transcripts.

E coli BL21 (DE3) cells transformed with either pCOP (contains *caf1* operon) or pT7-COP (contains *caf1* operon preceded by an upstream T7 promoter), were grown at 35^{o}C for 16 h to produce Caf1 polymer and then total RNA extraction was followed by cDNA synthesis. The cDNA was used as a template for amplification (PCR) of intergenic regions 1, 2 and 3 using primers complementary to the sequence either side of the intergenic regions (**Figure 4-1 A**), in order to examine the size of the *caf1* operon multigenic transcript. Therefore, the amplification of intergenic regions 2 and 3 in pCOP transformant would indicate that *caf1* operon genes are transcribed as a one large transcription unit, whereas, the lack of amplification would mean that these genes are transcribed as three small RNA transcripts (*caf1M*, *caf1A* and *caf1*). In the culture of *E. coli* BL21 (DE3) transformed with pT7-COPT7 (with T7 promoter upstream to *caf1R*), the combined amplification of all three

intergenic regions 1, 2 and 3 suggests that T7 RNA polymerase transcribes all three genes together.

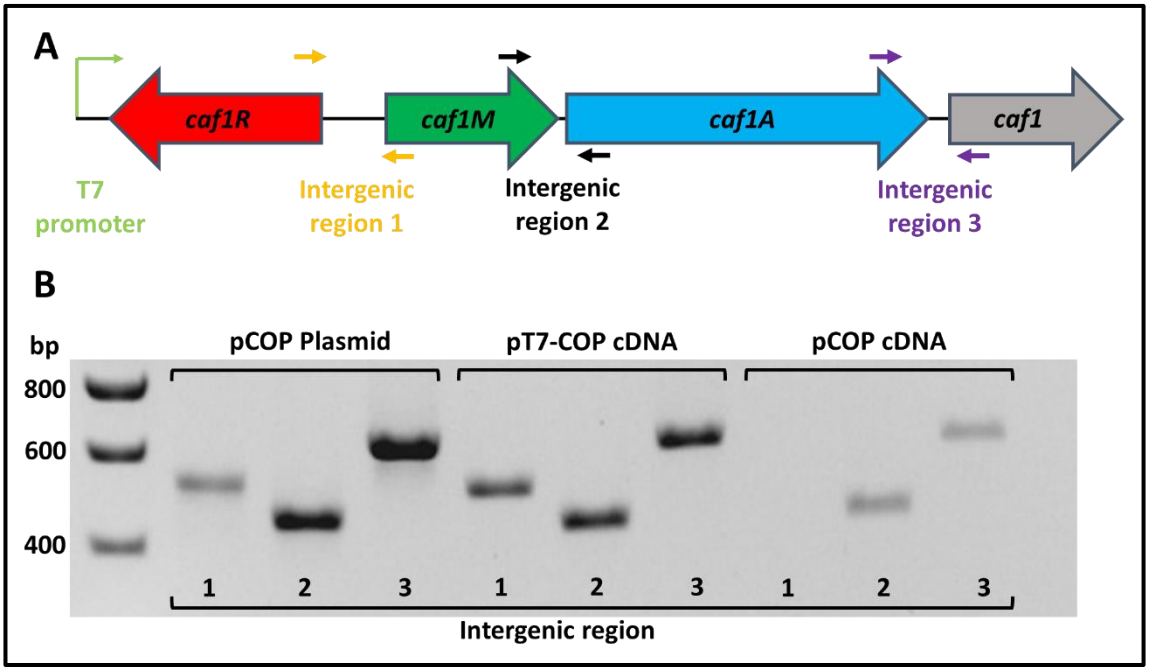

Figure 4-1. Schematic of primer design. (A) A diagram of the *caf1* operon present in the plasmids used in this study is shown, with arrows corresponding to the forward and reverse primers and placed in the approximate position where they bind. Forward primers are shown on top of the genes and reverse primers shown beneath. Primers used for detecting the presence of intergenic regions in cDNA are shown: orange for intergenic region 1, black for intergenic region 2 and purple for intergenic region 3. (B) Agarose gel showing the PCR amplification products corresponding to the three intergenic regions of the *caf1* operon generated from either the pCOP plasmid, or cDNA made using mRNA from cultures of *E. coli* transformed with pT7-COP or pCOP plasmids and grown at 35°C for 16 h. Both plasmids contain the *caf1* operon as shown in (A).

The agarose gel of PCR products (**Figure 4-1 B**) demonstrates that when using pCOP plasmid as a positive control, all intergenic regions 1, 2 and 3 were amplified. This was same for pT7-COP cDNA implying that the transcription was driven by T7 RNA polymerase starting from T7 promoter upstream of *caf1R*. However, the intergenic regions 2 and 3 were amplified when pCOP cDNA used as a template suggesting that both *caf1M* and *caf1A* and *caf1A*

and *caf1* are co-transcribed together. Consequently, *caf1M*, *caf1A* and *caf1* genes are most likely to be transcribed as a long, single, multigenic mRNA (of approximately 3900 nt) in the natural system. Although intergenic region 2 is unlikely to contain a promoter due to its short length (only 24 nt), intergenic region 1, being upstream of all three genes and relatively long, is likely to contain both promoter and Caf1R-binding site.

4.2.2. Identification of *caf1* operon promoter

The *intergenic region 1* which is a 328 nt long DNA sequence located between *caf1R* and *caf1M* was used to predict the *caf1* operon genes promoter, three promoters were predicted using Neural Network Promoter Prediction 2.2 (NNPP2.2) (Reese, 2001) webserver. The prediction process in this webserver is based on two features: the recognition of the TATA-box and the transcription start site. We hypothesized that one of the identified promoters P1, P2 and P3 **(Table 4-1)** upstream of *caf1M* is likely to drive the transcription of the *caf1* operon genes as a one mRNA transcript. Therefore, each one of these promoters was deleted separately from pCOPF (a pCOP plasmid in which Caf1R, M and A have C-terminal FLAG tags) and *E. coli* BL21 (DE3) cells were transformed either with pCOPFΔP1 (promoter 1 knocked-out), pCOPFΔP2 (promoter 2 knocked-out) or pCOPFΔP3 (promoter 3 knocked-out) and grown at 35^{o}C for 16 h.

Table 4-1. The sequences and scores of the predicted promoters

Name	Score	Sequence of the predicted promoter
P1	0.94	aattgttctcagtgaggctgtgctacggatataaaaatccccttcatttg
P2	0.95	ccttcatttgttacccacctttttacgcatatcgtcgatatgaaatgatg
P3	0.84	ccaccttttacgcatatcgtcgatatgaaatgatggggaggggggtggga

The fold-change in expression of *caf1* operon genes in each culture was evaluated by RT-PCR relative to β-*lactamase* (as a house keeping gene), where the fold-change in expression = ($2^{-\Delta Ct}$) and ΔCt = (Ct value of target gene – Ct value of β-*lactamase* gene) (Livak and Schmittgen, 2001; Dheda *et al.*, 2004) and the proteins levels were analysed by western blot using anti-FLAG and anti-Caf1 antibodies. The means of expression fold-change in the cultures of pCOPF, pCOPFΔP1, pCOPFΔP2 and pCOPFΔP3 transformants are shown in **Table 4-2** and statistical analysis is shown in **Table 4-3**.

Table 4-2. The RT-PCR analysis data of *caf1* operon predicted promoters. Showing the means of expression fold-change of the *caf1* operon genes relative to β-*lactamase* as a house keeping gene. ± value is the standard error of the mean.

Gene transcript	The mean values of the expression fold-change			
	pCOPF	**pCOPFΔP1**	**pCOPFΔP2**	**pCOPFΔP3**
caf1M	1.2 ± 0.08	0.84 ± 0.09	0.88 ± 0.04	1.18 ± 0.09
caf1A	1.07 ± 0.1	0.7 ± 0.06	0.13 ± 0.01	0.99 ± 0.09
caf1	2.23 ± 0.1	1.66 ± 0.1	0.62 ± 0.02	2.15 ± 0.1

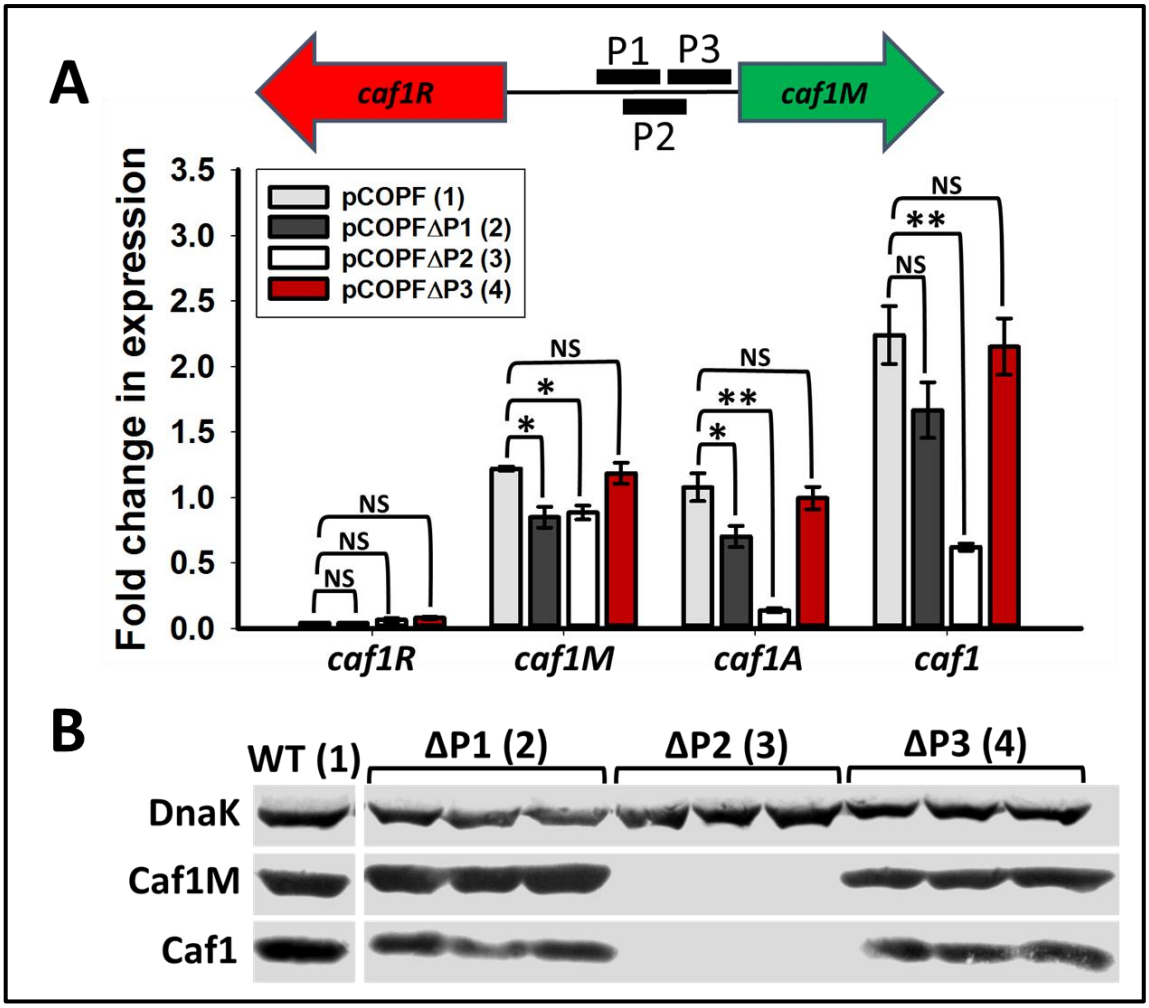

Figure 4-2. Determination of promoter site. (A) Transcript levels of the *caf1* operon genes as determined by RT-PCR, from cultures grown for 16 h of *E. coli* transformed with either pCOPF (full *caf1* operon, *caf1R*, *caf1M* and *caf1A* have FLAG tags), or pCOPF with the proposed promoter regions P1,2 or 3 deleted. Three cultures of each condition were grown, with RT-PCR reactions run in duplicate for each culture. Bar heights correspond to mean fold-change in expression relative to β-*lactamase*. Error bars represent standard error of the mean (S.E.M). A diagram detailing the positions of the P1, 2 and 3 regions is shown in the top of the graph. (B) Western blot of the above cultures showing the levels of Caf1M and Caf1 (detected by anti-FLAG tag and anti-Caf1 antibodies respectively), and using DnaK (detected by an anti-DnaK antibody) as a loading control. NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). Biological replicates = 3. Statistics analysis was performed using One-Way ANOVA test.

Table 4-3. The statistics of the RT-PCR analysis data of *caf1* operon predicted promoters. Showing the P-values of the comparisons among the different cultures using One-Way ANOVA test.

Gene transcript	The statistical differences among the mean values (P-values)		
	pCOPF vs. pCOPFΔP1	pCOPF vs. pCOPFΔP2	pCOPF vs. pCOPFΔP3
caf1M	0.01	0.004	0.786
caf1A	0.009	<0.001	0.573
caf1	0.04	<0.001	0.732

Although the deletion of the P1 sequence resulted in a slight decrease in expression of all the *caf1* genes **(Figure 4-2 A)**, the P2 region knock-out resulted in the largest decrease in the transcript levels, whereas, the levels the *caf1* operon transcripts obtained from P3 deletion were very close to the levels of the pCOPF transformant (the natural system). On the other hand, the western blot analysis (**figure 4-2 B**) revealed that although, both pCOPFΔP1 and pCOPFΔP3 transformants produced Ca1M and Caf1 bands of similar intensity to the pCOPF transformant (the natural system), the P2 region deletion prevents Caf1M and Caf1 production. Therefore, the predicted region 2 is most likely to contain the promoter responsible for *caf1* gene transcription as a long single mRNA. The P2 region is 44 nt upstream from the potential *caf1M*-RBS (ribosome-binding site).

4.2.3. Identification of the *caf1* operon Ribosome-Binding Sites

In order to comprehend the translation of such a multigenic mRNA of *caf1* operon, the location of the ribosome-binding sites (which are required for proteins synthesis) of the *caf1* operon were determined by manual sequence analysis by comparing with known consensus sequences. Three RBS sites were recognised close to the translational start codons of each gene (**Figure**

4-3 A), the *caf1M*-RBS is a consensus Shine-Dalgarno sequence (AGGAGGT), whereas both *caf1A*-RBS and *caf1*-RBS are more likely to be the non-consensus RBS sites (AGGGGAC) and (TAGAGGT) respectively. To investigate these RBS sequences, they were substituted with random DNA sequences of the same lengths.

E. coli BL21 (DE3) cells were transformed with either pCOPF, pCOPFΔM-RBS (where *caf1M*-RBS is substituted with a random sequence), pCOPFΔA-RBS (where *caf1A*-RBS is substituted with a random sequence) or pCOPFΔF-RBS (where *caf1*-RBS is substituted with a random sequence), three single colonies from each transformant were used to inoculate 5 ml liquid cultures and grown at 35^{o}C for 16 h, total RNA was extracted and cDNA synthesized to be analysed by RT-PCR, the western blot was performed to test the effect of RBS sites substitution on protein production using samples from the same cultures used for RT-PCR.

Although the deletion of the RBS sites is not supposed to affect the patterns of RNA transcripts, interestingly, the transcripts levels were altered (**Figure 4-3 A**) as compared to the pCOPF transformant (which contains the natural *caf1* operon), the deletion of *caf1M*-RBS leads to large decrease in all RNA transcripts levels with statistically significant difference when comparing with pCOPF cultures and similar reductions were obtained in pCOPFΔA-RBS cultures, whereas, the levels of transcripts in pCOPFΔF-RBS culture were similar to pCOPF transformant except for *caf1* transcript there was a clear decline as well.

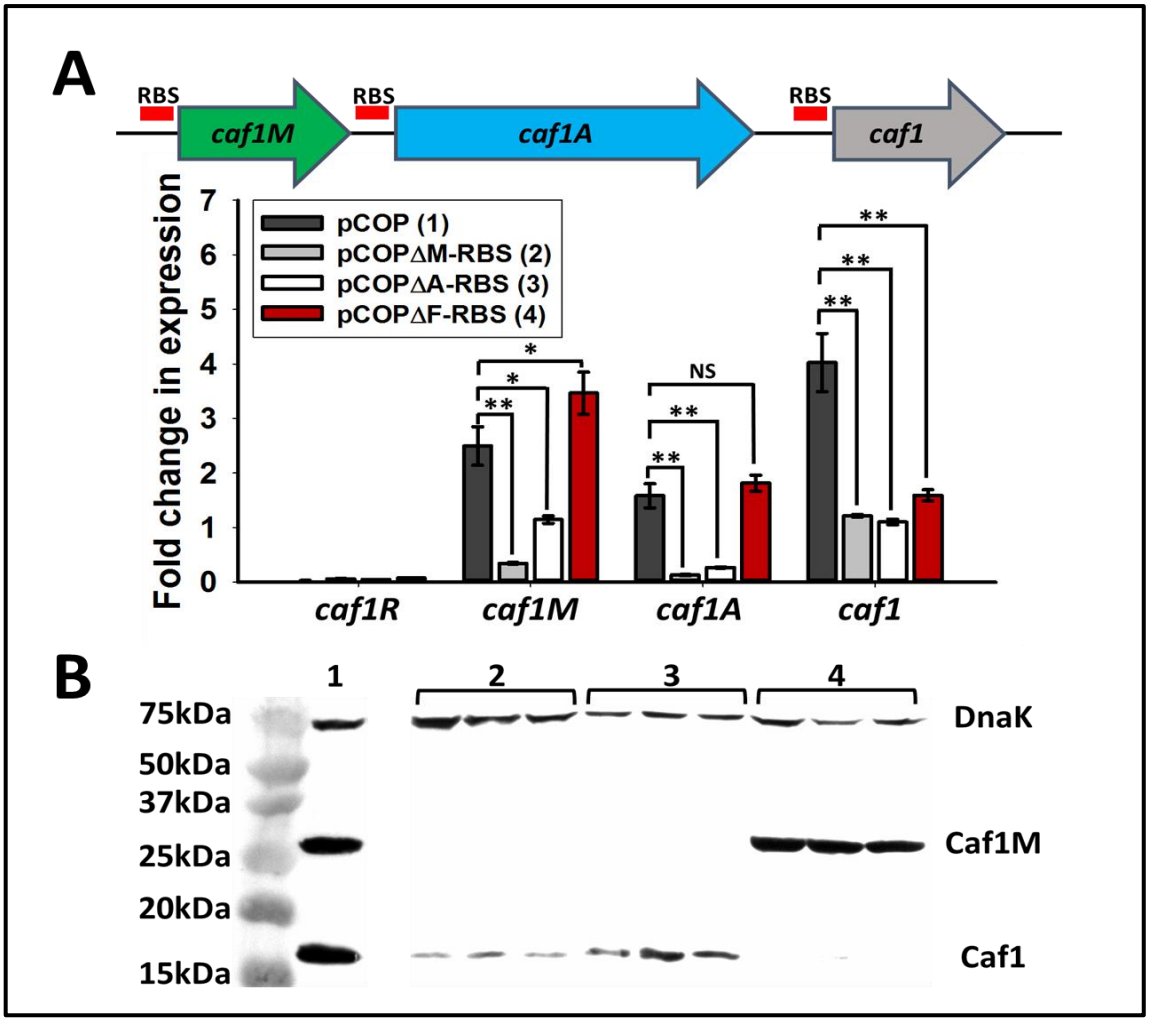

Figure 4-3. Determination of *caf1* operon ribosome-binding sites. (A) Transcript levels of the *caf1* operon genes as determined by RT-PCR, from cultures grown for 16 h of *E. coli* transformed with either pCOPF (full *caf1* operon, *caf1R*, *M* and *A* have FLAG tags), or pCOPF with the detected RBS of *caf1M*, *caf1A* and *caf1* deleted separately. Three cultures of each condition were grown, with RT-PCR reactions run in duplicate for each culture. Bar heights correspond to mean fold-change in expression relative to β-*lactamase*. Error bars represent standard error of the mean (S.E.M). A diagram detailing the positions of the *caf1M*-RBS, *caf1A*-RBS and *caf1*-RBS is shown in the top of the graph. (B) Western blot of the cultures mentioned in (A). *E. coli* BL21 transformed with either pCOPF (1), pCOPFΔM-RBS (2), pCOPFΔA-RBS (3) or pCOPFΔF-RBS (4) grown at 35°C for 16 h. Showing the levels of Caf1M and Caf1 (detected by anti-FLAG tag and anti-Caf1 antibodies respectively), and using DnaK (detected by an anti-DnaK antibody) as a loading control. NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). Biological replicates = 3. Statistics analysis was performed using One-Way ANOVA test.

Table 4-4. The RT-PCR analysis data of the predicted ribosome binding-sites in *caf1* operon. Showing the means of expression fold-change of the *caf1* operon genes relative to β-*lactamase* as a house keeping gene. ± value is the standard error of the mean.

Gene transcript	The means of the expression fold-change			
	pCOPF	pCOPFΔ*M*-RBS	pCOPFΔ*A*-RBS	pCOPFΔ*F*-RBS
caf1R	0.02 ± 0.002	0.05 ± 0.006	0.04 ± 0.003	0.07 ± 0.005
caf1M	2.49 ± 0.3	0.34 ± 0.01	1.14 ± 0.06	3.46 ± 0.3
caf1A	1.58 ± 0.2	0.13 ± 0.01	0.26 ± 0.01	1.81 ± 0.1
caf1	4.02 ± 0.5	1.21 ± 0.02	1.10 ± 0.05	1.58 ± 0.1

The western blot results (**Figure 4-3 B**) showed that the Caf1M production ceased in both pCOPFΔM-RBS and pCOPFΔA-RBS, whilst, the Caf1 bands were less intense in the *caf1M*-RBS deletion even when the *caf1* transcript levels obtained by RT-PCR are similar for these two transformants. This suggests that the absence of Caf1M (chaperone) which protects the Caf1 subunits from aggregation and protease digestion during the periplasm cause a decrease in Caf1 levels (Galyov *et al.*, 1991a; Zavyalov *et al.*, 1997; Chapman *et al.*, 1999). Interestingly, the levels of *caf1* operon transcripts decreased after RBS deletion implying that the stability of transcripts of the *caf1* operon were reduced. This might actually be as a result of the deletion of the RBS sites, since the binding of mRNA to ribosomes gives it more stability. Although the levels of RNA transcripts of the *caf1* gene are similar in cultures of pCOPFΔM-RBS, pCOPFΔA-RBS and pCOPFΔF-RBS or even a slightly more in pCOPFΔF-RBS, the Caf1 protein bands disappeared in the *caf1*-RBS deletion indicating that the RBS site of *caf1* identified here is more likely the right Shine-Dalgarno sequence of *caf1* gene. Additional investigations of promoters, RBS sites and Caf1R-binding sites need to be achieved to know their sequence exactly which is very important to

determine the molecular mechanism of transcription and translation in this system.

4.2.4. Analysing of *caf1* transcripts size using rapid amplification of 5' cDNA ends (5' RACE)

The determining of RNA transcripts full sequence is an efficient way to identify the transcription start site and the responsible promoter(s) for *caf1* operon transcription. *E. coli* BL21 (DE3) cells were transformed with pCOPF, grown at 35^{o}C for 16 h, total RNA was extracted and the cDNA was synthesized by reverse transcription (**Figure 4-4 A**) using a gene specific primer of 9 nt (GSP-RT) (RT stands for reverse transcription) complementary to a sequence in the *caf1* transcript. The cDNA was purified to remove the primers and nucleotides to be used in terminal transferase reaction. This reaction was done to add a poly adenine tail to the 3' termini of cDNA which will provide then a primer-binding site at cDNA 3' end.

Next, the amplification of these cDNA was done using the poly A tailed-cDNA, a gene specific primer (GSP1) which is complementary to a site on the *caf1* gene in the antiparallel strand of cDNA upstream to the (GSP-RT) primer, $(dT)_{17}$-adapter primer which contains tail of 17 thymine at 3' end and other 18 nt and the adapter primer which resembles the $(dT)_{17}$-adapter primer in the 18 nt without thymine tail and other PCR components. The PCR products were analysed by agarose gel electrophoresis (**Figure 4-4 B**) and the DNA bands were extracted and sent for sequencing.

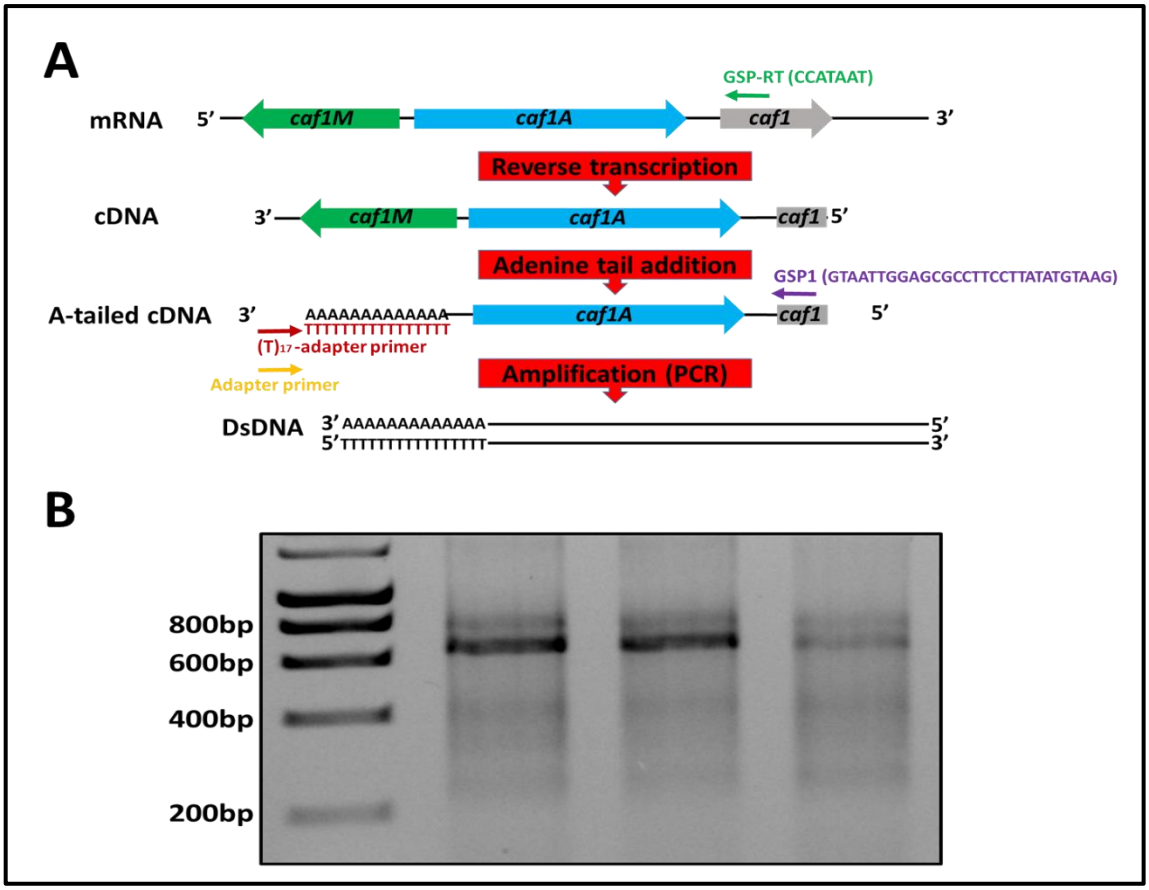

Figure 4-4. Rapid amplification of 5' cDNA ends (5' RACE). (A) A diagram showing the workflow of 5' RACE. The main steps are: (1) Total RNA extraction from pCOP transformants culture, (2) reverse transcription reaction using the extracted RNA as a template and GSP-RT primer to synthesise cDNA, (3) adenine-tail addition to 3' end of the pure cDNA by terminal transferase reaction and finally (4) the amplification of the adenine-tailed cDNA by PCR using GSP1 primer, (T)17-adapter and adapter primers. The green arrow (GSP-RT) is the gene specific primer of 9 nt used in reverse transcription reaction, the purple arrow (GSP1) is the gene specific primer used in PCR, it is complementary to a small part of *caf1* gene on the antiparallel strand of cDNA, the dark red arrow $(dT)_{17}$-adapter primer contains 17 thymine tail complementary to the adenine tail in cDNA strand and the orange arrow refers to adapter primer which resemble the $(dT)_{17}$-adapter primer without thymine tail. (B) Agarose gel electrophoresis of 5' RACE PCR products using adenine tailed cDNA as a template DNA, GSP1, $(dT)_{17}$-adapter and adapter primers mentioned in (A) the shown bands are the amplified adenine-tailed cDNA from the cultures of *E. coli* BL21 (DE3) transformed with pCOP grown at 35°C for 16 h.

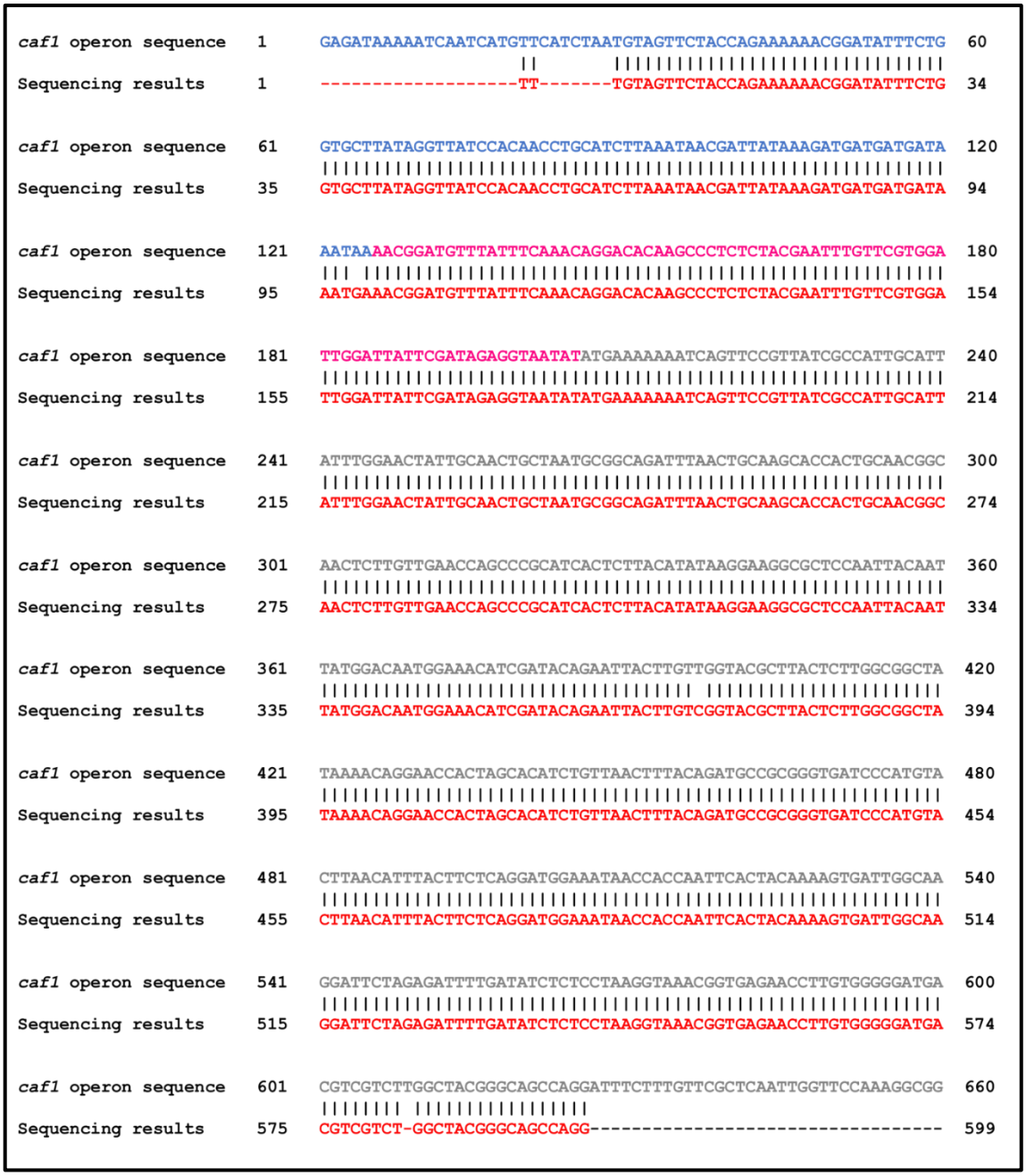

Figure 4-5. The sequence alignment of 5' RACE PCR products sequencing results and *caf1* operon. The PCR products were purified from agarose gel shown in (Figure 4-4 B) and sent to (Eurofins Genomics) using GSP1 primer which is complementary to a short sequence in the *caf1* gene. The sequence 1 is a part of *caf1* operon includes a small part of *caf1A* in blue colour and intergenic sequence 3 in pink colour and a part of *caf1* in grey colour. The sequence 2 is the sequencing results of 5' RACE PCR products obtained from Eurofins Genomics.

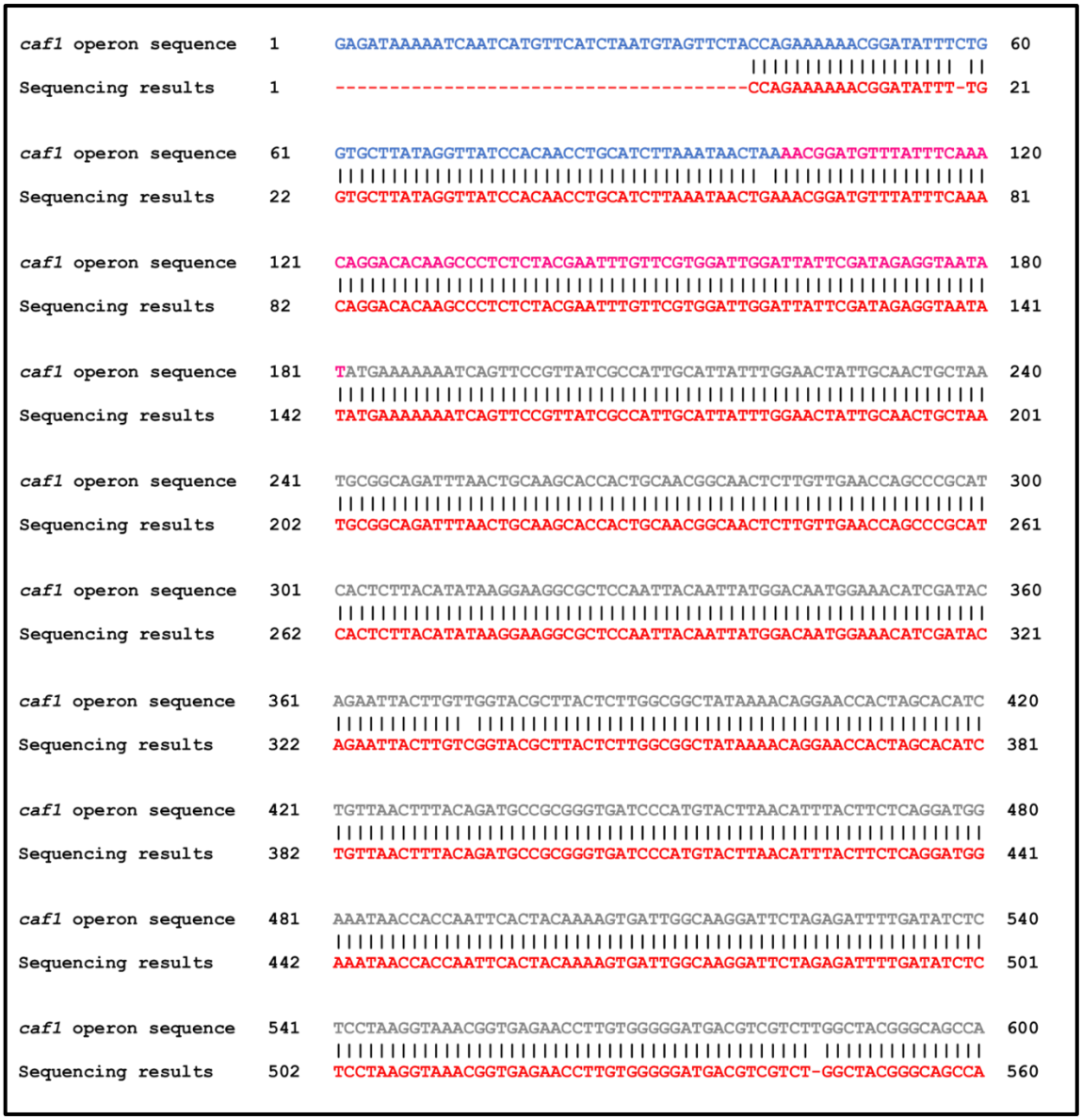

Figure 4-6. The sequence alignment of 5' RACE PCR products sequencing results and *caf1* operon from different sample. The PCR products were purified from agarose gel shown in (Figure 4-4 B) and sent to (Eurofins Genomics) using GSP1 primer which is complementary to a short sequence in the *caf1* gene. The sequence 1 is a part of *caf1* operon includes a small part of *caf1A* in blue colour and intergenic sequence 3 in pink colour and a part of *caf1* in grey colour. The sequence 2 is the sequencing results of 5' RACE PCR products obtained from Eurofins Genomics.

The obtained DNA sequencing results (**Figures 4-5 and 4-6**) confirm the presence of polycistronic transcript, since the sequence of PCR products from using poly adenine-tailed cDNA as a template DNA were complementary to a part of *caf1* operon started from *caf1* gene to *caf1A* gene including *intergenic region 3* (**Figure 4-4 A**) and because the intergenic region 2 is very short sequence only 24 nt, it would be unlikely to contain a promoter, at least 8 different 5' RACE samples from different pCOPF transformants cultures were sequenced and showed similar results. Therefore, these results provide another evidence to verify the polycistronic transcription of *caf1* operon genes.

Even though the 5' RACE gave an additional proof of the presence of polycistronic mRNA transcript of *caf1* genes, the exact promoter(s) or transcription starting site could not be identified due to lack of sufficient time in this project to carry out the long optimisation process required to obtain successful results from 5' RACE technique. The transcription of *caf1* genes is thermo-responsive (Perry and Fetherston, 1997; Han *et al.*, 2004; Motin *et al.*, 2004) and we next sought to clarify if Caf1R is likely to be the responsible factor for such temperature-sensitive transcription.

4.2.5. Caf1R is a thermo-responsive transcriptional regulator

The production of Caf1 polymer in *Y. pestis* is a temperature responsive process, the bacteria do not produce the capsule in the flea where the ambient temperature is around 25^{o}C (Anisimov *et al.*, 2005), but when the bacteria infect the human body the capsule secretion is induced in response to human body temperature 37^{o}C (Perry and Fetherston, 1997; Han *et al.*, 2004) in order to resist the phagocytosis by immune cells (Du, 2002). To

investigate the role of Caf1R in such a temperature-sensitive regulation mechanism at molecular level, *E. coli* BL21 (DE3) cells were transformed with either pCOPF or pCOPFΔR, three different colonies from each transformants grown in 5 ml liquid medium at two different temperatures 25°C and 35°C for 16 h, the total RNA was extracted and cDNA was synthesized for all cultures. RT-PCR technique was used to assess the levels of *caf1* gene's transcripts in each separate case and the protein synthesis was analysed by western blot, the means of expression-fold change were calculated relative to β-*lactamase* (as a house keeping gene) (**Figure 4-7**) for pCOPF transformants cultures grown at 25°C, 35°C, together with pCOPFΔR transformants cultures grown at 25°C and 35°C as shown in **Table 4-5**.

Table 4-5. The RT-PCR data of the thermal-responsive regulation by Caf1R. Showing the means of expression fold-change of the pCOPF and pCOPFΔR cultures grown either at 25°C or 35°C relative to β-*lactamase* as a house keeping gene. ± value is the standard error of the mean.

Gene transcript	The mean values of the expression fold-change			
	pCOPF at 25°C	pCOPF at 35°C	pCOPFΔR at 25°C	pCOPFΔR at 35°C
caf1R	0.15 ± 0.006	0.02 ± 0.002	0.0001 ± 0	0.0007 ± 0
caf1M	0.81 ± 0.05	2.49 ± 0.3	0.21 ± 0.01	0.25 ± 0.01
caf1A	0.29 ± 0.03	1.58 ± 0.2	0.07 ± 0.004	0.11 ± 0.007
caf1	3.86 ± 0.2	4.02 ± 0.5	1.32 ± 0.1	0.76 ± 0.08

The deletion of *caf1R* resulted in large decrease in all RNA transcripts with statistically significant differences as compared to intact *caf1* operon. The P-values of the statistical analysis are shown in **Table 4-6**. Although the *caf1M* and *caf1A* transcripts levels increased significantly by higher temperature in the natural *caf1* operon. Neither of the transcript levels (*caf1M* and *caf1A*)

were influenced by temperature change in the absence of *caf1R* in pCOPFΔR transformant. Surprisingly, there were no differences in *caf1* RNA transcript levels in response to thermal change for both pCOPF and pCOPFΔR transformant.

Table 4-6. The statistics of the RT-PCR analysis data of the thermal responsive transcriptional regulation by Caf1R. Showing the P-values of the comparisons among the pCOPF and pCOPFΔR cultures grown at 25°C and 35°C using One-Way ANOVA test.

Gene transcript	The statistical differences among the mean values (P-values)	
	pCOPF at 25°C vs. at 35°C	pCOPFΔR at 25°C vs. at 35°C
caf1M	<0.001	0.88
caf1A	<0.001	0.83
caf1	0.72	0.40

In addition, western blot analysis (**Figure 4-7 C**) of the same samples used in RT-PCR experiment confirm the production of both Caf1M and Caf1 in pCOPF culture at 35°C, while, pCOPFΔR transformants grown at 35°C produce very low amount of Caf1 which might be the outcome of leaky expression by non-specific binding of T7 RNA polymerase (Smeekens and Romano, 1986). Though the *caf1* gene transcript levels in pCOPF transformant culture were similar at both 25°C and 35°C, the western blot revealed obvious Caf1 band at 35°C but not at 25°C suggesting the occurrence of post-transcriptional downregulation at 25°C which could be by RNA thermometer (Krajewski and Narberhaus, 2014; Righetti *et al.*, 2016).

Therefore, these results indicated that Caf1R is the responsible transcriptional factor for upregulation of both *caf1M* and *caf1A* genes. To investigate the role of Caf1R in *caf1* operon gene expression in response to immediate thermal change, a new experiment was performed to mimic the

instant alteration in temperature in the nature when *Y. pestis* moves from flea where the ambient temperature is around 25°C to the human body at ≈ 37°C (Anisimov *et al.*, 2005).

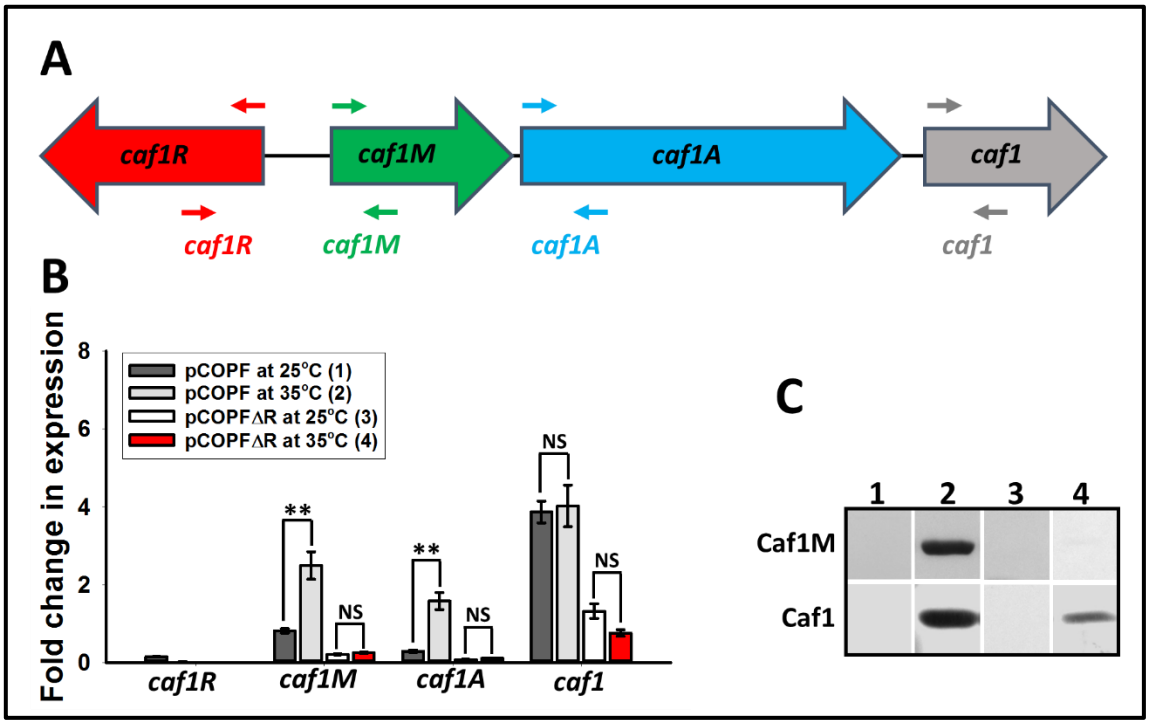

Figure 4-7. Temperature-responsive regulation of *caf1* operon gene expression. (A) A diagram of the *caf1* operon present in the plasmids used in this study is shown, with arrows corresponding to the forward and reverse primers and placed in the approximate position where they bind, forward primers are shown on top of the genes and reverse primers shown beneath, primers used for RT-PCR are shown in red for *caf1R*, green for *caf1M*, blue for *caf1A* and grey for *caf1*. (B) Graph of transcript levels, determined by RT-PCR, of each caf1 gene are shown for cultures of *E. coli* transformed with pCOPF (full *caf1* operon) and pCOPFΔR (*caf1* operon, *caf1R* deleted), grown at different temperatures 25°C and 35 °C overnight (≈16 h). Three cultures of each condition were grown, with RT-PCR reactions run in duplicate for each culture. Bar heights correspond to mean fold-change in expression relative to β-*lactamase*. Error bars represent standard error of the mean (S.E.M). (C) Western blot of the cultures mentioned in (A), *E. coli* BL21 (DE3) transformed with either pCOPF grown at 25°C for 16 h (1) and at 35°C for 16 h (2) or pCOPFΔR grown at 25°C for 16 h (3) and at 35°C for 16 h (4). Showing the levels of Caf1M and Caf1 (detected by anti-FLAG tag and anti-Caf1 antibodies respectively) in the cell pellets of the expression cultures. NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). Biological replicates = 3. Statistical analysis was performed using One-Way ANOVA test.

4.2.6. The immediate thermo-responsive transcriptional regulation by Caf1R

The molecular mechanism of the prompt temperature-sensitive activation of *caf1* operon has not been investigated. Therefore, the hypothesis that Caf1R is responsible for this instant thermoregulation was examined. *E. coli* BL21 (DE3) cells were transformed with either pCOPF or pCOPFΔR, grown at the non-inducing temperature of 25^{o}C for 16 h. The stationary phase cultures were then diluted to 0.5 OD_{600} and then grown for a further 1 h at either 25°C or 35°C. The total RNA was then extracted and cDNA was synthesised for all cultures to be analysed by RT-PCR (**Figure 4-8 A**). The mean values of the expression fold-change of *caf1* operon genes were calculated relative to β-*lactamase* (as a house keeping gene) for pCOPF transformant cultures grown at 25^{o}C for 16 h, at 25^{o}C for extra 1 h and at 35^{o}C for extra 1 h and for pCOPFΔR transformant cultures grown at 25^{o}C for 16 h, at 25^{o}C for extra 1 h and at 35^{o}C for extra 1 h **(Table 4-7)**.

The transcript levels *caf1* genes were similar at 25^{o}C and 35^{o}C for pCOPFΔR transformants (in the absence of *caf1R*), the statistics using One-Way ANOVA test **(Table 4-8)** revealed no significant differences at these two temperatures for all *caf1* operon transcripts except *caf1R* with P-values 0.821, 0.930 and 0.206 for *caf1M*, *caf1A* and *caf1* respectively. Remarkably, the presence of *caf1R* in wild type pCOPF transformant cultures caused a substantial alteration in transcription in response to temperature, since the mRNA transcripts of all *caf1* genes were increased significantly.

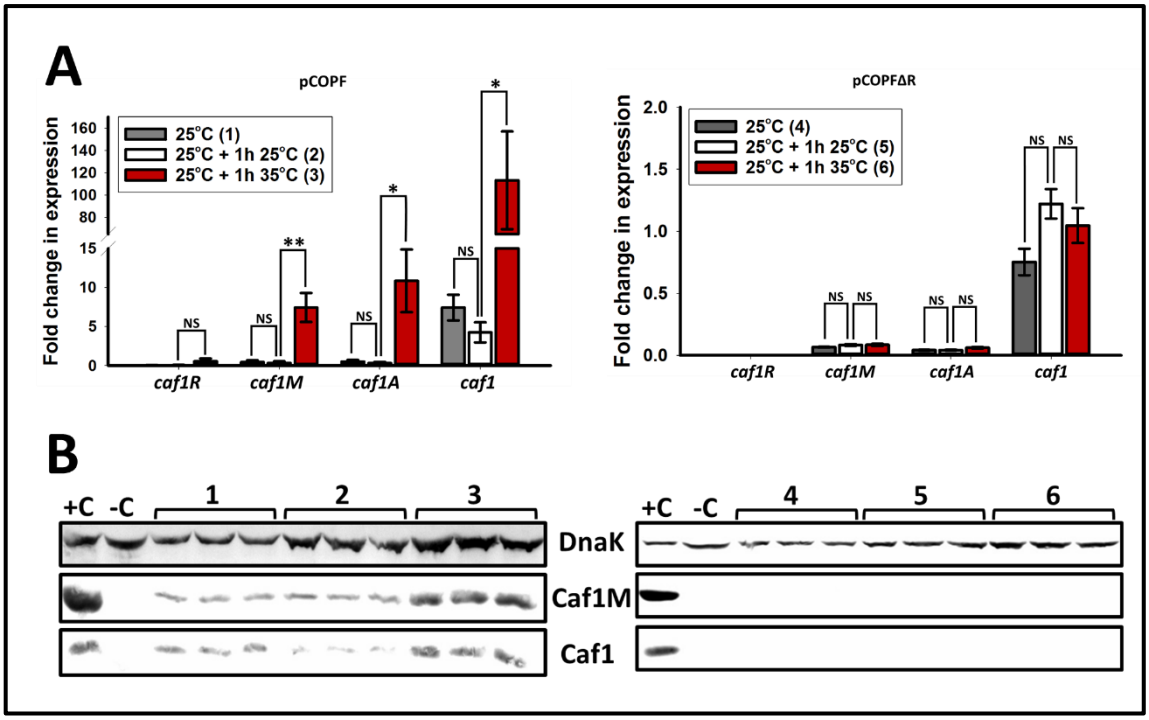

Figure 4-8. Immediate temperature-responsive regulation of *caf1* operon gene expression. (A) Transcript levels, determined by RT-PCR, of each *caf1* gene are shown for cultures of *E. coli* BL21 (DE3) cells transformed with pCOPF (full *caf1* operon) and pCOPFΔR (*caf1* operon, *caf1R* deleted, grown at 25°C overnight (≈ 16 h), then either 25°C or 35°C for 1 further hour. Three cultures of each condition were grown, with RT-PCR reactions run in duplicate for each culture. Bar heights correspond to mean fold-change in expression relative to β-*lactamase*. Error bars represent standard error of the mean (S.E.M). (B) Western blots showing the levels of Caf1M and Caf1 (detected by anti-FLAG tag and anti-Caf1 antibodies respectively) in the cell pellets of the expression cultures are displayed underneath each graph, using DnaK probed with an anti-DnaK antibody as a loading control. +C and -C represent the pellets of *E. coli* BL21(DE3) cells transformed with pCOPF and untransformed respectively, and grown for 16 h at 35°C. NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). Biological replicates = 3. Statistical analysis was performed using One-Way ANOVA test.

Table 4-7. The RT-PCR data of the immediate thermal-responsive regulation by Caf1R. Showing the means of expression fold-change of the pCOPF and pCOPFΔR cultures grown either at 25°C for 16 h, at 25°C for 1 h or at 35°C for 1 h relative to β-*lactamase* as a house keeping gene. ± value is the standard error of the mean.

Gene transcript	The mean values of the expression fold-change					
	pCOPF cultures			pCOPFΔR cultures		
	At 25°C for 16 h	At 25°C for 1 h	At 35°C for 1 h	At 25°C for 16 h	At 25°C for 1 h	At 35°C for 1 h
caf1R	0.06 ± 0.006	0.06 ± 0.02	0.53 ± 0.3	0	0	0
caf1M	0.43 ± 0.1	0.31 ± 0.1	7.42 ± 1.4	0.06 ± 0.004	0.08 ± 0.004	0.08 ± 0.006
caf1A	0.5 ± 0.1	0.32 ± 0.09	10.84 ± 3	0.04 ± 0.003	0.04 ± 0.002	0.06 ± 0.004
caf1	7.41 ± 1	4.24 ± 0.9	113 ± 45	0.75 ± 0.08	1.22 ± 0.1	1.04 ± 0.08

Table 4-8. The statistics of the RT-PCR analysis data of the immediate thermal responsive transcriptional regulation by Caf1R. Showing the P-values of the comparison among the pCOPF and pCOPFΔR cultures grown either at 25oC for 16 h, at 25°C for 1 h or at 35°C for 1 h using One-Way ANOVA test.

Gene transcript	The statistical differences among the mean values (P-values)			
	pCOPF cultures grown at		pCOPFΔR cultures grown at	
	25°C for 16 h vs. 25°C for 1 h	25°C for 1 h vs. 35°C for 1 h	25°C for 16 h vs. 25°C for 1 h	25°C for 1 h vs. 35°C for 1 h
caf1R	1.00	0.12	1.00	1.00
caf1M	1.00	<0.001	0.99	1.00
caf1A	1.00	<0.001	1.00	1.00
caf1	1.00	0.003	1.00	0.99

There are statistically significant differences using One-Way ANOVA test between pCOPF cultures grown either at 25°C or at 35°C. The analysis of the

same cultures used in RT-PCR investigation by western blot (**Figure 4-8 B)** showed an obvious increase in Caf1M and Caf1 production at 35^{o}C as compared to 25^{o}C for pCOPF transformant and undetectable production of Caf1M and Caf1 was shown at both temperatures for pCOPFΔR transformants. The similarity in patterns of gene transcripts and protein levels implying that this regulation happens at the transcription not translation levels. Therefore, Caf1R is a thermo-responsive transcription factor which is responsible for the upregulation of *caf1* operon genes in response to immediate change in temperature.

4.2.7. Structural modelling predicts a model of DNA binding by Caf1R

Caf1R shows homology with the Rob, MarA and SoxS proteins which are non-canonical transcriptional activators from AraC family involved in the response to the superoxide and resistance to antibiotics and heavy metals (Gallegos *et al.*, 1997). The high sequence similarity of Caf1R's helix-turn-helix motifs with the non-canonical AraC transcription regulators supports the sequence alignment and structural modelling of Caf1R based on the crystal structure of the Rob protein (Hyock Joo *et al.*, 2000) and the MarA protein (Rhee *et al.*, 1998). The sequence alignment of Caf1R was performed using the Clustal Omega webserver (Sievers *et al.*, 2011) and visualised using the ESPript webserver (Robert and Gouet, 2014).

Caf1R protein sequence was used as an input using SWISS-MODEL Workspace webserver (Waterhouse *et al.*, 2018) to generate two structural models of Caf1R based on either Rob or MarA crystal structures. The degree of homology of the N-terminus in Caf1R model to the N-terminus of Rob and MarA in their crystal structures are shown in **Table 4-9**. The alignment results

(**Figure 4-9 A**) revealed that most of residues involved in DNA binding in both Rob and MarA are conserved in Caf1R's HTH motifs which are located at the N-terminus. 23 residues were fully conserved throughout these four proteins in the N-terminal domain.

Table 4-9. The homology of Caf1R N-terminus to Rob and MarA srystal structures. Showing the identity % and GMQE score, which is a number between (0-1), reflects the accuracy of the built model.

Protein	PDB code	Identity	GMQE
Rob	1D5Y	52.05	0.48
MarA	1BL0	35.82	0.41

The N-terminal HTH motif in Rob protein binds to the major groove of DNA (**Figure 4-9 B**) to form specific interactions with nitrogen bases, Trp-36 and Gln-39 contact DNA bases through van der Waals interaction, at the same time, Trp-36 is also hydrogen bonded to the DNA bases (Hyock Joo *et al.*, 2000). Arg-40 and Trp-36 in the N-terminal HTH motif of Rob is mutated to Ile-41 and Arg-37 in Caf1R model respectively and Gln-39 is conserved in Caf1R (Gln-40). On the other hand, the C-terminal HTH motif in Rob protein interacts with the DNA backbone using Arg-90 and Lys-94 which are conserved in Caf1R's N-terminal HTH motif (Arg-91 and Lys-95), these positively charged amino acids form non-specific hydrogen bonds with the phosphate group at the backbone (Hyock Joo *et al.*, 2000).

Additionally, the second model of Caf1R based on MarA crystal structure (**Figure 4-10**) demonstrated the homology of Caf1R with MarA in both N-terminal and C-terminal HTH motifs. The N-terminal HTH motif in MarA interacts the DNA backbone (**Figure 4-10 A**) through Gly-57 and Gln-58 by

forming hydrogen bonds with the phosphate group, whereas Trp-42, Gln-45 and Arg-46 can make contact with the nitrogen bases specifically, as Trp-42 and Gln-45 interact nitrogen bases through van der Waals forces and Arg-46 is hydrogen bonded to three DNA bases particularly (Rhee *et al.*, 1998).

Although two residues are conserved in Caf1R's N-terminal HTH motif which are Gln-45 and Gly-57 at positions 40 and 52 in Caf1R sequence respectively, the other MarA residues were substituted as follows: Trp-42 to Arg-37, Arg-46 to Ile-41 and Gln-58 to Thr-53. Unlike Rob's C-terminal HTH motif, MarA's one lies in the DNA major groove and contacts the nitrogen bases specifically (**Figure 4-10 B**) through Gln-91, Gln-92, Thr-93, Thr-95 and Arg-96 in helix 6. While, Lys-99 at the end of helix 6 and His-107 and Lys-108 in helix 7 contact the phosphate group of the DNA backbone. Interestingly, the C-terminal HTH motif of Caf1R has 6 conserved DNA-binding residues Gln-86, Gln-87, Thr-88, Thr-90, Arg-91 and Lys-99. The substituted residues in Caf1R HTH motif as compared to Rob and MarA indicate that Caf1R should have a different binding site.

The Caf1R C-terminus did not show any homology to the AraC family which implies significant differences in the regulation mechanism. This domain has been identified by InterPro webserver (Finn *et al.*, 2017) as a member of GyrI-like small molecule binding domain family, InterPro accession number is (IPR029442). The main feature of this superfamily is to have two strand-helix-strand-strand (SHS2) motifs opposite to each other and these motifs are involved in the binding of small molecule effectors (Anantharaman and Aravind, 2004) which implies that Caf1R function could be dependent on small molecule binding. Therefore, the effect of the Caf1R C-terminal domain

deletion was examined to determine the importance of C-terminus in Caf1 production.

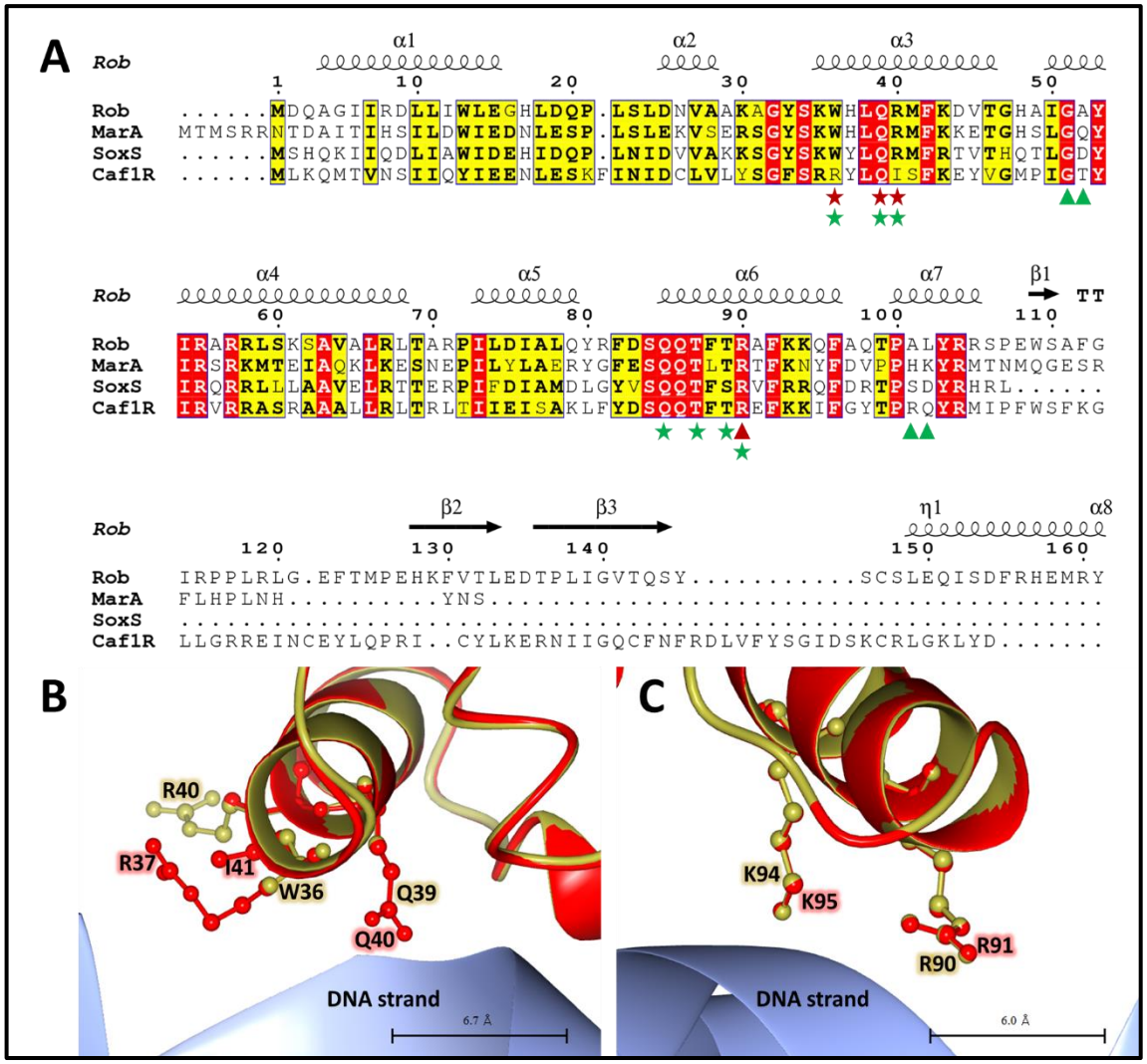

Figure 4-9. Sequence analysis and structural modelling of Caf1R based on Rob crystal structure. (A) Sequence alignment of Caf1R and homologous AraC family members. Residues from Rob that interact with the backbone of the DNA are highlighted with dark red triangles, while residues that make sequence specific contacts are highlighted with dark red stars. Residues from MarA that interact with the backbone of the DNA are highlighted green triangles, while residues that make sequence specific contacts are highlighted with green stars. Regions of identity are highlighted in red, and regions of similarity in yellow. The secondary structures of Rob are depicted above the alignment. (B and C) Superposition of a Rob:DNA complex (gold:light blue, PDB code: 1D5Y) (Hyock Joo *et al.*, 2000) with the homology model of Caf1R's 3D structure obtained from SWISS-MODEL Workspace (red) are shown with focus on the helix-turn-helix motif 1 (B) and 2 (C). Key DNA binding residues are shown as balls and sticks, with the same colours of their models, the residues of Rob are shown in gold and Caf1R in red.

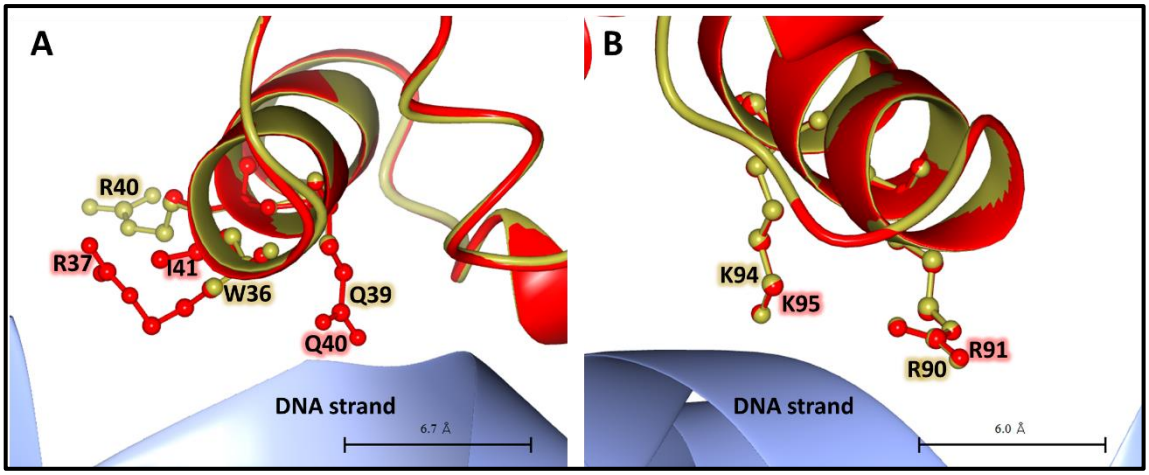

Figure 4-10. Structural modelling of Caf1R based on MarA crystal structure. Superposition of a MarA:DNA complex (silver:light blue, PDB code: 1BL0) (Rhee *et al.*, 1998) with the homology model of Caf1R's 3D structure obtained from SWISS-MODEL Workspace (red) are shown with focus on the helix-turn-helix motif 1 (A) and 2 (B). Key DNA binding residues are shown as balls and sticks, with the same colours of their models, the residues of MarA are shown in silver and Caf1R in red.

4.2.8. The Caf1R C-terminus is required for Caf1 polymer production

Karlyshev *et al.* (1992a) mentioned that C-terminus of Caf1R is not required for *caf1* expression and when bacterial cells were transformed with a plasmid containing only 81 residues from Caf1R's N-terminus with an additional tail of 19 amino acids, very large amounts of Caf1 could be produced. According to the structural modelling of Caf1R, it was found that these 81 residues from N-terminus do not include the second HTH DNA-binding motif which is important for DNA binding of both Rob and MarA as mentioned above. Hence, we decided to delete the C-terminus of Caf1R from pCOP plasmid (containing the whole *caf1* operon) by inserting three stop codons upstream to the codon of tyrosine 132 to generate pCOPΔRCT (containing the whole *caf1* operon with deleted Caf1R's C-terminus).

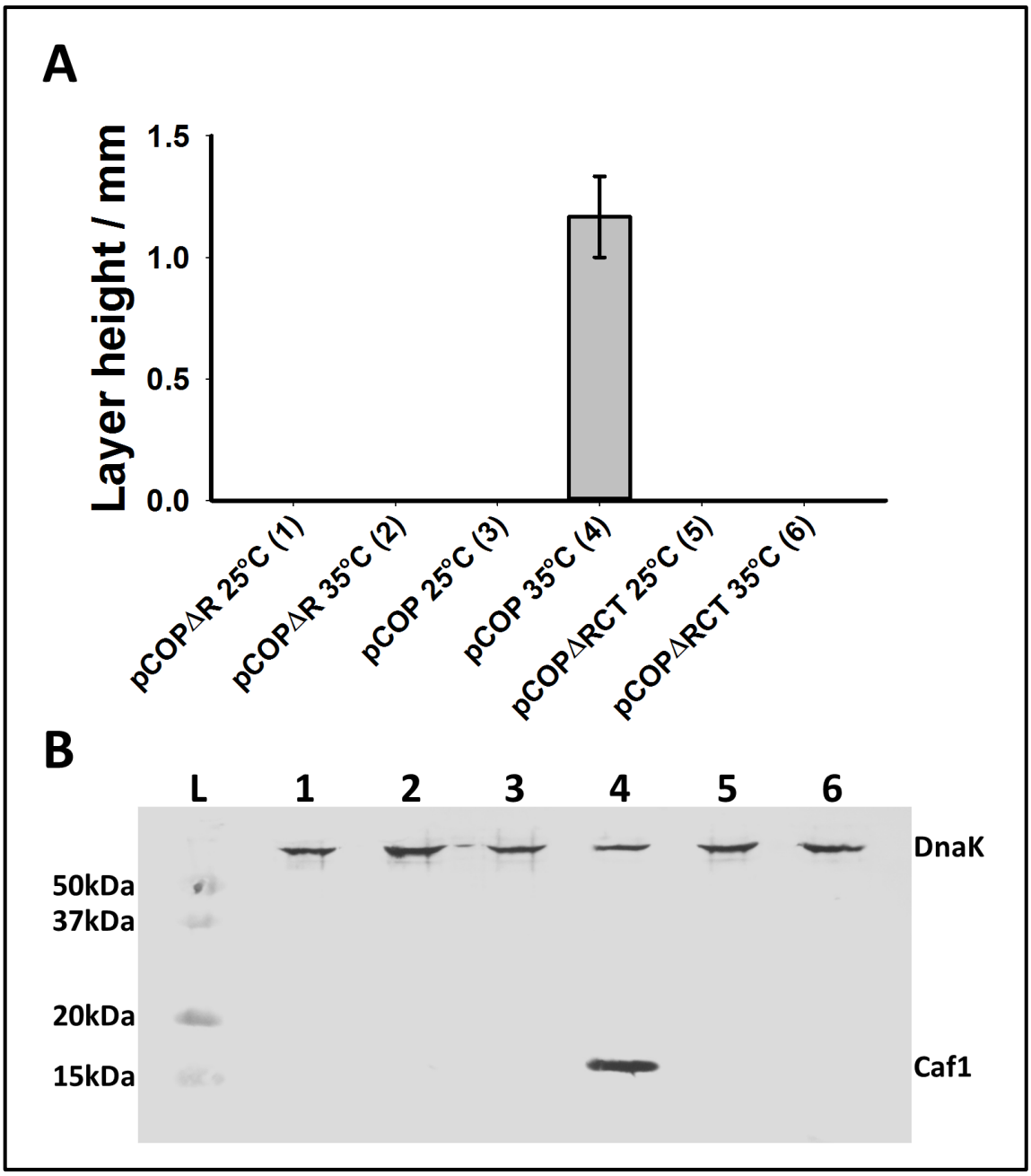

Figure 4-11. The Caf1R's C-terminus knock-out halts Caf1 polymer production. (A) Graph of flocculent height obtained for cultures of *E. coli* BL21 (DE3) transformed with either pCOPΔR (1) grown at 25°C and (2) at 35°C, pCOP (3) grown at 25°C and (4) at 35°C or pCOPΔRCT (5) grown at 25°C and (6) at 35°C all for 16 hours. Bar heights correspond to mean flocculent layer height obtained from three separate cultures from each individual sample, where error bars represent the standard error of the mean (S.E.M). (B) Western blot of the samples mentioned in (A) using anti-Caf1 and anti-DnaK antibodies showing Caf1 band at ≈ 15.5 kDa in the sample 4 only and DnaK bands at ≈ 70 kDa in all samples.

E. coli BL21 (DE3) were transformed with either pCOPΔR, pCOP or pCOPΔRCT, three single colonies from each transformant were grown in 5 ml liquid media either at 25^{o}C or 35^{o}C for 16 h, centrifuged in capillary tubes to measure the flocculent layer height for each culture and these cultures were analysed in western blotting using anti-Caf1 and anti-DnaK antibodies. Although the culture of pCOP at 35^{o}C produced Caf1 polymer (**Figure 4-11 A**), all other cultures at both 25^{o}C and 35^{o}C including pCOPΔR and pCOPΔRCT cultures did not produce any Caf1. Western blot results (**Figure 4-11 B**) were corresponding to the results stated above and shown in **Figure 4-11 A**, all samples revealed the presence of DnaK band at ≈ 70 kDa as a loading control, whereas, the Caf1 band was detected only in the culture of pCOP grown at 35^{o}C. Therefore, the pCOPΔRCT transformants behaved exactly like pCOPΔR transformants demonstrating that the C-terminus of Caf1R is required for Caf1 polymer production. There are two potential hypotheses could explain this action as follows: The C-terminus of Caf1R might be involved in its function directly as a positive transcriptional activator or it could be crucial for Caf1R protein folding.

4.3. Discussion

It was shown by this study that *caf1* operon is transcribed as a single long polycistronic transcript (4.2.1) depending on the Caf1R transcription factor. The presence of regulatory regions, a putative promoter and a Shine-Dalgarno sequence 7 bases upstream of the *caf1M* start codon were predicted by Galyov *et al.* (1991a). An independent software tools was used to identify potential promoters which may be responsible for the transcription of *caf1* operon as a single polycistronic transcript. The responsible promoter for this transcription is located in P2 region as shown in (4.2.2). The P2 region identified in this study overlaps with the putative one which was predicted by Galyov *et al.* (1991a). Therefore, it was shown that this region is responsible for *caf1* operon transcription through prediction and experimental confirmation.

Although *caf1M*, *caf1A* and *caf1* genes are transcribed as a single polycistronic unit, it was surprising that the levels *caf1M* and *caf1A* transcripts were similar and *caf1* transcript was much higher as shown by RT-PCR results. This difference could be assigned to two potential reasons: either the presence of a second promoter close to *caf1* gene under the regulation of Caf1R or the difference in the level of RNA stability of these three units in the polycistronic transcript. 5' RACE (Rapid Amplification of cDNA Ends) was used to search for any additional promoter upstream of the *caf1* gene in *intergenic region 3.* If the majority of *caf1* transcript sequences terminate at the transcriptional start site inside *intergenic region 3*, this means that there is a second promoter exists in this region. Corresponding to the detected polycistronic transcript, 7 out of 8 sequences obtained continued into *caf1A* through the *intergenic region 3* and stopped at

different locations. This indicates that the dissociation of DNA polymerase enzyme is the likely reason rather than a presence of a transcriptional start site in the sequence of the *caf1A* gene.

Therefore, the P2 region contains the promoter responsible for polycistronic transcription of *caf1* operon and its deletion abolished the *caf1* operon expression. The presence of a second promoter upstream of the *caf1* gene in *intergenic region 3* was predicted by Galyov *et al.* (1990). However, this region did not enable the expression of *caf1* in isolation (Galyov *et al.*, 1990) implying that it does not contain a functional promoter which agrees with the results of this study. So, the more convenient explanation for such difference in transcripts levels *caf1M* and *caf1A* and *caf1* observed in RT-PCR results is the variance in stability of translational units within the polycistronic mRNA which has been reported previously (Hui *et al.*, 2014). In other words, the individual RNA of the *caf1* operon genes in the polycistronic unit might have different levels of stability, for example, the half-life of the *caf1A* RNA seems to be less than that of the *caf1* RNA, so, the *caf1A* RNA will be degraded faster. Such a system would allow bacterial cells to produce optimal levels of each protein, Caf1M, Caf1A and Caf1, since, the polycistronic transcription of *caf1* operon is activated by Caf1R in response to temperature at 35^{o}C. Thus, the difference in stability of *caf1M*, *caf1A* and *caf1* mRNA leads to differential degradation of these transcripts to produce different amounts of translational units. Caf1 polymer production requires fewer amounts of chaperones and ushers as compared with Caf1 subunits.

When *Y. pestis* infects mainly two hosts with different body temperatures, ≈ 25^{o}C for flea (Perry and Fetherston, 1997) and ≈ 37^{o}C for mammals, it is important for it to have a thermo responsive transcription system. Such a

system will enable the bacterium to produce the required virulence factors for each environment. The rapid upregulation of *caf1M*, *caf1A* and *caf1* gene expression in response to temperature increase was clearly shown in the temperature switching experiment (4.2.6) which correlates well with the results of Han *et al.* (2004). Crucially, this kind of regulation occurs at the transcriptional level and completely depends on the Caf1R transcription activator.

Two other lines of evidence indicated that the temperature-sensitive regulation of *caf1* operon expression occurs at the transcriptional level. The first one is that the levels of the Caf1M and Caf1 proteins observed by western blot correlates with the levels of their transcripts detected by RT-PCR. Whereas, the introduction of a T7 promoter in order to bypass the thermo regulation permitted the *caf1* expression at 30^{o}C, which did not happen with the native *caf1* operon. Therefore, these could imply that either the translational and post-transcriptional regulation (e.g. RNA thermometers) are not present at all, having a very weak effect or this translational and post-transcriptional regulation could only be effective at very low temperatures (25^{o}C) as will be mentioned later in (5.3).

As shown in (4.2.7), Caf1R N-terminus showed a high sequence similarity to the DNA binding domains of both MarA and Rob (Rhee *et al.*, 1998; Hyock Joo *et al.*, 2000) suggesting that Caf1 R can bind DNA as a monomer like MarA and Rob with some differences in DNA binding mode as a result of W40R, R44I and H/A105R substitutions. The transcription factor of AraC family often contains another non-conserved domain involved in their function. This domain is not well-known in this family apart from AraC and XylS which use this domain in effector binding to modulate their function in response to

the presence of such effectors (Gallegos *et al.*, 1997; Soisson *et al.*, 1997; Kaldalu *et al.*, 2000). The Caf1R homolog, Rob uses its non-conserved C-terminal domain to be sequestered in distinct foci where it cannot induce transcription (Griffith *et al.*, 2009). When the Rob C-terminus binds to dipyridyl or decanoate the sequestration will be reversed to induce the transcriptional activator (Griffith *et al.*, 2009).

The deletion of the GyrI domain from Rob led it to lose the ability to sequester in the absence of dipyridyl or decanoate and activate the transcription without inducers (Ariza *et al.*, 1995; Griffith *et al.*, 2009). In this study, a truncated Caf1R protein containing only the first 132 amino acids which include both predicted helix-turn-helix DNA binding domains according to the model shown in (4.2.7) was seen to abolish the production of Caf1 implying the importance of the C-terminus in Caf1R function. The explanation of this negative effect could be attributed either to the direct involvement of C-terminus in Caf1R function or to the misfolding of Caf1R as a result of this deletion. On the contrary, Karlyshev *et al.* (1992b) reported that the using of N-terminus of Caf1R which contains only 82 amino acids led to increase the expression of *caf1* operon. As the Caf1R C-terminus was identified as a member of GyrI-like small molecule binding domain family using InterPro webserver (Finn *et al.*, 2017), this could then outweigh the hypothesis of the C-terminus significance in Caf1R function.

The role of Caf1R in the immunoevasive ability of such highly dangerous bacteria make it of interest. It was revealed by PSI-BLAST searches that many unannotated sequences from various Gram-negative were similar proteins to Caf1R. One annotated example is LdaA, the putative regulatory protein which is part of the *lda* locus (locus of diffuse adherence) in

enteropathogenic *E. coli* (EPEC) which functions to attenuate the adherence of *E. coli* to HEp-2 cells (Scaletsky *et al.*, 2005). In addition, Cantey *et al.* (1999) identified that AfrR, a putative transcriptional regulator in the AF/R1 operon showed high degree of homology with Caf1R. The AF/R1 operon codes for a pilus in a rabbit diarrheagenic *E. coli* (RDEC) which is responsible for its adhesion and pathogenicity (Cantey *et al.*, 1999). If these proteins certainly regulate the transcription in a similar way to Caf1R, it may hint to the new conserved thermo responsive transcriptional activation system in Gram-negative bacteria.

Chapter Five: Using the Caf1R transcription factor to create a novel thermo-responsive expression system

5. Chapter Five: Using the Caf1R transcription factor to create a novel thermo-responsive expression system

5.1. Introduction

The natural biological systems have been exploited to generate useful products in biotechnology. For example, the *lac* regulation system which regulates the lactose metabolism in *E. coli* (Kennell and Riezman, 1977) is one of the most common systems used in synthetic expression vectors, the expression in this system can be induced using lactose or its chemical analog, Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Kennell and Riezman, 1977; Donovan *et al.*, 1996; Mads Kaern *et al.*, 2003; Marbach and Bettenbrock, 2012). The *araBAD* operon in *E. coli*, which is responsible for L-arabinose uptake by the bacterium as an alternative carbon and energy source (Schleif, 2000) was utilized as an artificial inducible expression system (Guzman *et al.*, 1995; Siegele and Hu, 1997). The gene expression by *araBAD* promoter-based vectors can be induced by adding L-arabinose to the culture media (Guzman *et al.*, 1995; Siegele and Hu, 1997).

Recently, a new temperature-sensitive expression system has been developed to produce a distinct protein in response to different temperature ranges selectively (Lermant *et al.*, 2018). This system was built up based on the *E. coli* Cold Shock Protein A (CspA) as a cold-responsive system (Gualerzi *et al.*, 2003; Lermant *et al.*, 2018) and the λcl repressor from lambda phage and its regulator site (*pL*) as a heat-responsive system (Dodd and Egan, 2002; Lermant *et al.*, 2018). Thermo-responsive regulation occurs at different levels which might be transcriptional or translational (Shapiro and Cowen, 2012) and regulation of this expression system happens at both transcriptional and translational levels (Lermant *et al.*, 2018). Generally,

temperature-sensitive regulation can be mediated by DNA thermosensors (López-García and Forterre, 2000), RNA thermometers (Jens and Franz, 2012) or by thermo-responsive transcription activators. The temperature-sensitive transcriptional activation of *caf1* operon gene expression mediated by the Caf1R transcription factor has already been observed in this study.

In this chapter, the temperature-sensitive Caf1R transcription activator was exploited to create a novel thermo-responsive expression system to be used in biotechnology. The plasmid pR-GFP (which is composed of pGEM-T vector backbone (without T7 promoter), *caf1R* gene, *intergenic region 1* from *caf1* operon, ribosome binding site of *caf1M* and *gfp* gene) was constructed to examine the thermo-responsive expression of *gfp* under the regulation of Caf1R. The GFP synthesis rate per cell was zero at 25^{o}C and small amounts of GFP were produced at 35^{o}C. Several development steps were achieved in order to enhance the expression levels at 35^{o}C. Interestingly, the insertion of a strong constitutive promoter downstream to the P2 (the promoter of *caf1* operon) in the *intergenic region 1* (pRC2-GFP) did not increase the GFP production at 25^{o}C which might indicate that there is a post-transcriptional regulation element such as RNA thermometers. Whereas, the presence of this constitutive promoter resulted in an obvious increase in GFP production at 35^{o}C.

The *gfp* expression was further improved at 35^{o}C by insertion of *intergenic region 3* from the *caf1* operon (the region which separates *caf1A* and *caf1* genes) downstream of *intergenic region 1* containing the mentioned constitutive promoter (pRC3-GFP vector). The expression at 25^{o}C was still low but higher than that obtained from pRC2-GFP cultures. The temperature-sensitive induction of *gfp* expression by pRC3-GFP vector was

confirmed again at four different time points. *caf1R* knock-out from pRC3-GFP vector led to large reduction in GFP production. In addition, some issues in this study were addressed in this chapter including the effect of CT-FLAG-tags on the Caf1 polymer production. It was found that the CT-FLAG-tag in Caf1A has a significant negative effect on Caf1 production and it was not detected by western blot using an anti-FLAG antibody. The replacement of FLAG-tag position to a flexible loop in β-barrel in Caf1A restored the amounts of produced Caf1 polymer to wild type construct levels and the new FLAG-tagged Caf1A has been detected by western blot.

5.2. Results

5.2.1. The thermal induction of green fluorescence protein (*gfp*) gene expression

The previous results shown in this study (2.2.5) proved that Caf1R, the transcriptional regulator is responsible for the activation of *caf1* operon transcription in response to temperature change. The transcriptional regulation by Caf1R in response to the temperature change, suggested that a novel expression system which can be induced by switching temperature from 25^{o}C to 35^{o}C instead of using a chemical inducer like IPTG or L-arabinose could be developed. Two plasmid constructs pR-GFP and pR-T7 GFP **(Figure 5-1 A)** were examined for the efficiency of thermal-sensitive induction of *gfp* gene expression.

E. coli BL21 (DE3) cells were transformed with either pGEM-T vector as a negative control (which does not produce GFP), pET-GFP as a positive control for GFP production (which is composed of pET28A vector backbone, strong constitutive promoter and *gfp* gene), pR-GFP (which is composed of pGEM-T vector backbone without T7 promoter, *caf1R* gene, *intergenic region 1* from *caf1* operon, ribosome binding site of *caf1M* and *gfp* gene), pR-T7 GFP (which is composed pGEM-T vector backbone, T7 promoter, *caf1R* gene, *intergenic region 1* from *caf1* operon, ribosome binding site of *caf1M* and *gfp* gene) as shown in **Figure 5-1 A**. The cultures were incubated in a plate reader (n=9, three biological and three technical replicates for each culture) at either 25^{o}C or 35^{o}C for 16 h using 200 µl inoculated TB media with appropriate antibiotics in 96 well plates, both OD_{600} and fluorescence were measured every 10 minutes.

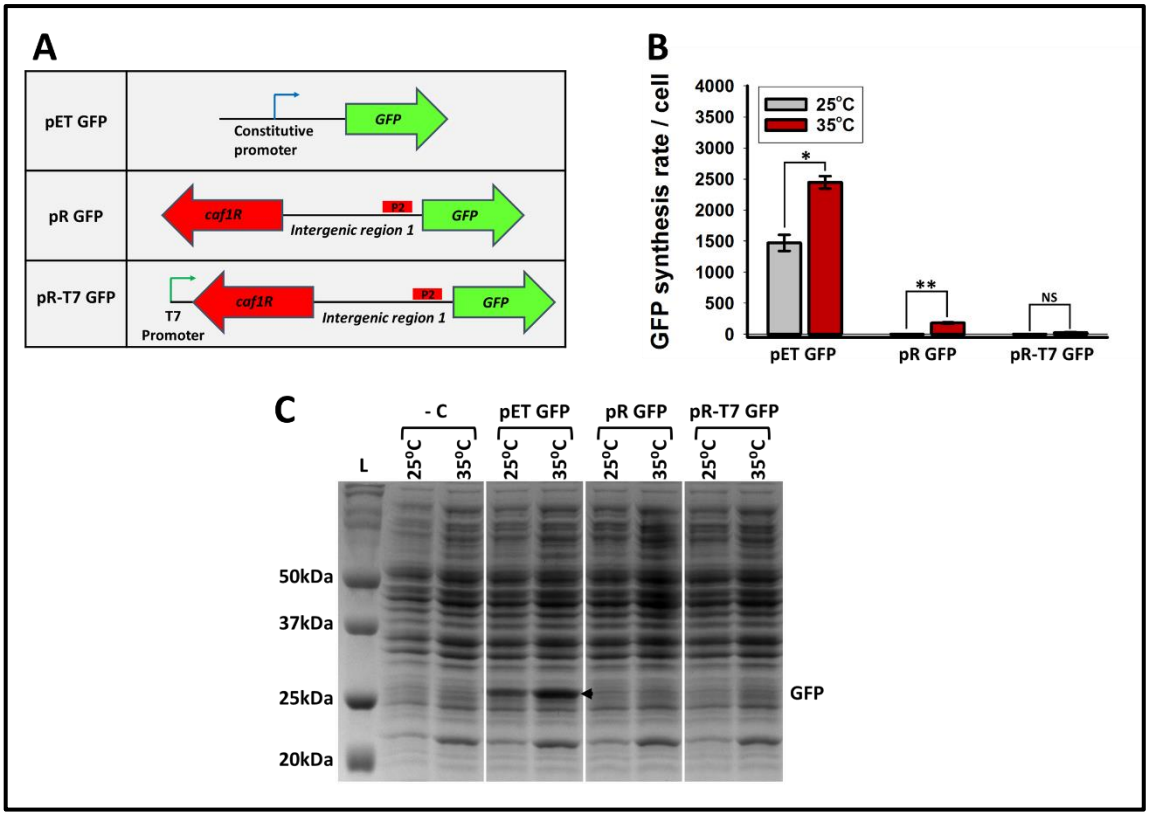

Figure 5-1. The temperature-sensitive induction of *gfp* expression. (A) Diagrams of the plasmids constructs used. pET GFP is composed of pET28A vector backbone, insulated strong constitutive promoter, medium strength ribosome binding site and *gfp* gene. pR GFP is composed of pGEM-T vector backbone without T7 promoter, *caf1R* gene, Intergenic region 1 from *caf1* operon, ribosome binding site of *caf1M* and *gfp* gene. pR-T7 GFP is composed of pGEM-T vector backbone with T7 promoter, *caf1R* gene, intergenic region 1 from *caf1* operon, ribosome binding site of *caf1M* and *gfp* gene. P2 is the predicted *caf1* operon promoter. The blue bent arrow represents insulated constitutive promoter. (B) Graph of GFP production in response to thermal change for three different cultures of *E. coli* BL21 (DE3) transformed with either pET GFP, pR GFP or pR-T7 GFP. The liquid cultures were grown for 16 hours at either 25°C or 35°C separately. The GFP synthesis rate were measured between two time points (14 and 16 hours) as mentioned in Materials and Methods. Bar heights correspond to GFP synthesis rate. Error bars represent standard error of the mean (S.E.M). NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). (C) SDS-PAGE analysis of the same cultures stated in (A) showing the GFP bands at ≈ 27 kDa in the lanes of pET GFP transformant cultures at both temperatures 25°C and 35°C with less intensity at 25°C. The negative control (-C) is *E. coli* BL21 (DE3) transformed with pGEM-T vector. Biological replicates = 3. Statistical analysis was performed using T-test.

Two time points were used to calculate the GFP synthesis rate during 2 h (between 14 h and 16 h) for each culture at both temperatures 25°C and 35°C subtracted from the GFP synthesis rate of the negative control (which is *E. coli* BL21 (DE3) cells transformed with pGEM-T vector) as mentioned in methodology section. These cultures were also analysed by SDS-PAGE to examine the production of GFP in each culture. The mean values of GFP synthesis rate per cell after 16 h growth of the stated cultures are shown in **Table 5-1.**

Table 5-1. The mean values of GFP synthesis rate per cell. Produced by E. coli BL21 (DE3) transformed with either pET-GFP, pR-GFP or pR-T7 GFP grown at 25°C and 35°C separately. ± value is the standard error of the mean.

***E. coli* BL21 (DE3) transformed with**	**The mean values of GFP synthesis rate per cell above that of negative control**	
	at 25°C	**at 35°C**
pET-GFP	1471 ± 129	2447 ± 99
pR GFP	0	182 ± 9
pR-T7 GFP	0	25 ± 12

Although pET-GFP transformants cultures produced GFP at both temperatures 25°C and 35°C, the GFP synthesis rate per cell in pR-GFP and pR-T7 GFP transformants cultures were zero at 25°C **(Figure 5-1)**. Based on the *caf1R* expression system, pR-GFP transformants cultures produced low amounts of GFP as compared to that produced by the cultures of pET-GFP transformants. The statistical analysis using T-test showed there is a significant difference between pR-GFP transformants cultures grown at 25°C and at 35°C as shown in **Table 5-2**.

Table 5-2. The P-values of statistical comparisons between the grown cultures of pET-GFP, pR-GFP and pR-T7 GFP at two different temperatures 25°C and 35°C using T-test.

The comparison between cultures of	P-values
pET-GFP at 25°C vs. pET-GFP at 35°C	0.003
pR-GFP at 25°C vs. pR-GFP at 35°C	<0.001
pR-T7 GFP at 25°C vs. pR-T7 GFP at 35°C	0.1

The SDS-PAGE results **(Figure 5-1 C)** showed that pET-GFP transformants cultures produced obvious GFP bands of ≈ 27 kDa at both 25°C and 35°C. Whereas, the cultures of pR-GFP and pR-T7 GFP transformants did not produce a clear GFP bands at either 25°C or 35°C. The SDS-PAGE results thus match the GFP synthesis rates. The GFP produced by pR-GFP transformants culture at 35°C was not detected in SDS-PAGE. Although the levels of *gfp* expression by pR-GFP were low at 35°C, it showed promising results to start with developing this expression system for biotechnology applications.

5.2.2. The insertion of a strong insulated constitutive promoter enhances the *gfp* expression at 35°C

The molecular mechanism of thermal responsive transcription of *caf1* operon which is driven by the Caf1R transcription activator is not well-known yet. To examine whether this temperature-sensitive regulation happens only at transcriptional level as in section (4.2.6) or also happens at post-transcriptional level as well, since, temperature-sensitive regulation has been seen in *Yersinia pseudotuberculosis* and RNA structuromes were responsible for such regulation (Righetti *et al.*, 2016).

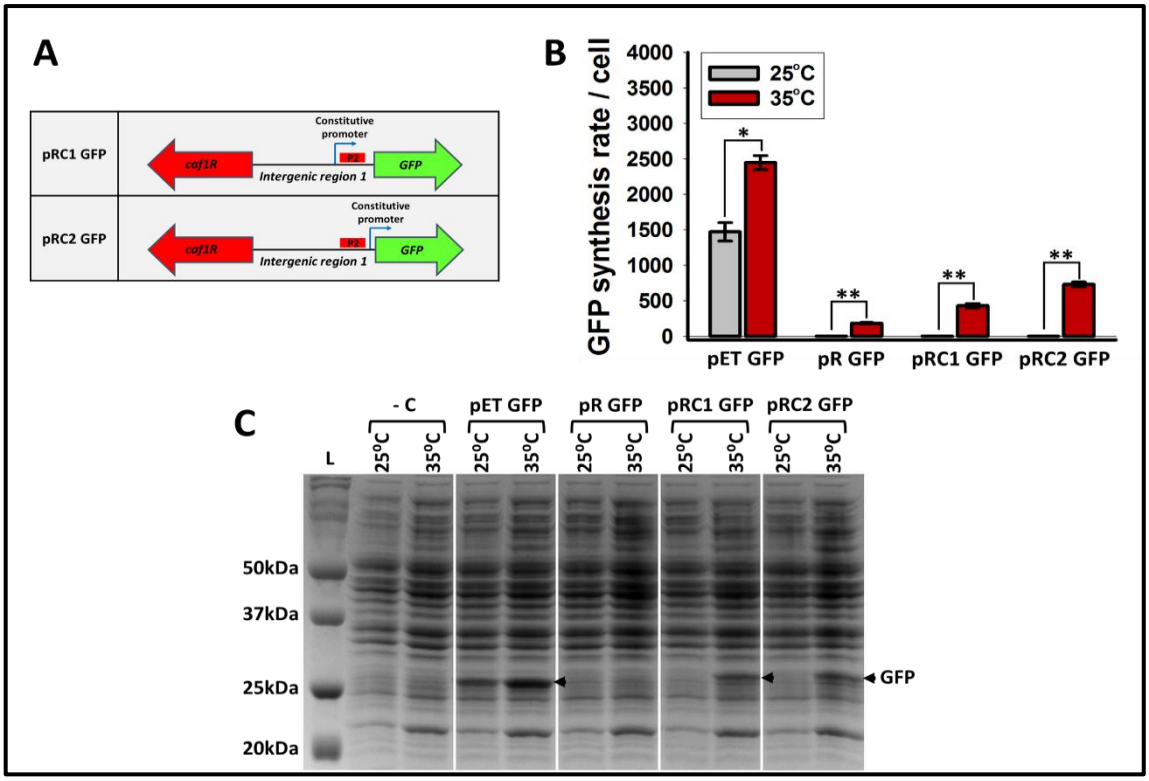

Figure 5-2. The effect of insulated constitutive promoter insertion on the thermal-responsive induction system. (A) Diagrams of the plasmids constructs used in this experiment. pRC1-GFP is composed of pGEM-T vector backbone without T7 promoter, *caf1R* gene, Intergenic region 1 from *caf1* operon, insulated strong constitutive promoter upstream to the predicted *caf1* operon promoter, ribosome binding site of *caf1M* and *gfp* gene. pRC2-GFP is composed of pGEM-T vector backbone without T7 promoter, *caf1R* gene, Intergenic region 1 from *caf1* operon, insulated strong constitutive promoter downstream to the predicted *caf1* operon promoter, ribosome binding site of *caf1M* and *gfp* gene. P2 is the predicted *caf1* operon promoter. The blue bent arrow represents the insulated constitutive promoter (B) Graph of GFP production in response to thermal change for four different cultures of *E. coli* BL21 (DE3) transformed with either pET-GFP, pR-GFP, pRC1-GFP or pRC2-GFP. The liquid cultures were grown for 16 hours at either 25°C or 35°C separately. The GFP synthesis rates were measured between two time points (14 and 16 hours) as mentioned in Materials and Methods. Bar heights correspond to GFP synthesis rate. Error bars represent standard error of the mean (S.E.M). NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). (C) SDS-PAGE analysis of the same cultures stated in (A) showing the GFP bands at ≈ 27 kDa in the lanes of pET-GFP transformant cultures at both temperatures 25°C and 35°C with less intensity at 25°C. Both pRC1-GFP and pRC2-GFP transformant cultures produce GFP bands at 35°C not at 25°C. The negative control (-C) is *E. coli* BL21 (DE3) transformed with pGEM-T vector. Biological replicates = 3. Statistical analysis was performed using T-test.

Therefore, the insertion of an insulated strong constitutive promoter used in pET-GFP which is designed in a way that give it an insulation from any stimulatory or repressive effects of 3'- or 5'- sequence elements (Davis *et al.*, 2011) in *intergenic region 1* **(Figure 5-1 A)** is a preliminary examination to show if there is a post-transcriptional thermo-responsive regulation. If there is no post-transcriptional regulation, the presence of strong constitutive promoter will activate the transcription at both temperatures 25^{o}C and 35^{o}C.

E. coli BL21 (DE3) cells were transformed with either pGEM-T vector as a negative control (which does not produce GFP), pET-GFP as a positive control for GFP production, pR-GFP, pRC1-GFP (which is composed of pGEM-T vector backbone without T7 promoter, *caf1R* gene, *Intergenic region 1* from *caf1* operon, insulated strong constitutive promoter upstream to the predicted *caf1* operon promoter, ribosome binding site of *caf1M* and *gfp* gene) **(Figure 5-2 A)** or pRC2-GFP (which is composed of pGEM-T vector backbone without T7 promoter, *caf1R* gene, *Intergenic region 1* from *caf1* operon, insulated strong constitutive promoter downstream to the predicted *caf1* operon promoter, ribosome binding site of *caf1M* and *gfp* gene) **(Figure 5-2 A)**.

The cultures were incubated in a plate reader at either 25^{o}C or 35^{o}C for 16 h using 200 µl inoculated TB media with appropriate antibiotics in a 96 well plate (n=9, three biological and three technical replicates for each culture). Both OD_{600} and fluorescence were measured every 10 minutes. Two time points were used to calculate the GFP synthesis rate during 2 h (between 14 h and 16 h) for each culture at both temperatures 25^{o}C and 35^{o}C **(Figure 5-2 B)**, the value for the negative control was subtracted from the GFP synthesis rate (where the negative control is *E. coli* BL21 (DE3) cells transformed with pGEM-T vector) as mentioned in methodology section. These cultures were

analysed by SDS-PAGE **(Figure 5-2 C)** to examine the production of GFP in each culture. The mean values of GFP synthesis rate per cell after 16 h growth of these cultures are shown in **Table 5-3**. The statistics showed significant differences among the cultures grown at 25^{o}C and 35^{o}C as shown in **Table 5-4**.

The *gfp* expression both cultures of pRC1-GFP and pRC2-GFP at 35^{o}C as a result of the insulated constitutive promoter insertion **(Figure 5-2)**. The GFP production was higher in pRC2-GFP transformants cultures as compared to cultures of pRC1-GFP. The location of the constitutive promoter could be the reason of such difference between pRC1-GFP and pRC2-GFP transformants cultures. Since, the constitutive promoter in pRC1-GFP is located upstream to the predicted *caf1* operon promoter (P2), where Caf1R might bind to regulate the transcription. Consequently, the bound Caf1R in pRC1-GFP system at could block the way of the RNA polymerase recruited by the constitutive promoter.

Table 5-3. The mean values of GFP synthesis rate per cell. Produced by *E. coli* BL21 (DE3) transformed with either pET-GFP, pR GFP, pRC1-GFP or pRC2-GFP grown at 25^{o}C and 35^{o}C separately. ± value is the standard error of the mean.

***E. coli* BL21 (DE3) transformed with**	**The mean values of GFP synthesis rate per cell above that of negative control**	
	at 25^{o}C	**at 35^{o}C**
pET-GFP	1471 ± 129	2447 ± 99
pR GFP	0	182 ± 9
pRC1-GFP	0	430 ± 31
pRC2-GFP	0	730 ± 33

Table 5-4. The P-values of statistical comparisons between the grown cultures of pET-GFP, pR-GFP, pRC1-GFP and pRC2-GFP at two different temperatures 25°C and 35°C using T-test.

The comparison between cultures of	P-values
pET-GFP at 25°C vs. pET-GFP at 35°C	0.003
pR-GFP at 25°C vs. pR-GFP at 35°C	<0.001
pRC1-GFP at 25°C vs. pRC1-GFP at 35°C	<0.001
pRC2-GFP at 25°C vs. pRC2-GFP at 35°C	<0.001

Interestingly, the *gfp* expression in the cultures of pRC1-GFP and pRC2-GFP transformants were zero at 25°C, the presence a strong constitutive promoter does not increase the GFP production. This could indicate that there is a post-transcriptional regulation, which could be due to an RNA thermometer. The SDS-PGAE results **(Figure 5-2 C)** confirmed the presence of GFP band at ≈ 27 kDa in the cultures of pRC1-GFP and pRC2-GFP at 35°C but not at 25°C and that matches the results of the GFP synthesis rate. Although the insertion of the constitutive promoter led to increase the *gfp* expression, the levels of GFP are still low as compared to the cultures of pET-GFP.

5.2.3. The insertion of *intergenic region 3* from *caf1* operon improve the GFP production at 35°C

The *intergenic region 3* is an 80 nt long sequence between *caf1A* and *caf1* genes in the *caf1* operon which includes the RBS of *caf1*. It was suggested previously in this study that there is no potential functional promoter in *intergenic region 3* as confirmed by RACE experiment in (4.2.4) and reported

by Galyov *et al.* (1990). The main aim of the insertion of this sequence is to examine the effect of using the *caf1* RBS instead of the *caf1M* RBS. Therefore, this sequence was inserted upstream to *intergenic region 1* in pRC2-GFP with deleted *caf1M* RBS to generate a new plasmid called pRC3-GFP **(Figure 5-3 A)**.

E. coli BL21 (DE3) cells were transformed with either pGEM-T vector as a negative control, pET-GFP as a positive control for GFP production, pRC2-GFP or pRC3-GFP (which is composed of pGEM-T vector backbone, *caf1R* gene, *intergenic region 1* from *caf1* operon, insulated strong constitutive promoter downstream of the predicted *caf1* operon promoter, *intergenic region 3* from *caf1* operon, ribosome binding site of *caf1* and *gfp* gene). The cultures were incubated in plate reader at either 25^{o}C or 35^{o}C for 16 h using 200 µl inoculated TB media with appropriate antibiotics in 96 well plate (n=9, three biological and three technical replicates for each culture). Both OD_{600} and fluorescence were measured every 10 minutes.

Graphs of GFP synthesis rate per cell and SDS-PAGE results **(Figure 5-3)** showed that the insertion of *intergenic region 3* increases the production of GFP at 35^{o}C with a little leaky expression at 25^{o}C. The mean values of the GFP synthesis rate per cell subtracted from the GFP synthesis rate of the negative control (which is *E. coli* BL21 (DE3) transformed with pGEM-T vector) are shown in **Table 5-5** and statistical analysis results using T-test are shown in **Table 5-6**. Further examinations are required to determine the mechanism of such regulation achieved by this expression system.

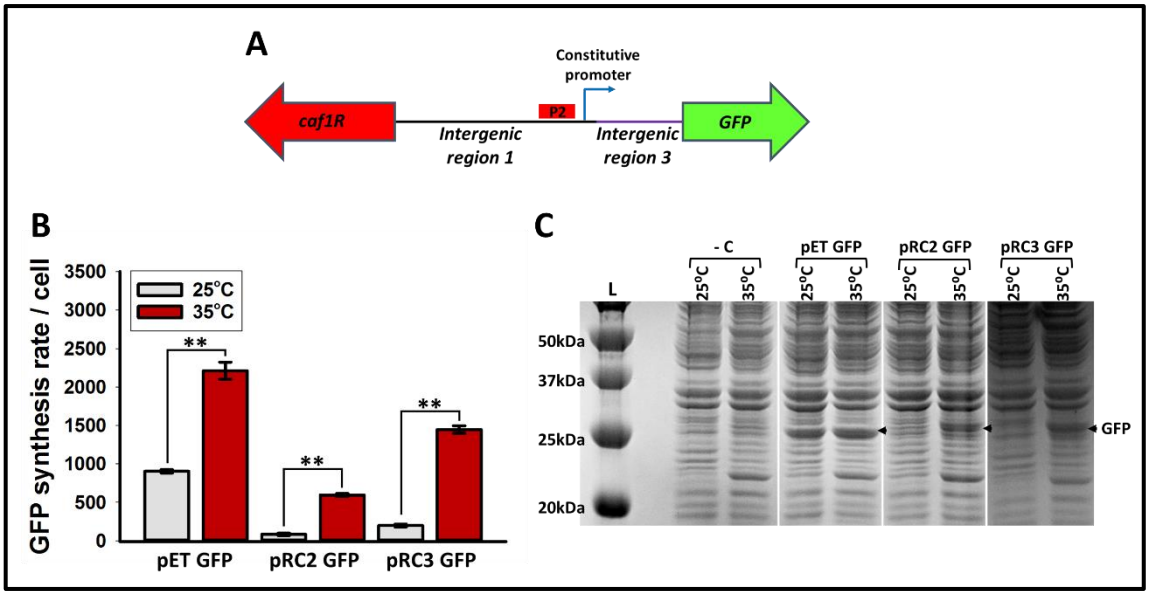

Figure 5-3. The effect of *intergenic sequence 3* insertion (which is the intergenic region between *caf1A* and *caf1* in *caf1* operon) on the thermal-responsive induction system. (A) A diagram of pRC3 GFP which is composed of pGEM-T vector backbone, *caf1R gene*, *intergenic region 1* from *caf1* operon, insulated strong constitutive promoter downstream to the predicted *caf1* operon promoter, *intergenic region 3* from *caf1* operon, ribosome binding site of *caf1* and *gfp* gene. (B) Graph of GFP production in response to thermal change for three different cultures of *E. coli* BL21 (DE3) transformed with either pET GFP, pRC2 GFP or pRC3 GFP. The liquid cultures were grown for 16 hours at either 25°C or 35°C separately. The GFP synthesis rates were measured between two time points (14 and 16 hours) as mentioned in materials and methods. Bar heights correspond to GFP synthesis rate. Error bars represent standard error of the mean (S.E.M). NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). (C) SDS-PAGE analysis of the same cultures stated in (A) showing the GFP bands at ≈ 27 kDa in the lanes of pET GFP transformant cultures at both temperatures 25°C and 35°C with less intensity at 25°C. Both pRC2 GFP and pRC3 GFP transformant cultures produce GFP bands at 35°C not at 25°C. The negative control (-C) is culture of *E. coli* BL21 (DE3) transformed with pGEM-T vector. Biological replicates = 3. Statistical analysis was performed using T-test.

Table 5-5. The mean values of GFP synthesis rate per cell. Produced by *E. coli* BL21 (DE3) transformed with either pET-GFP, pRC2-GFP or pRC3-GFP grown at 25°C and 35°C separately. ± value is the standard error of the mean.

E. coli BL21 (DE3) transformed with	The mean values of GFP synthesis rate per cell above that of negative control	
	at 25°C	at 35°C
pET-GFP	903 ± 20	2212 ± 110
pRC2-GFP	84 ± 14	595 ± 19
pRC3-GFP	197 ± 18	1446 ± 49

Table 5-6. The P-values of statistical comparisons between the grown cultures of pET-GFP, pRC2-GFP and pRC3-GFP at two different temperatures 25°C and 35°C using T-test.

The comparison between cultures of	P-values
pET-GFP at 25°C vs. pET-GFP at 35°C	<0.001
pRC2-GFP at 25°C vs. pRC2-GFP at 35°C	<0.001
pRC3-GFP at 25°C vs. pRC3-GFP at 35°C	<0.001

5.2.4. The thermo-responsive *gfp* expression by pRC3-GFP vector during different time points

In order to confirm the ability of pRC3-GFP vector to express *gfp* in response to temperature, the GFP synthesis rate per cell of the *gfp* expressing cultures were calculated in different time points. *E. coli* BL21 (DE3) cells were transformed with either pET-GFP, pRC2-GFP or pRC2-GFP, grown in plate reader for 16 h at either 25°C or 35°C using 200 µl of inoculated TB media with appropriate antibiotics in the 96 well plate (n=9, three biological and three technical replicates for each culture). OD_{600} and fluorescence were measured every 10 minutes. Eight time points were used to calculate the

GFP synthesis rate of 2 h (between 4 h and 6 h, 6 h and 8 h, 8 h and 10 h and 10 h and 12 h) for each culture at both temperatures 25°C and 35°C subtracted from the GFP synthesis rate of the negative control. The mean values of GFP synthesis rate per cell for these culture in different time points subtracted from the mean value of the negative control are shown in **Table 5-7**. The statistics were performed using T-test as shown in **Table 5-8**.

Table 5-7. The mean values of GFP synthesis rate per cell. Produced by *E. coli* BL21 (DE3) transformed with either pET-GFP, pRC2-GFP or pRC3-GFP grown at 25°C and 35°C separately at different time points. ± value is the standard error of the mean.

***E. coli* BL21 (DE3) transformed with**	**The mean values of GFP synthesis rate per cell above that of negative control**	
	at 25°C	**at 35°C**
pET-GFP (6 h)	5362 ± 259	6553 ± 1294
pRC2-GFP (6 h)	15 ± 8	260 ± 29
pRC3-GFP (6 h)	188 ± 49	550 ± 81
pET-GFP (8 h)	2832 ± 239	3259 ± 244
pRC2-GFP (8 h)	23 ± 2	516 ± 5
pRC3-GFP (8 h)	89 ± 26	1172 ± 102
pET-GFP (10 h)	1306 ± 126	2996 ± 116
pRC2-GFP (10 h)	58 ± 3	628 ± 12
pRC3-GFP (10 h)	104 ± 19	1333 ± 84
pET-GFP (12 h)	686 ± 147	2861 ± 101
pRC2-GFP (12 h)	82 ± 15	610 ± 14
pRC3-GFP (12 h)	132 ± 17	1423 ± 72

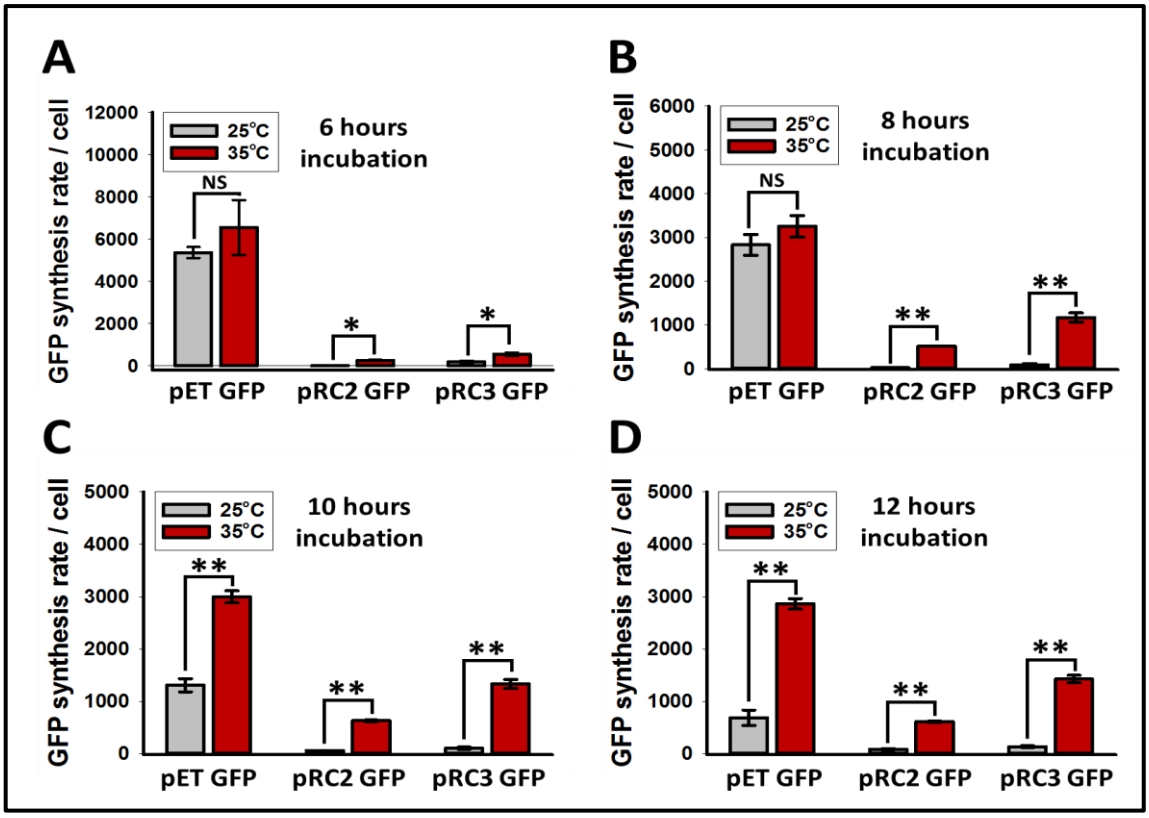

Figure 5-4. The effect of intergenic sequence 3 insertion from *caf1* operon on the thermal-responsive induction system in different time points. Showing the graphs of GFP production in response to thermal change for three different cultures of *E. coli* BL21 (DE3) transformed with either pET GFP, pRC2 GFP or pRC3 GFP. (A) The liquid cultures were grown for 6 hours at either 25°C or 35°C separately. The GFP synthesis rate were measured between two time points (4 and 6 hours). (B) The liquid cultures were grown for 8 hours at either 25°C or 35°C separately. The GFP synthesis rate were measured between two time points (6 and 8 hours). (C) The liquid cultures were grown for 10 hours at either 25°C or 35°C separately. The GFP synthesis rate were measured between two time points (8 and 10 hours). (D) The liquid cultures were grown for 12 hours at either 25°C or 35°C separately. The GFP synthesis rate were measured between two time points (10 and 12 hours). Bar heights correspond to GFP synthesis rate per cell subtracted from the negative control (*E. coli* BL21 (DE3) cells transformed with pGEM-T vector). Error bars represent standard error of the mean (S.E.M). NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). Biological replicates = 3. Statistical analysis was performed using T-test.

Table 5-8. The P-values of statistical comparisons between the grown cultures of pET-GFP, pRC2-GFP and pRC3-GFP at two different temperatures 25°C and 35°C in different time points using T-test.

The comparison between cultures of	P-values
pET-GFP at 25°C vs. pET-GFP at 35°C (6 h)	0.4
pRC2-GFP at 25°C vs. pRC2-GFP at 35°C (6 h)	0.001
pRC3-GFP at 25°C vs. pRC3-GFP at 35°C (6 h)	0.02
pET-GFP at 25°C vs. pET-GFP at 35°C (8 h)	0.28
pRC2-GFP at 25°C vs. pRC2-GFP at 35°C (8 h)	<0.001
pRC3-GFP at 25°C vs. pRC3-GFP at 35°C (8 h)	<0.001
pET-GFP at 25°C vs. pET-GFP at 35°C (10 h)	<0.001
pRC2-GFP at 25°C vs. pRC2-GFP at 35°C (10 h)	<0.001
pRC3-GFP at 25°C vs. pRC3-GFP at 35°C (10 h)	<0.001
pET-GFP at 25°C vs. pET-GFP at 35°C (12 h)	<0.001
pRC2-GFP at 25°C vs. pRC2-GFP at 35°C (12 h)	<0.001
pRC3-GFP at 25°C vs. pRC3-GFP at 35°C (12 h)	<0.001

The mean values of the GFP synthesis rate per cell of the cultures of pRC3-GFP transformants were higher than these of the cultures of pRC2-GFP at all different time point **(Figure 5-4)**. The GFP synthesis rate has increased in pRC3-GFP transformants cultures by approximately two times as compared to pRC2-GFP transformants cultures at 35°C in all cases. Although the *gfp* expression has increased at 25°C in the cultures of pRC3-GFP as compared with pRC2-GFP transformants cultures, the levels of expression are still low. The mean values of the GFP synthesis rate per cell of the pRC3-GFP transformants cultures grown at 35°C were increased by 2.9 times after 6 h incubation, 13.1 times after 8 h incubation, 12.8 times after 10 h incubation and 10.7 times after 12 h incubation as compared with these obtained at

25°C. From previous results, it was shown that the presence of the *caf1R* gene led to the activation of the *gfp* expression more efficiently at 35°C with a little leaky expression at 25°C. However, further examinations are needed to confirm the role of Caf1R in the temperature-sensitive *gfp* expression by the pRC3-GFP system.

5.2.5. *caf1R* knock-out from pRC3-GFP results in a large reduction in *gfp* expression

According to our previous results in this study, the thermo-responsive expression of *gfp* is supposed to be a Caf1R-dependent process. Therefore, the substitution of *caf1R* start codon by a random 3 nt and insertion of three stop codons at the beginning of *caf1R* gene should be enough to stop the expression of *caf1R*. So, a large decrease in *gfp* expression is a hypothetical outcome for *caf1R* deletion. *E. coli* BL21 (DE3) cells transformed with either pGEM-T vector, pET-GFP, pRC3-GFP or pRC3ΔR GFP, grown at either 25°C or 35°C for 16 h in plate reader using 200 µl of inoculated TB media per each well with appropriate antibiotics in 96 well plate (n=9, three biological and three technical replicates for each culture). OD_{600} and fluorescence were measured every 10 minutes. Two time points were used to calculate the GFP synthesis rate of 2 h (between 14 h and 16 h) for each culture at both temperatures 25°C and 35°C subtracted from the GFP synthesis rate of the negative control. The SDS-PAGE analysis was performed to examine the production of GFP in all above cultures. The mean values of GFP synthesis rate per cell above that of negative control are stated in **Table 5-9** and the statistical analysis was performed using T-test as shown in **Table 5-10**.

Table 5-9. The mean values of GFP synthesis rate per cell. Produced by *E. coli* BL21 (DE3) transformed with either pET-GFP, pRC3-GFP or pRC3ΔR GFP grown at 25°C and 35°C separately. ± value is the standard error of the mean.

***E. coli* BL21 (DE3) transformed with**	**The mean values of GFP synthesis rate per cell above that of negative control**	
	at 25°C	**at 35°C**
pET-GFP	903 ± 20	2212 ± 110
pRC3-GFP	198 ± 18	1446 ± 49
pRC3ΔR GFP	18 ± 3	513 ± 112

Table 5-10. The P-values of statistical comparisons between the grown cultures of pET-GFP, pRC3-GFP and pRC3ΔR GFP at two different temperatures 25°C and 35°C using T-test.

The comparison between cultures of	**P-values**
pET-GFP at 25°C vs. pET-GFP at 35°C	<0.001
pRC3-GFP at 25°C vs. pRC3-GFP at 35°C	<0.001
pRC3ΔR GFP at 25°C vs. pRC3ΔR GFP at 35°C	0.01

The graph of GFP synthesis rate per cells of the cultures of pRC3-GFP and pRC3ΔR GFP transformants and their SDS-PAGE results **(Figure 5-5)** showed a large decrease in the *gfp* expression at 25°C and 35°C after *caf1R* deletion. The mean value of the GFP synthesis rate per cell of the pRC3-GFP transformants cultures grown at 25°C was decreased by 11 times after *caf1R* deletion. Whereas, it was decreased by approximately 7.4 times when the same cultures grown at 35°C after *caf1R* knock-out. Therefore, the Caf1R transcription factor is responsible for most of the *gfp* transcription in this system at 35°C. In addition, the majority of gfp expression was lost for the cultures of pRC3-GFP grown at 25°C when *caf1R* deleted.

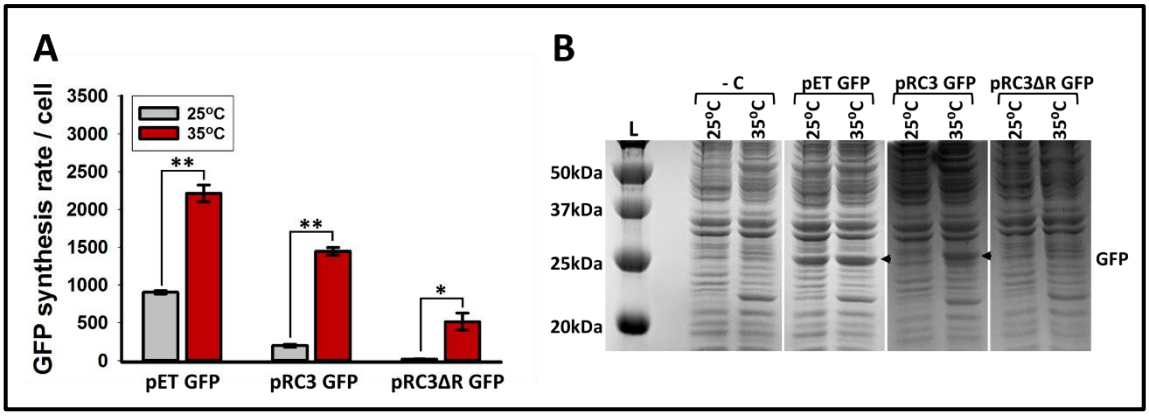

Figure 5-5. The effect of *caf1R* knock-out on the thermal-responsive *gfp* expression system. (A) Graph of GFP production in response to thermal change for three different cultures of *E. coli* BL21 (DE3) transformed with either pET GFP, pRC3 GFP or pRC3ΔR GFP (same as pRC3 GFP with deleted *caf1R* gene). The liquid cultures were grown for 16 hours at either 25°C or 35°C separately. The GFP synthesis rate were measured between two time points (14 and 16 hours) as mentioned in materials and methods. Bar heights correspond to GFP synthesis rate per cell subtracted from the negative control (*E. coli* BL21 (DE3) cells transformed with pGEM-T vector). Error bars represent standard error of the mean (S.E.M). NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). (B) SDS-PAGE analysis of the same cultures stated in (A) showing the GFP bands at ≈ 27 kDa in the lanes of pET GFP transformant cultures at both temperatures 25°C and 35°C with less intensity at 25°C, clear GFP band in the lane of pRC3 GFP transformant culture at 35°C only. Whereas, pRC3ΔR GFP transformant cultures did not produce GFP bands. Biological replicates = 3. Statistical analysis was performed using T-test.

5.2.6. The suggested models for thermo-responsive transcriptional control of gene expression by Caf1R

There are three potential mechanisms by which Caf1R could function to activate the *caf1* operon gene expression **(Figure 5-6)**. In the first model, regardless of its binding state to DNA, Caf1R is present at 25°C but its unable to activate the gene expression. The conformation of Caf1R changes when the temperature rises to 35°C immediately to highly increase the RNA transcript levels of *caf1* operon genes **(Figure 5-6 A)**. The second model suggests that the abundance of Caf1R is itself thermo-responsive, the Caf1R

transcription is not temperature-sensitive but the translation or stabilisation of Caf1R protein happens only at 35°C to induce the gene expression **(Figure 5-6 B)**.

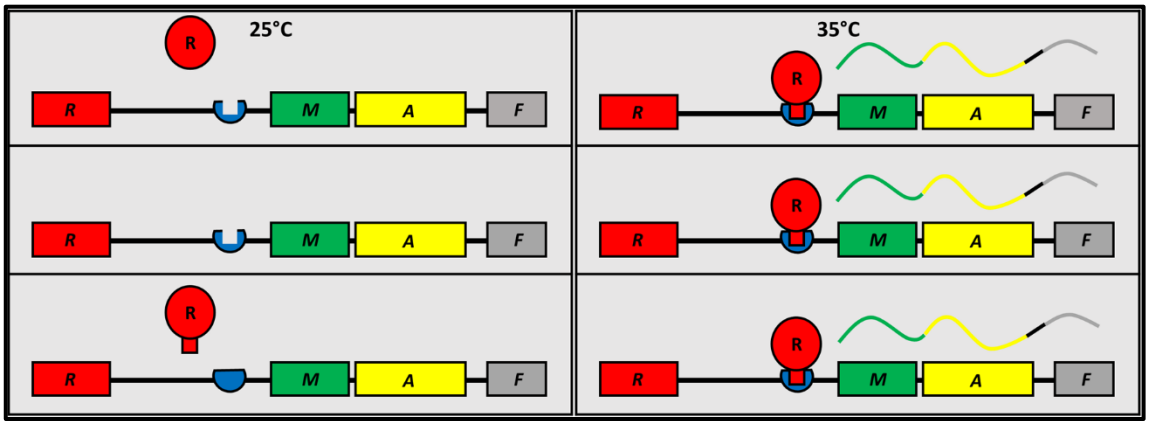

Figure 5-6. Models of thermo-responsive Caf1R-dependent transcriptional control of gene expression. Three models that potentially describe the Caf1R dependent thermo-responsive increase in *caf1* operon transcription are shown, with their states at 25°C and 35°C depicted in the left- and right-hand side panels respectively. In model (A), the Caf1R protein is the thermo-responsive element. At 25°C it is unable to activate transcription of the *caf1* operon, regardless of its DNA binding state. At 35°C, a conformational change allows the protein to activate transcription. Inactive Caf1R is shown as a red circle and active Caf1R as a red circle with square basement. In (B), the Caf1R protein abundance is thermo-responsive. Transcription is not thermally regulated so either Caf1R is always transcribed but only translated at 35°C or Caf1R is stabilised at 35°C. The increased Caf1R protein level activates transcription of the operon. In (C), the DNA is the thermo-responsive element. At 25°C, the DNA is not in an optimal conformation for Caf1R binding, and so transcription is not activated. At 35°C, a change in the DNA facilitates optimal Caf1R binding, and so transcription of the operon is activated. Genes are shown as coloured rectangles (red, green, yellow and grey for *caf1R*, *caf1M*, *caf1A* and *caf1* respectively), with the Caf1R DNA binding site shown in blue. mRNA transcripts are represented as wavy lines coloured according to the gene they are transcribed from.

Whereas, the third potential mechanism proposes that the Caf1R-binding sequence in DNA is responsible for such temperature-sensitive regulation. The DNA conformation is not optimal for Caf1R binding at 25°C and once the temperature rises to 35°C, the DNA conformation will be change to facilitate the Caf1R binding and activate the transcription of *caf1* operon (Al-Jawdah

et al., 2019). The temperature-sensitive regulation by Caf1R could be a useful process in biotechnology and synthetic biology, since it could be employed to create a thermo-responsive expression system induced by changing temperature to 35^{o}C after non-expressing growth at 25^{o}C to obtain the optimal OD_{600} before induction.

5.2.7. Troubleshooting of some issues related to this study

5.2.7.1. C-terminal FLAG-Tag of the *caf1* operon gene products affect the production of Caf1 polymer

In order to examine the effect of C-terminal FLAG-tag of either Caf1R, Caf1M and Caf1A on the Caf1 polymer production by *E. coli* BL21 (DE3), a comparison was performed between pT7-COP (which is composed of pGEM-T vector with T7 promoter and *caf1* operon) and pT7-COPF (which is composed of pGEM-T vector with T7 promoter and *caf1* operon with C-terminal FLAG-tag). The main advantage of using FLAG-tagged *caf1* operon gene products is to be able to detect them by western blot. *E. coli* BL21 (DE3) cells were transformed with either pT7-COP or pT7-COPF, grown in liquid culture media with appropriate antibiotics at 35^{o}C for 16 h (n=3 for both cultures). The thickness of flocculent layer was measured for these cultures per (mm) after centrifugation in capillary tubes as stated in (materials and methods section).

SDS-PAGE analysis was performed to examine the production of Caf1 by these cultures which confirmed the presence of Caf1 bands in all mentioned cultures **(Figure 5-7 B)**. The mean values of the flocculent layer height of the pT7-COP and pT7-COPF transformants cultures were: 32.6 mm and 16.3 mm respectively **(Figure 5-7 A)** and there is a statistically significant difference

between them with a P-value < 0.001. Although the Caf1 polymer production is still active, the amount of flocculent layer has decreased clearly in the presence of FLAG-tag at the C-terminus of either Caf1R, Caf1M or Caf1A. It has been concluded that the C-terminal FLAG-tags have a negative effect on the Caf1 polymer production process. Therefore, it is required to determine which individual FLAG-tagged protein is the reason behind such drop in Caf1 production.

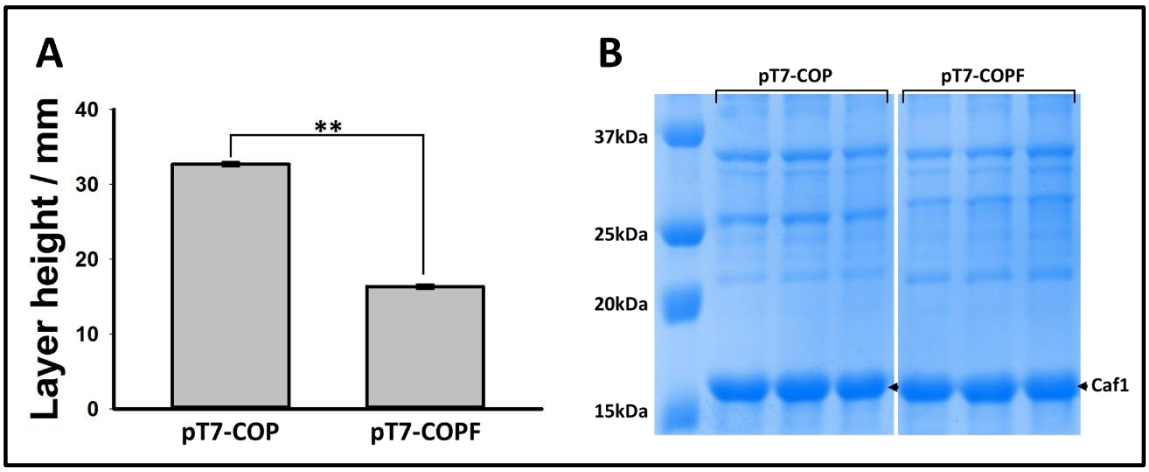

Figure 5-7. The effect of the C-terminal FLAG-tags of the *caf1* operon gene products on the Caf1 polymer production. (A) Graph of flocculent layer heights obtained for cultures of *E. coli* BL21 (DE3) transformed with either pT7-COP or pT7-COPF (Caf1R^{FLAG}, Caf1M^{FLAG}, Caf1A^{FLAG}) grown for 16 h at 35°C. Bar heights correspond to mean values of the flocculent layer heights obtained from three separate cultures from each transformants, where error bars represent the standard error of the mean (S.E.M). (B) SDS-PAGE of the same cultures (biological triplicate) mentioned in (A), heated flocculent layer with same amounts in each (at 100 °C for 10 minutes). The Caf1 bands are obvious at ≈ 15.5 kDa in all samples. NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). Biological replicates = 3. Statistical analysis was performed using T-test.

5.2.7.2. The C-terminal FLAG-tag of the Caf1A is the main reason for the reduction in Caf1 polymer production

To determine which FLAG-tagged protein exactly affect the production of Caf1 polymer, the pT7-COP used to generate either pT7-COP R^{FLAG}, pT7-COP M^{FLAG} or pT7-COP A^{FLAG} by In-Fusion HD cloning as stated in Materials and Methods section. *E. coli* BL21 (DE3) cells were transformed with either pT7-COP, pT7-COP R^{FLAG}, pT7-COP M^{FLAG} or pT7-COP A^{FLAG}. The inoculated liquid cultures were grown with appropriate antibiotics at 35^{o}C for 16 h to assess the amount of flocculent layer in each culture (biological triplicate of each culture). The mean values of the flocculent layer thickness **(Table 5- 11)** were measured by centrifugation of liquid cultures in capillary tubes and SDS-PAGE was achieved to examine *caf1* expression in the mentioned cultures.

Therefore, the presence of FLAG-tag at the C-terminus of the Caf1A led to a significant decrease in the Caf1 polymer production indicating that this tag affected the function of Caf1A negatively by involving in the biogenesis of Caf1.

Table 5-11. The mean values of the flocculent height. Produced by the cultures of pT7-COP, pT7-COP R^{FLAG}, pT7-COP M^{FLAG} and pT7-COP A^{FLAG}, grown at 35^{o}C for 16 h. ± value is the standard error of the mean.

***E. coli* BL21 (DE3) transformed with**	**The mean values of the flocculent layer thickness**
pT7-COP	37 mm ± 0.5
pT7-COP R^{FLAG}	33 mm ± 1
pT7-COP M^{FLAG}	28 mm ± 0.3
pT7-COP A^{FLAG}	20 mm ± 0.5

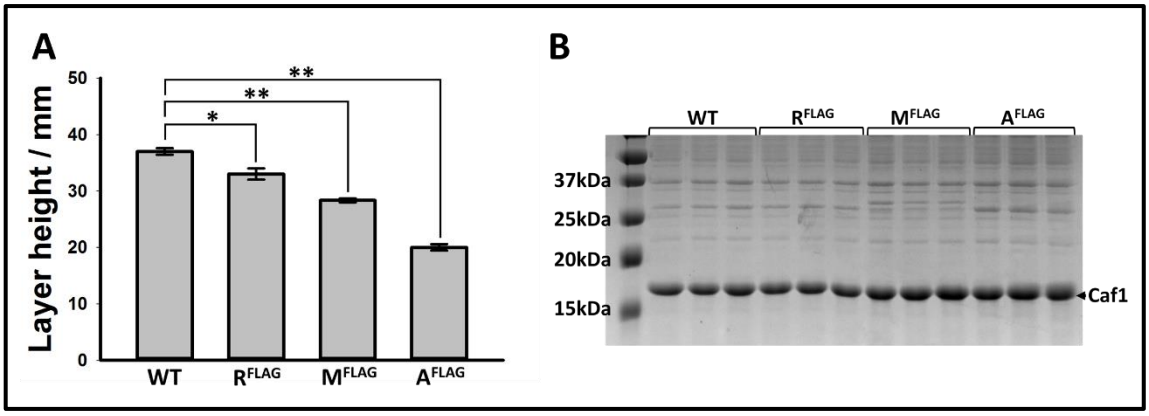

Figure 5-8. The effect of the C-terminal FLAG-tagged *caf1* operon gene products on the Caf1 polymer production. (A) Graph of flocculent layer height obtained for cultures of *E. coli* BL21 (DE3) transformed with either pT7-COP (WT *caf1* operon), pT7-COP R^{FLAG}, pT7-COP M^{FLAG} or pT7-COP A^{FLAG}, grown for 16 h at 35°C. Bar heights correspond to mean values of the flocculent layer height obtained from three separate cultures of each transformant, where error bars represent the standard error of the mean (S.E.M). (B) SDS-PAGE of the heated flocculent layer (at 100 °C for 10 minutes) obtained from the same cultures mentioned in (A) (biological triplicate). The Caf1 bands are obvious at ≈ 15.5 kDa in all samples. NS means that the difference is statistically not significant (when P-value > 0.05), * means that there is a statistically significant difference (when P-value ranges from 0.05 to 0.001) and ** means that there is a statistically significant difference (when P-value <0.001). Biological replicates = 3. Statistical analysis was performed using One-Way ANOVA test.

5.2.7.3. The insertion of FLAG-tag in a flexible loop in the central β-barrel of Caf1A has no effect on the Caf1 production

First of all, in all western blot results shown in this study, the FLAG-tagged Caf1A could not be detected, whereas, both the FLAG-tagged Caf1R and Caf1M were detected many times using anti-FLAG antibodies as stated in materials and methods section. This could indicate that the C-terminal FLAG-tag of Caf1A could be cleaved off during the Caf1 biogenesis process. Therefore, a new position was selected for FLAG-tag insertion in the central β-barrel facing the outside. A model of Caf1A was generated based on the crystal structure of FimD usher (PDB code: 3RFZ) to determine which is the best location of the FLAG-tag insertion **(Figure 5-9 A)** to avoid any interference with the Caf1 polymer secretion process. The new FLAG-tag

containing plasmid (pT7-COP $A^{new\text{-}FLAG}$) was generated by In-Fusion HD cloning technique as stated in materials and methods section. The FLAG-tag was inserted between the codons of proline 467 and asparagine 468.

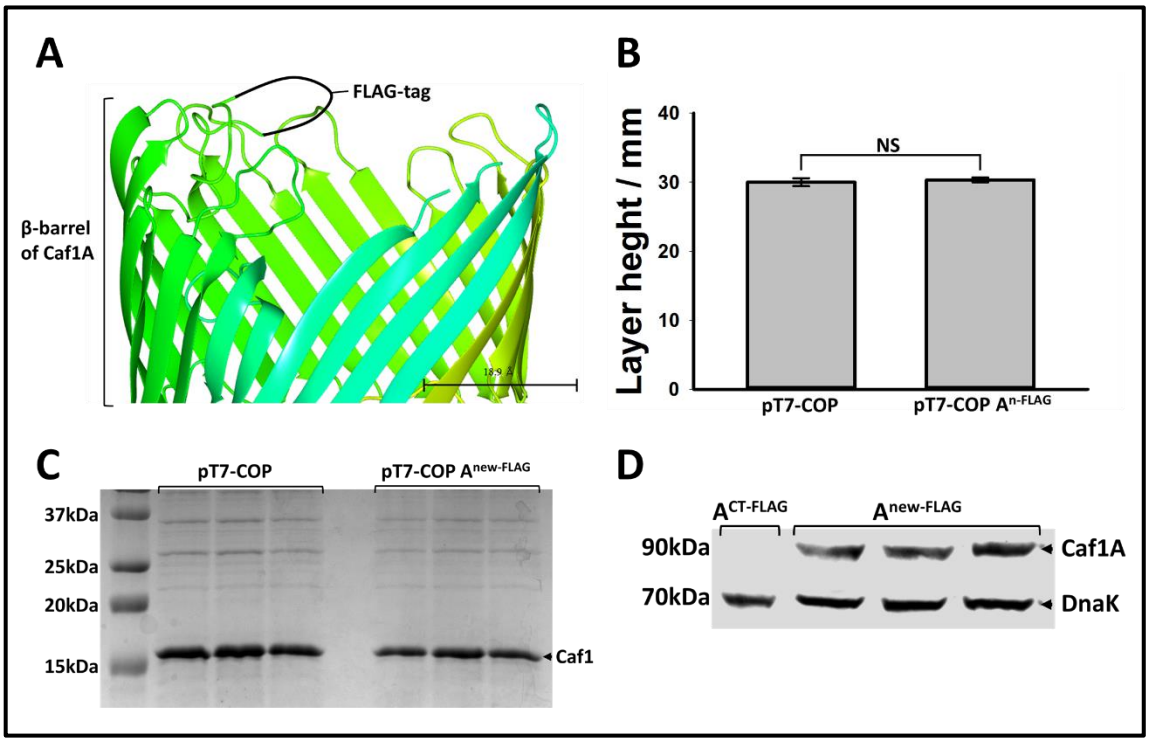

Figure 5-9. The effect insertion of FLAG-tag insertion in a flexible loop in the central β-barrel of Caf1A. (A) Structural modelling of the Caf1A usher outer membrane protein. Based on the crystal structure of FimD usher (PDB code: 3RFZ) (Phan *et al.*, 2011). Showing the central β-barrel with FLAG-tag located in a flexible loop facing the outer surface in black color. (B) Graph of the flocculent layer heights obtained for cultures of *E. coli* BL21 (DE3) transformed with either pT7-COP or pT7-COP $A^{n\text{-}FLAG}$, grown for 16 h at 35°C. Bar heights correspond to mean values of the flocculent layer heights obtained from three separate cultures from each transformant, where error bars represent the standard error of the mean (S.E.M). NS means that the difference is statistically not significant (when P-value > 0.05) (C) SDS-PAGE of the flocculent layer obtained from the same cultures (biological triplicate) mentioned in (B), heated at 100 °C for 10 minutes. The Caf1 bands are obvious at ≈ 15.5 kDa in all samples. (D) Western blot analysis of the cultures of *E. coli* BL21 (DE3) transformed with pT7-COP $A^{CT\text{-}FLAG}$ or pT7-COP $A^{new\text{-}FLAG}$, grown for 16 h at 35°C using anti-FLAG tag and anti-DnaK antibodies. The cell pellets were used to detect the FLAG-tagged Caf1A and DnaK as a loading control. Showing DnaK bands at ≈ 70 kDa and Caf1A at ≈ 90 kDa. Biological replicates = 3. Statistical analysis was performed using T-test.

E. coli BL21 (DE3) cells were transformed with either pT7-COP $A^{CT\text{-}FLAG}$ or pT7-COP $A^{new\text{-}FLAG}$, the inoculated culture media (biological triplicate from each were grown at 35^{o}C for 16 h. The measurements of the flocculent layer heights were taken per (mm) after cultures centrifugation in capillary tubes as stated in materials and methods section. The mean values of the flocculent layer thickness are: 30 mm and 30.3 mm for pT7-COP $A^{CT\text{-}FLAG}$ and pT7-COP $A^{new\text{-}FLAG}$ respectively. The flocculent layer samples obtained from these cultures were analysed by western blot using anti-FLAG and anti-DnaK antibodies. The new position of the FLAG-tag in Caf1A did not affect the production of Caf1 polymer **(Figure 5-9 B and C)** as observed in the old FLAG-tagged Caf1A. The western blot analysis could not detect the old FLAG-tagged Caf1A **(Figure 5-9 D)**, whereas, the new FLAG-tagged Caf1A bands were clearly visible.

5.3. Discussion

As shown in this study (5.2.2), Caf1R was used to create a thermo responsive expression system. Although the GFP production from pR-GFP expression vector was low, the expression of *gfp* was temperature sensitive. It seems that the native Caf1R system is a weak expression system. The *gfp* expression from pR-GFP vector is entirely driven by Caf1R which agrees with what was seen in this study (3.2.5, 3.2.6 and 3.2.8). The level of *caf1* operon gene expression driven by Caf1R in the pCOP vector was very low as compared to the level of *caf1* operon transcripts produced from the leaky expression of T7 RNA polymerase in pT7-COPΔR. The levels of *gfp* expression was enhanced at 35^{o}C by insertion of the strong insulated constitutive promoter taken from Davis *et al.* (2011).

Interestingly, the GFP production at 25^{o}C was very low even in the presence of the strong insulated constitutive promoter which points to the attendance of post-transcriptional regulation alongside the Caf1R-dependent regulation. According to our results it was suggested by this study in (4.3) that the temperature regulation of the *caf1* operon happens mainly at the transcriptional level with post-transcriptional regulation having very little effect at 30^{o}C and 35^{o}C, if any, role. Therefore, the post-transcriptional down regulation of *caf1* operon genes, *caf1M*, *caf1A* and *caf1* could only occur at low temperatures such as 25^{o}C but not at intermediate and high temperature degrees 30^{o}C and 35^{o}C, respectively. Wan *et al.* (2012) reported that many RNA secondary structures in yeast melt between 30^{o}C and 35^{o}C which could agree with the hypothesis that RNA secondary structures exist at 25^{o}C but not at 30^{o}C.

Additionally, the *gfp* expression was increased from pRC3-GFP as a result of the insertion of *intergenic region 3* (which separates between *caf1A* and *caf1*) and the ribosome binding site (RBS) of *caf1* gene. It was reported previously in this study that *intergenic region 3* is not likely to have a functional promoter, since 7 out of 8 of sequences obtained from RACE experiments continued through the *intergenic region 3* into *caf1A* gene indicating that the termination is due to dissociation of polymerase enzyme rather than a transcriptional start site with in this region. Galyov *et al.* (1990) also stated that this region in isolation did not enable the expression of *caf1*. Therefore, this increase in *gfp* expression by pRC3-GFP as compared to that obtained from pRC2-GFP cultures is a consequence of replacing the RBS of *caf1M* by that of *caf1*. It seems from these results that the RBS of *caf1* is stronger than that of *caf1M*. However, this could harmonise with the requirement of the Caf1 biogenesis system, larger amount of Caf1 subunits are demanded than chaperones and ushers.

The most similar artificial temperature sensitive expression system to our system is that developed by Lermant *et al.* (2018). This expression system can produce different proteins selectively in response to different ranges of temperature (Lermant *et al.*, 2018). The regulatory sequence of the *E. coli* Cold Shock Protein A (Gualerzi *et al.*, 2003) was utilised as a cold responsive transcriptional activator element. Whereas, the termo-labile mutant of λcl repressor and its pL promoter λ phage (Lieb, 1981; Dodd and Egan, 2002) were exploited as a factor to activate the expression in response to high temperature. In terms of complexity, the system of Lermant *et al.* (2018) seems to be more complicated than our system. Since, two biological systems from different organisms, *E. coli* and λ phage were functionalised to

generate this system. While, in our system we aimed to use *caf1R*, its regulatory site and promoter.

However, the main goal from developing the Caf1R-dependent system was to generate an expression vector for protein purification purposes. Whereas, the other thermo responsive was developed by Lermant *et al.* (2018) for the purpose of maintaining the productivity of the crops alongside with the issue of the climate change, since, this system allows the selective production of different compounds which protect the crops over these differences in temperature ranges.

In addition, it is obvious from the results of this study that Caf1R is the factor that transduces the thermal change into a transcriptional response. Three models have been suggested to clarify the potential mechanism of temperature sensitive regulation of *caf1* operon by Caf1R. Briefly, either Caf1R can only activate the transcription at 35^{o}C due to the conformational changes in its structure which does not occur at 25^{o}C (model A), the amount of translated Caf1R itself is increased at 35^{o}C to activate the *caf1* operon transcription directly. Since, *caf1R* mRNA forms secondary structures at 25^{o}C which block its binding to ribosomes (model B) or Caf1R can recognise its DNA binding site only at 35^{o}C to activate the transcription. Whereas, at 25^{o}C special DNA structures (e.g. supercoiling or promoter topology) prevent Caf1R from binding to DNA (model C).

The RovA global transcription factor in *Yersinia* species is a well-studied example for model A of temperature sensitive regulation (Cathelyn *et al.*, 2006; Quade *et al.*, 2012). RovA activates the expression of invasin, the internalization factor in enteropathogenic *Yersinia* species (Revell and Miller, 2000; Heroven *et al.*, 2004). The conformational changes in RovA structure

in response to temperature increase lead to partial unfolding of this protein which in turns lowers its affinity to DNA and increases its susceptibility to protease, thus the release of RovA from DNA results in de-repressing of the target genes (Quade *et al.*, 2012).

The LcrF transcription activator in *Y. pestis* is an example for model B of temperature sensitive regulation, which activates the expression of type III secretion system in response to temperature (Hoe and Goguen, 1993). The presence of secondary structure in *lcrF* mRNA at 26^{o}C decreases the efficiency of translation by sequestering the Shine-Dalgarno sequence and preventing mRNA from binding to ribosomes (Hoe and Goguen, 1993). When the temperature rises such secondary structures will disappear to induce the translation of *lcrF* mRNA and the LcrF transcription factor will then activate the transcription of type III secretion system gene cluster (Hoe and Goguen, 1993). The thermoregulation of *Shigella* species and enteroinvasive *E. coli* pathogenicity is considered as an example for model C of temperature sensitive regulation (Falconi *et al.*, 1998). Since, the structure of DNA undergoes a transition in response to temperature change and these structural changes modulate the accessibility of the H-NS transcriptional repressor to the *virF* promoter (Falconi *et al.*, 1998).

Chapter Six: Conclusions and Future Perspective

6. Chapter Six: Conclusions and Future Perspective

6.1. Conclusions

Y. pestis produces Caf1 polymer as antiphagocytic factor only when it senses human body temperature. The temperature sensitive upregulation of the *caf1* operon transcription has been reported (Han *et al.*, 2004; Motin *et al.*, 2004). Although, Caf1R was described as a positive transcription factor (Karlyshev *et al.*, 1992b), there were no clear results describing the role of Caf1R in the thermo-responsive transcriptional regulation of *caf1* operon. This study has shown that Caf1R is the responsible transcriptional regulator responsible for the activation of *caf1* operon gene transcription in response to temperature change. A putative promoter and regulatory elements were previously predicted in the intergenic region 1 upstream of *caf1M* (Galyov *et al.*, 1990), but the presence of them had not been confirmed experimentally. Interestingly, our results confirmed that this putative promoter region contains the responsible promoter for the transcription of *caf1M*, *caf1A* and *caf1* genes as a single polycistronic mRNA unit. The data, however did not clarify the mode of action of Caf1R. Therefore, the main conclusions obtained from this study are as follows:

6.1.1. Caf1R control of the Caf1 polymer secretion

According to the results achieved in this study, it should be concluded that the use of capillary tubes in the measurement of flocculent layer thickness allows to minimise the bacterial culture volume used in the measurement process and enhance the accuracy of separation. The Caf1 polymer amount is constant per volume of the flocculent layer even when that flocculent obtained from different time points. Although the *caf1R* overexpression led

to a stepwise decline in Caf1 polymer production when six different concentrations of L-arabinose used separately, the overexpression of other *caf1* operon genes did not affect the production of the Caf1. The *caf1R* gene was required for the expression of *caf1* operon. Whereas, the Caf1 polymer production was rescued either by using a plasmid carrying *caf1R* or by insertion of T7 promoter in the same plasmid containing *caf1* operon apart from *caf1R*. The deletion of *caf1R* in the presence of T7 promoter led to enhance the expression of *caf1* operon at both 30^{o}C and 35^{o}C indicating that Caf1R could bind in the way of T7 RNA polymerase to reduce its transcription efficiency. Although different *E. coli* strains and different plasmid constructs were used to purify Caf1R with long optimisation process, no pure Caf1R was obtained at the end as a result of instability and insolubility of this protein.

6.1.2. The transcriptional regulation of *caf1* operon by Caf1R in response to temperature

It was implied that *caf1* operon, *caf1M*, *caf1A* and *caf1* genes are transcribed as a single polycistronic mRNA as both *intergenic 2* (between *caf1M* and *caf1A*) and *intergenic 3* (between *caf1A* and *caf1*) were detected in the cDNA obtained from the purified RNA of *caf1* expression culture. The polycistronic transcription of the *caf1* operon was confirmed by Rapid Amplification of cDNA Ends (RACE) using RNA purified from *caf1* expressing cultures, since, the most sequences obtained by RACE continued the *intergenic region 3* into *caf1A* gene. The P2 sequence predicted in the *intergenic region 1* (between *caf1R* and *caf1M*) had a significant effect on the *caf1* operon expression. The deletion of this sequence led to a large decrease in the levels of the transcripts and halted the production of the Caf1M and Caf1 proteins

completely. Therefore, it was suggested that the P2 sequence contains the promoter responsible for the polycistronic transcription of the *caf1* operon. Although the ribosome binding sites (RBS) of the *caf1M*, *caf1A* and *caf1* were identified manually based on the Shine-Dalgarno consensus sequence, only the *caf1* RBS was confirmed by RT-PCR western blot analysis. Caf1R is the responsible transcriptional activator for *caf1* operon expression in response to the temperature change. Additionally, the sequence alignment and structural modelling of Caf1R revealed the high homology of its N-terminus with the helix-turn-helix DNA binding domains of the MarA and Rob from AraC family. The Caf1R C-terminus did not show any homology with the mentioned proteins suggesting that the mechanism of Caf1R regulation is different from MarA and Rob. It is a GyrI-like structure as in Rob with a strand-helix-strand-strand (SHS2) domains. The identification of Caf1R C-terminus to have a SHS2 domains could imply that Caf1R function depends on a small molecule binding. To confirm the importance of Caf1R C-terminus, the deletion of the C-terminus led a stop in the production of Caf1 polymer.

6.1.3. The Caf1R-dependent expression of *gfp* in response to temperature

Three models were suggested to clarify the molecular mechanism of Caf1R regulation in response to temperature. In the first model, the Caf1R itself is the thermo-responsive factor, in the second one the increase in *caf1R* translation is temperature sensitive and in the third model, a change in the DNA structure at the regulatory site mediates the regulation in response to temperature. In the constructed plasmid pR-GFP, the production of GFP was temperature sensitive. Although the Caf1R-dependent expression was low at 35^{o}C, no GFP production was obtained at 25^{o}C. The insertion of a strong

insulated promoter in *intergenic region 1* downstream of the predicted *caf1* operon promoter led to improved expression of *gfp* at 35^{o}C with very low expression at 25^{o}C indicating that there could be a post-transcriptional thermoregulation along with the Caf1R-dependent transcriptional regulation. The insertion of *intergenic region 3* from *caf1* operon which includes the RBS of *caf1* increased the production of GFP at 35^{o}C with low expression at 25^{o}C. Therefore, this increase might be a result of utilising *caf1* RBS instead of *caf1M* RBS.

6.2. Future perspective

Future experiment should be carried out to purify both N-terminus and C-terminus of Caf1R separately, in order to achieve molecular biological and biophysical examination using pure N-terminus and C-terminus. For example, using an Electrophoretic Mobility Shift Assay (EMSA) to determine the Caf1R binding site and using circular dichroism spectroscopy to investigate if there are conformational changes in Caf1R structure at 25^{o}C and 35^{o}C. Further experiments are needed to examine the importance of the Caf1R C-terminus in *caf1* operon expression using RT-PCR and western blot analysis. In addition, the Caf1R-dependant thermo-responsive expression system needs for further developments in order to be usable in biotechnology. Minimizing the number of elements involved in this system is one of the crucial priorities in the future. If the regulatory sites and promoters in *caf1* operon are determined, the artificial thermo-responsive expression system can be composed of *caf1R*, its promoter, its binding site and the promoter of *caf1* operon. The promoter of *caf1* operon could be replaced by a strong promoter to enhance the expression levels.

Bibliography

Al-Jawdah, A.D., Ivanova, I.G., Waller, H., Perkins, N.D., Lakey, J.H. and Peters, D.T. (2019) 'Induction of the immunoprotective coat of *Yersinia pestis* at body temperature is mediated by the Caf1R transcription factor', *BMC Microbiology*, 19(1), p. 68.

Anantharaman, V. and Aravind, L. (2004) 'The SHS2 module is a common structural theme in functionally diverse protein groups, like Rpb7p, FtsA, GyrI, and MTH1598/TM1083 superfamilies', *Proteins: Structure, Function, and Bioinformatics*, 56(4), pp. 795-807.

Anderluh, G., Gökçe, I. and Lakey, J.H. (2003) 'Expression of proteins using the third domain of the *Escherichia coli* periplasmic-protein TolA as a fusion partner', *Protein Expression and Purification*, 28(1), pp. 173-181.

Anisimov, A.P., Dentovskaya, S.V., Titareva, G.M., Bakhteeva, I.V., Shaikhutdinova, R.Z., Balakhonov, S.V., Lindner, B., Kocharova, N.A., Senchenkova, S.N., Holst, O., Pier, G.B. and Knirel, Y.A. (2005) 'Intraspecies and temperature-dependent variations in susceptibility of *Yersinia pestis* to the bactericidal action of serum and to polymyxin B', *Infect Immun*, 73(11), pp. 7324-31.

Ariza, R.R., Li, Z., Ringstad, N. and Demple, B. (1995) 'Activation of multiple antibiotic resistance and binding of stress-inducible promoters by *Escherichia coli* Rob protein', *Journal of Bacteriology*, 177(7), p. 1655.

Asako, H., Nakajima, H., Kobayashi, K., Kobayashi, M. and Aono, R. (1997) 'Organic solvent tolerance and antibiotic resistance increased by overexpression of *marA* in *Escherichia coli*', *Applied and Environmental Microbiology*, 63(4), p. 1428.

Baker, E.E., Sommer, H., Foster, L.E., Meyer, E. and Meyer, K.F. (1952) 'Studies on Immunization Against Plague', *The Journal of Immunology*, 68(2), p. 131.

Beveridge, T.J. (2001) 'Use of the gram stain in microbiology', *Biotech Histochem*, 76(3), pp. 111-8.

Brooks, G.F., Jawetz, E., Melnick, J.L. and Adelberg, E.A. (2013) *Jawetz, Melnick & Adelberg's medical microbiology*. 26th ed. / George F. Brooks ... [et al.].. edn. New York : London: New York : McGraw-Hill Medical.

Broz, P., Mueller, C.A., Müller, S.A., Philippsen, A., Sorg, I., Engel, A. and Cornelis, G.R. (2007) 'Function and molecular architecture of the *Yersinia* injectisome tip complex', *Molecular Microbiology*, 65(5), pp. 1311-1320.

Cantey, J.R., Blake, R.K., Williford, J.R. and Moseley, S.L. (1999) 'Characterization of the *Escherichia coli* AF/R1 pilus operon: novel genes necessary for transcriptional regulation and for pilus-mediated adherence', *Infect Immun*, 67(5), pp. 2292-8.

Carroll, K.C. (2015) *Jawetz, Melnick, & Adelberg's medical microbiology*. 27th edition. edn. New York. London : McGraw-Hill Education.

Cathelyn, J.S., Crosby, S.D., Lathem, W.W., Goldman, W.E. and Miller, V.L. (2006) 'RovA, a global regulator of *Yersinia pestis* specifically required for bubonic plague', *Proceedings of the National Academy of Sciences*, 103(36), p. 13514.

Chalton, D.A., Musson, J.A., Flick-Smith, H., Walker, N., McGregor, A., Lamb, H.K., Williamson, E.D., Miller, J., Robinson, J.H. and Lakey, J.H. (2006) 'Immunogenicity of a *Yersinia pestis* Vaccine Antigen Monomerized by Circular Permutation', *Infection and Immunity*, 74(12), p. 6624.

Chapman, D.A.G., Zavialov, A.V., Chernovskaya, T.V., Karlyshev, A.V., Zav, apos, yalova, G.A., Vasiliev, A.M., Dudich, I.V., Abramov, V.M., yalov, V.P. and MacIntyre, S. (1999) 'Structural and Functional Significance of the FGL Sequence of the Periplasmic Chaperone Caf1M of *Yersinia pestis*', *The Journal of Bacteriology*, 181(8), p. 2422.

Chen, J.-S., Zidwick, M.J. and Rogers∗, P. (2013) 'Organic Acid and Solvent Production: Butanol, Acetone, and Isopropanol; 1,3- and 1,2-Propanediol Production; and 2,3-Butanediol Production', in Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E. and Thompson, F. (eds.) *The Prokaryotes: Applied Bacteriology and Biotechnology*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 77-134.

Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S.J. and Knight, S.D. (1999) 'X-ray Structure of the FimC-FimH Chaperone-Adhesin Complex from Uropathogenic *Escherichia coli*', *Science*, 285(5430), pp. 1061-1066.

Cornelis, G.R. (2002) 'The *Yersinia* Ysc-Yop 'Type III' weaponry', *Nature Reviews Molecular Cell Biology*, 3(10), p. 742.

Cornelis, G.R. and Wolf-Watz, H. (1997) 'The *Yersinia* Yop virulon: a bacterial system for subverting eukaryotic cells', *Molecular Microbiology*, 23(5), pp. 861-867.

Cowan, C., Jones, H.A., Kaya, Y.H., Perry, R.D. and Straley, S.C. (2000) 'Invasion of Epithelial Cells by *Yersinia pestis:* Evidence for a *Y. pestis*-Specific Invasin', *Infection and Immunity*, 68(8), pp. 4523-4530.

Dangi, B., Gronenborn, A.M., Rosner, J.L. and Martin, R.G. (2004) 'Versatility of the carboxy-terminal domain of the α subunit of RNA polymerase in transcriptional activation: use of the DNA contact site as a protein contact site for MarA', *Molecular Microbiology*, 54(1), pp. 45-59.

Darby, C., Ananth, S.L., Tan, L. and Hinnebusch, B.J. (2005) 'Identification of gmhA, a *Yersinia pestis* gene required for flea blockage, by using a *Caenorhabditis elegans* biofilm system', *Infection and immunity*, 73(11), pp. 7236-7242.

Darby, C., Hsu, J.W., Ghori, N. and Falkow, S. (2002) 'Plague bacteria biofilm blocks food intake', *Nature*, 417(6886), pp. 243-244.

Davis, J.H., Rubin, A.J. and Sauer, R.T. (2011) 'Design, construction and characterization of a set of insulated bacterial promoters', *Nucleic acids research*, 39(3), pp. 1131-1141.

Degiacomi, M.T., Iacovache, I., Pernot, L., Chami, M., Kudryashev, M., Stahlberg, H., Van Der Goot, F.G. and Dal Peraro, M. (2013) 'Molecular assembly of the aerolysin pore reveals a swirling membrane-insertion mechanism', *Nature Chemical Biology*, 9(10), pp. 623-9.

Dheda, K., Huggett, J.F., Bustin, S.A., Johnson, M.A., Rook, G. and Zumla, A. (2004) 'Validation of housekeeping genes for normalizing RNA expression in real-time PCR', *BioTechniques*, 37(1), p. 112.

Di Yu, X., Dubnovitsky, A., Pudney, Alex F., MacIntyre, S., Knight, Stefan D. and Zavialov, Anton V. (2012) 'Allosteric Mechanism Controls Traffic in the Chaperone/Usher Pathway', *Structure*, 20(11), pp. 1861-1871.

Dodd, I.B. and Egan, J.B. (2002) 'Action at a distance in CI repressor regulation of the bacteriophage 186 genetic switch', *Molecular Microbiology*, 45(3), pp. 697-710.

Dodson, K.W., Jacob-Dubuisson, F., Striker, R.T. and Hultgren, S.J. (1993) 'Outer-Membrane PapC Molecular Usher Discriminately Recognizes Periplasmic Chaperone-Pilus Subunit Complexes', *Proceedings of the National Academy of Sciences of the United States of America*, 90(8), pp. 3670-3674.

Donovan, R.S., Robinson, C.W. and Glick, B.R. (1996) 'Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of thelac promoter', *Journal of Industrial Microbiology*, 16(3), pp. 145-154.

Du, Y. (2002) 'Role of Fraction 1 Antigen of *Yersinia pestis* in Inhibition of Phagocytosis', *Infection and Immunity*, 70(3), pp. 1453-1460.

Du, Y., Rosqvist, R. and Forsberg, Å. (2002) 'Role of Fraction 1 Antigen of *Yersinia pestis* in Inhibition of Phagocytosis', *Infection and Immunity*, 70(3), pp. 1453-1460.

Dubnovitsky, A.P., Duck, Z., Kersley, J.E., Härd, T., Macintyre, S. and Knight, S.D. (2010) 'Conserved Hydrophobic Clusters on the Surface of the Caf1A Usher C-Terminal Domain Are Important for F1 Antigen Assembly', *Journal of Molecular Biology*, 403(2), pp. 243-259.

Dufresne, K., Saulnier-Bellemare, J. and Daigle, F. (2018) 'Functional Analysis of the Chaperone-Usher Fimbrial Gene Clusters of *Salmonella enterica serovar Typhi*', *Frontiers in cellular and infection microbiology*, 8, pp. 26-26.

Dumon-Seignovert, L., Cariot, G. and Vuillard, L. (2004) 'The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3)', *Protein Expression and Purification*, 37(1), pp. 203-206.

Egan, S.M. and Schleif, R.F. (1993) 'A Regulatory Cascade in the Induction of *rhaBAD*', *Journal of Molecular Biology*, 234(1), pp. 87-98.

Falconi, M., Colonna, B., Prosseda, G., Micheli, G. and Gualerzi, C.O. (1998) 'Thermoregulation of *Shigella* and *Escherichia coli* EIEC pathogenicity. A temperature-dependent structural transition of DNA modulates accessibility of promoter to transcriptional repressor H-NS', *The EMBO Journal*, 17(23), p. 7033.

Fawcett, W.P. and Wolf, R.E., Jr. (1995) 'Genetic definition of the *Escherichia coli zwf* "*soxbox*," the DNA binding site for SoxS-mediated induction of glucose 6-phosphate dehydrogenase in response to superoxide', *J Bacteriol*, 177(7), pp. 1742-50.

Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.-Y., Dosztányi, Z., El-Gebali, S., Fraser, M., Gough, J., Haft, D., Holliday, G.L., Huang, H., Huang, X., Letunic, I., Lopez, R., Lu, S., Marchler-Bauer, A., Mi, H., Mistry, J., Natale, D.A., Necci, M., Nuka, G., Orengo, C.A., Park, Y., Pesseat, S., Piovesan, D., Potter, S.C., Rawlings, N.D., Redaschi, N., Richardson, L., Rivoire, C., Sangrador-Vegas, A., Sigrist, C., Sillitoe, I.,

Smithers, B., Squizzato, S., Sutton, G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Xenarios, I., Yeh, L.-S., Young, S.-Y. and Mitchell, A.L. (2017) 'InterPro in 2017-beyond protein family and domain annotations', *Nucleic acids research*, 45(D1), p. D190.

Folkesson, A., Advani, A., Sukupolvi, S., Pfeifer, J.D., Normark, S. and Löfdahl, S. (1999) 'Multiple insertions of fimbrial operons correlate with the evolution of *Salmonella serovars* responsible for human disease', *Molecular Microbiology*, 33(3), pp. 612-622.

Foster, J.W. (2017) *Microbiology : an evolving science*. Fourth international.. edn. London: London : W W Norton.

Friedlander, A.M., Welkos, S.L., Patricia, L.W., Gerard, P.A., Heath, D.G., George W. Anderson, Jr., Margaret, L.M.P., James, E. and Kelly, D. (1995) 'Relationship between Virulence and Immunity as Revealed in Recent Studies of the F1 Capsule of *Yersinia pestis*', *Clinical Infectious Diseases*, 21, pp. S178-S181.

Friedman, B.A., Dugan, P.R., Pfister, R.M. and Remsen, C.C. (1969) 'Structure of Exocellular Polymers and Their Relationship to Bacterial Flocculation', *Journal of Bacteriology*, 98(3), p. 1328.

Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) 'Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific

oligonucleotide primer', *Proceedings of the National Academy of Sciences of the United States of America*, 85(23), p. 8998.

Gallegos, M., Schleif, R., Bairoch, A., Hofmann, K. and Ramos, J. (1997) 'AraC/XylS family of transcriptional regulators', *Microbiology And Molecular Biology Reviews*, 61(4), pp. 393-+.

Galyov, E.E., Karlishev, A.V., Chernovskaya, T.V., Dolgikh, D.A., Smirnov, O.Y., Volkovoy, K.I., Abramov, V.M. and Zav'Yalov, V.P. (1991a) 'Expression of the envelope antigen F1 of *Yersinia pestisis* mediated by the product of *caf1M* gene having homology with the chaperone protein PapD of *Escherichia coli*', *FEBS Letters*, 286(1-2), pp. 79-82.

Galyov, E.E., Karlishev, A.V., Chernovskaya, T.V., Dolgikh, D.A., Smirnov, O.Y., Volkovoy, K.I., Abramov, V.M. and Zav'Yalov, V.P. (1991b) 'Expression of the envelope antigen F1 ofYersinia pestisis mediated by the product ofcaf1Mgene having homology with the chaperone protein PapD ofEscherichia coli', *FEBS Letters*, 286(1-2), pp. 79-82.

Galyov, E.E., Smirnov, O.Y., Karlishev, A.V., Volkovoy, K.I., Denesyuk, A.I., Nazimov, I.V., Rubtsov, K.S., Abramov, V.M., Dalvadyanz, S.M., Zav, Amp, Apos and Yalov, V.P. (1990) 'Nucleotide sequence of the *Yersinia pestis* gene encoding F1 antigen and the primary structure of the protein: Putative T and B cell epitopes', *FEBS Letters*, 277(1), pp. 230-232.

Gambino, L., Gracheck, S.J. and Miller, P.F. (1993) 'Overexpression of the MarA positive regulator is sufficient to confer multiple antibiotic resistance in *Escherichia coli*', *J Bacteriol*, 175(10), pp. 2888-94.

Garciamartin, C., Baldoma, L., Badia, J. and Aguilar, J. (1992) 'Nucleotide-sequence of the *rhaR-sodA* interval specifying *rhaT* in *Escherichia coli*', *Journal Of General Microbiology*, 138, pp. 1109-1116.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.e., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005) 'Protein Identification and Analysis Tools on the ExPASy Server', in Walker, J.M. (ed.) *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press, pp. 571-607.

Gong, M. and Makowski, L. (1992) 'Helical structure of P pili from *Escherichia coli*: Evidence from X-ray fiber diffraction and scanning transmission electron microscopy', *Journal of Molecular Biology*, 228(3), pp. 735-742.

Goure, J., Broz, P., Attree, O., Cornelis, G.R. and Attree, I. (2005) 'Protective Anti-V Antibodies Inhibit *Pseudomonas* and *Yersinia* Translocon Assembly within Host Membranes', *The Journal of Infectious Diseases*, 192(2), pp. 218-225.

Greenwood, D. (2012) *Medical microbiology : a guide to microbial infections : pathogenesis, immunity, laboratory diagnosis and control*. 18th ed.. edn. Edinburgh: Edinburgh : Churchill Livingstone.

Griffith, K.L., Fitzpatrick, M.M., Keen, E.F., 3rd and Wolf, R.E., Jr. (2009) 'Two functions of the C-terminal domain of *Escherichia coli* Rob: mediating "sequestration-dispersal" as a novel off-on switch for regulating Rob's activity as a transcription activator and preventing degradation of Rob by Lon protease', *J Mol Biol*, 388(3), pp. 415-30.

Gu, J.-D. and Mitchell, R. (2013) 'Biodeterioration', in Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E. and Thompson, F. (eds.) *The Prokaryotes: Applied Bacteriology and Biotechnology*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 309-341.

Guajardo, R., Lopez, P., Dreyfus, M. and Sousa, R. (1998) 'NTP concentration effects on initial transcription by T7 RNAP indicate that translocation occurs through passive sliding and reveal that divergent promoters have distinct NTP concentration requirements for productive initiation', *J Mol Biol*, 281(5), pp. 777-92.

Gualerzi, C.O., Maria Giuliodori, A. and Pon, C.L. (2003) 'Transcriptional and post-transcriptional control of cold-shock genes', *Journal of Molecular Biology*, 331(3), pp. 527-539.

Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (1995) 'Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter', *J Bacteriol*, 177(14), pp. 4121-30.

Hächler, H., Cohen, S.P. and Levy, S.B. (1991) '*marA*, a regulated locus which controls expression of chromosomal multiple antibiotic resistance in *Escherichia coli*', *Journal of Bacteriology*, 173(17), p. 5532.

Han, Y., Zhou, D., Pang, X., Song, Y., Zhang, L., Bao, J., Tong, Z., Wang, J., Guo, Z., Zhai, J., Du, Z., Wang, X., Zhang, X., Wang, J., Huang, P. and Yang, R. (2004) 'Microarray Analysis of Temperature-Induced Transcriptome of *Yersinia pestis*', *Microbiology and Immunology*, 48(11), pp. 791-805.

Heptinstall, S., Archibald, A.R. and Baddiley, J. (1970) 'Teichoic Acids and Membrane Function in Bacteria', *Nature*, 225(5232), pp. 519-521.

Heroven, A.K., Nagel, G., Tran, H.J., Parr, S. and Dersch, P. (2004) 'RovA is autoregulated and antagonizes H-NS-mediated silencing of invasin and *rovA* expression in *Yersinia pseudotuberculosis*', *Molecular Microbiology*, 53(3), pp. 871-888.

Hinnebusch, B.J., Perry, R.D. and Schwan, T.G. (1996) 'Role of the *Yersinia pestis* hemin storage (*hms*) locus in the transmission of plague by fleas', *Science*, 273(5273), p. 367.

Hinnebusch, B.J., Rudolph, A.E., Cherepanov, P., Dixon, J.E., Schwan, T.G. and Forsberg, A. (2002) 'Role of *Yersinia* murine toxin in survival of *Yersinia pestis* in the midgut of the flea vector', *Science (New York, N.Y.)*, 296(5568), pp. 733-735.

Hoe, N.P. and Goguen, J.D. (1993) 'Temperature sensing in *Yersinia pestis*: translation of the LcrF activator protein is thermally regulated', *J Bacteriol*, 175(24), pp. 7901-9.

Hui, M.P., Foley, P.L. and Belasco, J.G. (2014) 'Messenger RNA degradation in bacterial cells', *Annu Rev Genet*, 48, pp. 537-59.

Hung, D., Knight, S., Woods, R., Pinkner, J. and Hultgren, S. (1996) 'Molecular basis of two subfamilies of immunoglobulin-like chaperones', *EMBO Journal*, 15(15), pp. 3792-3805.

Hyock Joo, K., Marjon, H.J.B., Bruce, D. and Tom, E. (2000) 'Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA', *Nature Structural Biology*, 7(5), p. 424.

Jair, K.W., Martin, R.G., Rosner, J.L., Fujita, N., Ishihama, A. and Wolf, R.E., Jr. (1995) 'Purification and regulatory properties of MarA protein, a transcriptional activator of *Escherichia coli* multiple antibiotic and superoxide resistance promoters', *J Bacteriol*, 177(24), pp. 7100-4.

Jarrett, C.O., Deak, E., Isherwood, K.E., Oyston, P.C., Fischer, E.R., Whitney, A.R., Kobayashi, S.D., DeLeo, F.R. and Hinnebusch, B.J. (2004) 'Transmission of *Yersinia pestis* from an Infectious Biofilm in the Flea Vector', *The Journal of Infectious Diseases*, 190(4), pp. 783-792.

Jens, K. and Franz, N. (2012) 'Bacterial RNA thermometers: molecular zippers and switches', *Nature Reviews Microbiology*, 10(4), p. 255.

Jones, C.H., Pinkner, J.S., Roth, R., Heuser, J., Nicholes, A.V., Abraham, S.N. and Hultgren, S.J. (1995) 'FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the *Enterobacteriaceae*', *Proceedings of the National Academy of Sciences*, 92(6), p. 2081.

Jones, H.A., Lillard, J.W., Jr. and Perry, R.D. (1999) 'HmsT, a protein essential for expression of the haemin storage (Hms+) phenotype of *Yersinia pestis*', *Microbiology*, 145 (Pt 8), pp. 2117-28.

Juturu, V. and Wu, J.C. (2014) 'Microbial cellulases: Engineering, production and applications', *Renewable and Sustainable Energy Reviews*, 33, pp. 188-203.

Kabir, M.M. and Shimizu, K. (2006) 'Investigation into the effect of *soxR* and *soxS* genes deletion on the central metabolism of *Escherichia coli* based on gene expressions and enzyme activities', *Biochemical Engineering Journal*, 30(1), pp. 39-47.

Kaldalu, N., Toots, U., de Lorenzo, V. and Ustav, M. (2000) 'Functional domains of the TOL plasmid transcription factor XylS', *Journal of bacteriology*, 182(4), pp. 1118-1126.

Karlyshev, A.V., Galyov, E.E., Abramov, V.M. and Zav' Yalov, V.P. (1992a) 'Caf1R gene and its role in the regulation of capsule formation of *Y. pestis*', *FEBS Letters*, 305(1), pp. 37-40.

Karlyshev, A.V., Galyov, E.E., Abramov, V.M., Zav and Yalov, V.P. (1992b) '*Caf1R* gene and its role in the regulation of capsule formation of *Y. pestis*', *FEBS Letters*, 305(1), pp. 37-40.

Karlyshev, A.V., Galyov, E.E., Smirnov, O.Y., Guzayev, A.P., Abramov, V.M. and Zav' Yalov, V.P. (1992c) 'A new gene of the ƒ1 operon of *Y. pestis* involved in the capsule biogenesis', *FEBS Letters*, 297(1-2), pp. 77-80.

Kataoka, K., Mizushima, T., Ogata, Y., Miki, T. and Sekimizu, K. (1996) 'Heat shock-induced DNA relaxation in vitro by DNA gyrase of *Escherichia coli* in the presence of ATP', *The Journal of biological chemistry*, 271(40), p. 24806.

Kennell, D. and Riezman, H. (1977) 'Transcription and translation initiation frequencies of the *Escherichia coli* lac operon', *Journal of Molecular Biology*, 114(1), pp. 1-21.

Klinkert, B. and Narberhaus, F. (2009) 'Microbial thermosensors', *Cellular and Molecular Life Sciences*, 66(16), pp. 2661-2676.

Knight, S.D. (2007) 'Structure and assembly of *Yersinia pestis* F1 antigen', *Genus Yersinia: From Genomics To Function*, 603, pp. 74-87.

Knight, S.D., Berglund, J. and Choudhury, D. (2000) 'Bacterial adhesins: structural studies reveal chaperone function and pilus biogenesis', *Current Opinion in Chemical Biology*, 4(6), pp. 653-660.

Kolodrubetz, D. and Schleif, R. (1981) 'Identification of *araC* protein on two-dimensional gels, its in vivo instability and normal level', *Journal of Molecular Biology*, 149(1), pp. 133-139.

Krajewski, S.S. and Narberhaus, F. (2014) 'Temperature-driven differential gene expression by RNA thermosensors', *BBA - Gene Regulatory Mechanisms*, 1839(10), pp. 978-988.

Kuehn, M.J., Ogg, D.J., Kihlberg, J., Slonim, L.N., Flemmer, K., Bergfors, T. and Hultgren, S.J. (1993) 'Structural basis of pilus subunit recognition by the PapD chaperone', *Science*, 262(5137), p. 1234.

Kukoyi, A.R. (2016) 'Economic Impacts of Natural Polymers', in Olatunji, O. (ed.) *Natural Polymers: Industry Techniques and Applications*. Cham: Springer International Publishing, pp. 339-362.

Kumagai, H. (2013) 'Amino Acid Production', in Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E. and Thompson, F. (eds.) *The Prokaryotes: Applied Bacteriology and Biotechnology*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 169-177.

Kumar, D. (2016) 'Caf1R-mediated regulation of the f1 surface antigen of *Yersinia pestis*'. ProQuest Dissertations Publishing.

Kutyrev, V.V., Filippov, A.A., Oparina, O.S. and Protsenko, O.A. (1992) 'Analysis of *Yersinia pestis* chromosomal determinants Pgm+ and Psts associated with virulence', *Microb Pathog*, 12(3), pp. 177-86.

Lambin, P., Rochu, D. and Fine, J.M. (1976) 'A new method for determination of molecular weights of proteins by electrophoresis across a sodium dodecyl sulfate (SDS)-polyacrylamide gradient gel', *Analytical Biochemistry*, 74(2), pp. 567-575.

Lancini, G. and Demain, A.L. (2013) 'Bacterial Pharmaceutical Products', in Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E. and Thompson, F. (eds.) *The Prokaryotes: Applied Bacteriology and Biotechnology*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 257-280.

Leahy, D.J., Aukhil, I. and Erickson, H.P. (1996) '2.0 Å Crystal Structure of a Four-Domain Segment of Human Fibronectin Encompassing the RGD Loop and Synergy Region', *Cell*, 84(1), pp. 155-164.

Leduc, M., Fréhel, C., Siegel, E. and Van Heijenoort, J. (1989) 'Multilayered Distribution of Peptidoglycan in the Periplasmic Space of *Escherichia coli*', *Microbiology*, 135(5), pp. 1243-1254.

Lermant, A., Magnanon, A., Silvain, A., Lubrano, P., Lhuissier, M.J., Dury, C., Follenfant, M., Guiot, Z., Christien, G., Coudert, P., Arvor, A., Sportich, M., Vuillaume, G., Delettre, N., Henry, J., Lafon, E., Lhernould, T., Richard, F. and Ismail, A. (2018) 'A thermo-responsive plasmid for biconditional protein expression', *bioRxiv*, p. 289264.

Li, Z. and Demple, B. (1994) 'SoxS, an activator of superoxide stress genes in *Escherichia coli*. Purification and interaction with DNA', *Journal of Biological Chemistry*, 269(28), pp. 18371-7.

Lieb, M. (1981) 'A fine structure map of spontaneous and induced mutations in the lambda repressor gene, including insertions of IS elements', *Mol Gen Genet*, 184(3), pp. 364-71.

Lillard, J.W., Jr., Bearden, S.W., Fetherston, J.D. and Perry, R.D. (1999) 'The haemin storage (Hms+) phenotype of *Yersinia pestis* is not essential for the pathogenesis of bubonic plague in mammals', *Microbiology*, 145 (Pt 1), pp. 197-209.

Lindberg, F., Lund, B., Johansson, L. and Normark, S. (1987) 'Localization of the receptor-binding protein adhesin at the tip of the bacterial pilus', *Nature*, 328(6125), pp. 84-87.

Liochev, S.I. and Fridovich, I. (1992) 'Fumarase C, the stable fumarase of *Escherichia coli,* is controlled by the *soxRS* regulon', *Proc Natl Acad Sci U S A*, 89(13), pp. 5892-6.

Liochev, S.I., Hausladen, A. and Fridovich, I. (1999) 'Nitroreductase A is regulated as a member of the *soxRS* regulon of *Escherichia coli*', *Proceedings of the National Academy of Sciences*, 96(7), p. 3537.

Liu, F., Chen, H., Galván, E.M., Lasaro, M.A. and Schifferli, D.M. (2006) 'Effects of Psa and F1 on the Adhesive and Invasive Interactions of *Yersinia pestis* with Human Respiratory Tract Epithelial Cells', *Infection and Immunity*, 74(10), p. 5636.

Livak, K.J. and Schmittgen, T.D. (2001) 'Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method', *Methods*, 25(4), pp. 402-408.

López-García, P. and Forterre, P. (2000) 'DNA topology and the thermal stress response, a tale from mesophiles and hyperthermophiles', *BioEssays*, 22(8), pp. 738-746.

Lopez, P.J., Guillerez, J., Sousa, R. and Dreyfus, M. (1998) 'On the mechanism of inhibition of phage T7 RNA polymerase by *lac* repressor', *J Mol Biol*, 276(5), pp. 861-75.

Machado Roque, A.I. (2013) *Protein scaffolds for cell culture*. Thesis (Ph. D.)--University of Newcastle upon Tyne, 2013.

Madan Babu, M. and Teichmann, S.A. (2003) 'Functional determinants of transcription factors in *Escherichia coli*: protein families and binding sites', *Trends in Genetics*, 19(2), pp. 75-79.

Madigan, M.T. (2015) *Brock biology of microorganisms*. Fourteenth edition edn. Boston : Pearson.

Mads Kaern, William J. Blake, a. and Collins, J.J. (2003) 'The Engineering of Gene Regulatory Networks', *Annual Review of Biomedical Engineering*, 5(1), pp. 179-206.

Marbach, A. and Bettenbrock, K. (2012) '*lac* operon induction in *Escherichia coli*: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA', *Journal of Biotechnology*, 157(1), pp. 82-88.

Marenne, M.-N., Journet, L., Mota, L.J. and Cornelis, G.R. (2003) 'Genetic analysis of the formation of the Ysc–Yop translocation pore in macrophages by *Yersinia enterocolitica*: role of LcrV, YscF and YopN', *Microbial Pathogenesis*, 35(6), pp. 243-258.

Martin, R.G. and Rosner, J.L. (2002) 'Genomics of the *marA*/*soxS*/*rob* regulon of *Escherichia coli*: identification of directly activated promoters by application of molecular genetics and informatics to microarray data', *Molecular Microbiology*, 44(6), pp. 1611-1624.

McMurry, L.M. and Levy, S.B. (2010) 'Evidence that Regulatory Protein MarA of *Escherichia coli* Represses rob by Steric Hindrance', *The Journal of Bacteriology*, 192(15), p. 3977.

McNicholas, S., Potterton, E., Wilson, K.S. and Noble, M.E.M. (2011) 'Presenting your structures: the CCP4mg molecular-graphics software', *Acta Crystallographica Section D*, 67(4), pp. 386-394.

Michan, C.M., Hyde, E.I. and Busby, S.J.W. (1995) 'The *Escherichia coli* MeIR transcription activator: production of a stable fragment containing the DNA-binding domain', *Nucleic Acids Research*, 23(9), pp. 1518-1523.

Miller, J., Williamson, E.D., Lakey, J.H., Pearce, M.J., Jones, S.M. and Titball, R.W. (1998) 'Macromolecular organisation of recombinant *Yersinia pestis* F1 antigen and the effect of structure on immunogenicity', *FEMS Immunology & Medical Microbiology*, 21(3), pp. 213-221.

Miroux, B. and Walker, J.E. (1996) 'Over-production of Proteins in *Escherichia coli* : Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels', *Journal of Molecular Biology*, 260(3), pp. 289-298.

Mizushima, T., Kataoka, K., Ogata, Y., Inoue, R.I. and Sekimizu, K. (1997) 'Increase in negative supercoiling of plasmid DNA in *Escherichia coli* exposed to cold shock', *Molecular Microbiology*, 23(2), pp. 381-386.

Mohanty, A.K., Mukhopadhyay, U.K., Grover, S. and Batish, V.K. (1999) 'Bovine chymosin: Production by rDNA technology and application in cheese manufacture', *Biotechnology Advances*, 17(2), pp. 205-217.

Moralejo, P., Egan, S.M., Hidalgo, E. and Aguilar, J. (1993) 'Sequencing and characterization of a gene cluster encoding the enzymes for L-rhamnose metabolism in *Escherichia coli*', *Journal of Bacteriology*, 175(17), p. 5585.

Motin, V.L., Georgescu, A.M., Fitch, J.P., Gu, P.P., Nelson, D.O., Mabery, S.L., Garnham, J.B., Sokhansanj, B.A., Ott, L.L., Coleman, M.A., Elliott, J.M., Kegelmeyer, L.M., Wyrobek, A.J., Slezak, T.R., Brubaker, R.R. and Garcia, E. (2004) 'Temporal Global Changes in Gene Expression during Temperature Transition in *Yersinia pestis*', *Journal of Bacteriology*, 186(18), pp. 6298-6305.

Murray, P.R. (2015) *Medical Microbiology*. 8th ed.. edn.: Elsevier Health Sciences.

Nakajima, H., Kobayashi, M., Negishi, T. and Aono, R. (1995) '*soxRS* gene increased the level of organic solvent tolerance in *Escherichia coli*', *Biosci Biotechnol Biochem*, 59(7), pp. 1323-5.

Ni, L., Tonthat, N.K., Chinnam, N. and Schumacher, M.A. (2013) 'Structures of the *Escherichia coli* transcription activator and regulator of diauxie, XylR: an AraC DNA-binding family member with a LacI/GalR ligand-binding domain', *Nucleic Acids Res*, 41(3), pp. 1998-2008.

Nikaido, H. and Vaara, M. (1985) 'Molecular basis of bacterial outer membrane permeability', *Microbiol Rev*, 49(1), pp. 1-32.

Özkan, E., Carrillo, Robert A., Eastman, Catharine L., Weiszmann, R., Waghray, D., Johnson, Karl G., Zinn, K., Celniker, Susan E. and Garcia, K.C. (2013) 'An Extracellular Interactome of Immunoglobulin and LRR Proteins Reveals Receptor-Ligand Networks', *Cell*, 154(1), pp. 228-239.

Perry, R.D. and Fetherston, J.D. (1997) '*Yersinia pestis*-etiologic agent of plague', *Clinical Microbiology Reviews*, 10(1), pp. 35-66.

Perry, R.D., Pendrak, M.L. and Schuetze, P. (1990) 'Identification and cloning of a hemin storage locus involved in the pigmentation phenotype of *Yersinia pestis*', *Journal of Bacteriology*, 172(10), p. 5929.

Phan, G., Remaut, H., Wang, T., Allen, W.J., Pirker, K.F., Lebedev, A., Henderson, N.S., Geibel, S., Volkan, E., Yan, J., Kunze, M.B.A., Pinkner, J.S., Ford, B., Kay, C.W.M., Li, H., Hultgren, S.J., Thanassi, D.G. and Waksman, G. (2011) 'Crystal structure of the FimD usher bound to its cognate FimC-FimH substrate', *Nature*, 474(7349), pp. 49-53.

Pinkner, J.S., Remaut, H., Buelens, F., Miller, E., xc, berg, V., Pemberton, N., Hedenstr, xf, m, M., Larsson, A., Seed, P., Waksman, G., Hultgren, S.J. and Almqvist, F. (2006) 'Rationally Designed Small Compounds Inhibit Pilus Biogenesis in Uropathogenic Bacteria', *Proceedings of the National Academy of Sciences of the United States of America*, 103(47), pp. 17897-17902.

Power, J. (1967) 'The L-rhamnose genetic system in *Escherichia coli* K-12', *Genetics*, 55(3), p. 557.

Quade, N., Mendonca, C., Herbst, K., Heroven, A.K., Ritter, C., Heinz, D.W. and Dersch, P. (2012) 'Structural basis for intrinsic thermosensing by the master virulence regulator RovA of *Yersinia*', *The Journal of biological chemistry*, 287(43), pp. 35796-35803.

Reese, M.G. (2001) 'Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome', *Computers & chemistry*, 26(1), p. 51.

Remaut, H., Tang, C., Henderson, N.S., Pinkner, J.S., Wang, T., Hultgren, S.J., Thanassi, D.G., Waksman, G. and Li, H. (2008) 'Fiber Formation across the Bacterial Outer Membrane by the Chaperone/Usher Pathway', *Cell*, 133(4), pp. 640-652.

Remaut, H. and Waksman, G. (2006) 'Protein–protein interaction through β-strand addition', *Trends in Biochemical Sciences*, 31(8), pp. 436-444.

Revell, P.A. and Miller, V.L. (2000) 'A chromosomally encoded regulator is required for expression of the *Yersinia enterocolitica inv* gene and for virulence', *Molecular Microbiology*, 35(3), pp. 677-685.

Rhee, S., Martin, R.G., Rosner, J.L. and Davies, D.R. (1998) 'A Novel DNA-Binding Motif in MarA: The First Structure for an AraC Family Transcriptional Activator', *Proceedings of the National Academy of Sciences of the United States of America*, 95(18), pp. 10413-10418.

Righetti, F., Nuss, A.M., Twittenhoff, C., Beele, S., Urban, K., Will, S., Bernhart, S.H., Stadler, P.F., Dersch, P. and Narberhaus, F. (2016) 'Temperature-responsive in vitro RNA structurome of *Yersinia pseudotuberculosis*', *Proceedings of the National Academy of Sciences of the United States of America*, 113(26), p. 7237.

Robert, X. and Gouet, P. (2014) 'Deciphering key features in protein structures with the new ENDscript server', *Nucleic acids research*, 42(Web Server issue), p. W320.

Rogers, P., Chen, J.-S. and Zidwick, M.J. (2013) 'Organic Acid and Solvent Production: Acetic, Lactic, Gluconic, Succinic, and Polyhydroxyalkanoic Acids', in Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E. and Thompson, F. (eds.) *The Prokaryotes: Applied Bacteriology and Biotechnology*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 3-75.

Roque, A.I., Soliakov, A., Birch, M.A., Philips, S.R., Shah, D.S. and Lakey, J.H. (2014) 'Reversible non-stick behaviour of a bacterial protein polymer provides a tuneable molecular mimic for cell and tissue engineering', *Adv Mater*, 26(17), pp. 2704-9, 2616.

Runco, L.M., Myrczek, S., Bliska, J.B. and Thanassi, D.G. (2008) 'Biogenesis of the Fraction 1 Capsule and Analysis of the Ultrastructure of Yersinia pestis', *The Journal of Bacteriology*, 190(9), p. 3381.

Sauer, F.G., Barnhart, M., Choudhury, D., Knight, S.D., Waksman, G. and Hultgren, S.J. (2000a) 'Chaperone-assisted pilus assembly and bacterial attachment', *Current Opinion in Structural Biology*, 10(5), pp. 548-556.

Sauer, F.G., Fütterer, K., Pinkner, J.S., Dodson, K.W., Hultgren, S.J. and Waksman, G. (1999) 'Structural Basis of Chaperone Function and Pilus Biogenesis', *Science*, 285(5430), pp. 1058-1061.

Sauer, F.G., Knight, S.D., Waksman and, G.J. and Hultgren, S.J. (2000b) 'PapD-like chaperones and pilus biogenesis', *Seminars in Cell & Developmental Biology*, 11(1), pp. 27-34.

Sauer, F.G., Remaut, H., Hultgren, S.J. and Waksman, G. (2004) 'Fiber assembly by the chaperone-usher pathway', *Biochim Biophys Acta*, 1694(1-3), pp. 259-67.

Scaletsky, I.C.A., Michalski, J., Torres, A.G., Dulguer, M.V. and Kaper, J.B. (2005) 'Identification and characterization of the locus for diffuse adherence, which encodes a novel afimbrial adhesin found in atypical enteropathogenic *Escherichia coli*', *Infection and Immunity*, 73(8), p. 4753.

Schaefer, B.C. (1995) 'Revolutions in Rapid Amplification of cDNA Ends: New Strategies for Polymerase Chain Reaction Cloning of Full-Length cDNA Ends', *Analytical Biochemistry*, 227(2), pp. 255-273.

Schleif, R. (2000) 'Regulation of the l-arabinose operon of *Escherichia coli*', *Trends in Genetics*, 16(12), pp. 559-565.

Schleif, R. 34 (2010) 'AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action'. Oxford, UK, pp. 779-796.

Schneiders, T., Barbosa, T., McMurry, L. and Levy, S. (2004) 'The *Escherichia coli* Transcriptional Regulator MarA Directly Represses Transcription of purA and hdeA', *Journal of Biological Chemistry*, 279(10), pp. 9037-9042.

Schneiders, T. and Levy, S.B. (2006) 'MarA-mediated Transcriptional Repression of the *rob* Promoter', *Journal of Biological Chemistry*, 281(15), pp. 10049-10055.

Shapiro, R.S. and Cowen, L.E. (2012) 'Thermal control of microbial development and virulence: molecular mechanisms of microbial temperature sensing', *mBio*, 3(5).

Siegele, D.A. and Hu, J.C. (1997) 'Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents

mixed populations', *Proceedings of the National Academy of Sciences*, 94(15), p. 8168.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D. and Higgins, D.G. (2011) 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Molecular Systems Biology*, 7(1), pp. n/a-n/a.

Simpson, W.J., Thomas, R.E. and Schwan, T.G. (1990) 'Recombinant capsular antigen (fraction 1) from *Yersinia pestis* induces a protective antibody response in BALB/c mice', *Am J Trop Med Hyg*, 43(4), pp. 389-96.

Slonczewski, J. (2011) *Microbiology : an evolving science*. 2nd ed.. edn. New York ; London : W.W. Norton.

Smeekens, S. and Romano, L. (1986) 'Promoter and nonspecific DNA binding by the T7 RNA polymerase', *Nucleic Acids Research*, 14(6), pp. 2811-2827.

Soisson, S.M., MacDougall-Shackleton, B., Schleif, R. and Wolberger, C. (1997) 'Structural basis for ligand-regulated oligomerization of AraC', *Science*, 276(5311), pp. 421-5.

Soliakov, A., Harris, J.R., Watkinson, A. and Lakey, J.H. (2010) 'The structure of *Yersinia pestis* Caf1 polymer in free and adjuvant bound states', *Vaccine*, 28(35), pp. 5746-5754.

Solovyev, V. and Salamov, A. (2011) *Automatic annotation of microbial genomes and metagenomic sequences*.

Spangler, B.D. (1992) 'Structure and function of cholera toxin and the related *Escherichia coli* heat-labile enterotoxin', *Microbiological Reviews*, 56(4), p. 622.

Stock, J.B., Rauch, B. and Roseman, S. (1977) 'Periplasmic space in *Salmonella typhimurium* and *Escherichia coli*', *Journal of Biological Chemistry*, 252(21), pp. 7850-61.

Suomalainen, M., Haiko, J., Kukkonen, M., Korhonen, T.K., Lähteenmäki, K., Virkola, R., Westerlund-Wikström, B., Lobo, L. and Ramu, P. (2007) 'Using Every Trick in the Book: The Pla Surface Protease of *Yersinia pestis*', in Perry, R.D. and Fetherston, J.D. (eds.) *The Genus Yersinia: From Genomics to Function*. New York, NY: Springer New York, pp. 268-278.

Sutherland, I.W. (2001) 'Microbial polysaccharides from Gram-negative bacteria', *International Dairy Journal*, 11(9), pp. 663-674.

Tate, C.G., Muiry, J.A. and Henderson, P.J. (1992) 'Mapping, cloning, expression, and sequencing of the *rhaT* gene, which encodes a novel L-rhamnose-H+ transport protein in *Salmonella typhimurium* and *Escherichia coli*', *Journal of Biological Chemistry*, 267(10), pp. 6923-6932.

Thanassi, D.G., Saulino, E.T. and Hultgren, S.J. (1998) 'The chaperone/usher pathway: a major terminal branch of the general secretory pathway', *Current Opinion in Microbiology*, 1(2), pp. 223-231.

Ulusu, Y., Dura, G., Waller, H., Benning, M.J., Fulton, D.A., Lakey, J.H. and Peters, D.T. (2017) 'Thermal stability and rheological properties of the 'non-stick' Caf1 biomaterial', *Biomedical Materials*, 12(5), p. 051001.

Vadyvaloo, V., Sebbane, F., Hinnebusch, B.J., Sturdevant, D. and Jarrett, C. (2007) 'Analysis of *Yersinia pestis* Gene Expression in the Flea Vector', in Perry, R.D. and Fetherston, J.D. (eds.) *The Genus Yersinia: From Genomics to Function*. New York, NY: Springer New York, pp. 192-200.

Via, P., Badia, J., Baldoma, L., Obradors, N. and Aguilar, J. (1996) 'Transcriptional regulation of the *Escherichia coli rhaT* gene', *Microbiology*, 142 (Pt 7), pp. 1833-40.

Vitagliano, L., Ruggiero, A., Pedone, C. and Berisio, R. (2008) 'Conformational states and association mechanism of *Yersinia pestis* Caf1 subunits', *Biochemical and Biophysical Research Communications*, 372(4), pp. 804-810.

Vorontsov, E.D., Dubichev, A.G., Serdobintsev, L.N. and Naumov, A.V. (1990) 'Association-dissociation processes and supermolecular organisation of the capsule antigen (protein F1) of *Yersinia pestis*', *Biomedical science*, 1(4), p. 391.

Waksman, G. and Hultgren, S.J. (2009) 'Structural biology of the chaperone-usher pathway of pilus biogenesis', *Nature Reviews: Microbiology*, 7(11), pp. 765-774.

Waldminghaus, T., Fippinger, A., Alfsmann, J. and Narberhaus, F. (2005) 'RNA thermometers are common in alpha- and gamma-proteobacteria', *Biological chemistry*, 386(12), p. 1279.

Wan, Y., Qu, K., Ouyang, Z., Kertesz, M., Li, J., Tibshirani, R., Makino, Debora L., Nutter, Robert C., Segal, E. and Chang, Howard Y. (2012) 'Genome-wide measurement of RNA folding energies', *Molecular Cell*, 48(2), pp. 169-181.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R. and Schwede, T. (2018) 'SWISS-MODEL: homology modelling of protein structures and complexes', *Nucleic acids research*, 46(W1), p. W296.

Welkos, S.L., Andrews, G.P., Lindler, L.E., Snellings, N.J. and Strachan, S.D. (2004) 'Mu dI1(Ap lac) mutagenesis of *Yersinia pestis* plasmid pFra and identification of temperature-regulated loci associated with virulence', *Plasmid*, 51(1), pp. 1-11.

Willey, J.M. (2014) *Prescott's microbiology*. Ninth edition. International edition. edn. New York, NY : McGraw Hill Education.

Willey, J.M., Sherwood, L. and Woolverton, C.J. (2008) *Prescott, Harley, and Klein's Microbiology*. McGraw-Hill Higher Education.

Williamson, E.D. and Oyston, P.C.F. (2013) 'Protecting against plague: towards a next-generation vaccine', *Clinical & Experimental Immunology*, 172(1), pp. 1-8.

Wilson, B.A. and Salyers, A.A. (2011) *Bacterial pathogenesis : a molecular approach*. 3rd ed.. edn. Washington, DC: Washington, DC : ASM Press.

Yu, X., Visweswaran, Ganeshram R., Duck, Z., Marupakula, S., MacIntyre, S., Knight, Stefan D. and Zavialov, Anton V. (2009) 'Caf1A usher possesses a Caf1 subunit-like domain that is crucial for Caf1 fibre secretion', *Biochemical Journal*, 418(3), p. 541.

Zav'yalov, V.P., Chernovskaya, T.V., Chapman, D.A., Karlyshev, A.V., MacIntyre, S., Zavialov, A.V., Vasiliev, A.M., Denesyuk, A.I., Zav'yalova, G.A., Dudich, I.V., Korpela, T. and Abramov, V.M. (1997) 'Influence of the conserved disulphide bond, exposed to the putative binding pocket, on the structure and function of the immunoglobulin-like molecular chaperone Caf1M of *Yersinia pestis*', *The Biochemical journal*, 324 (Pt 2)(Pt 2), pp. 571-578.

Zav'yalova, G.A., Zav'yalov, V.P., Denesyuk, A.I. and Korpela, T. (1995) 'Modelling of steric structure of a periplasmic molecular chaperone Caf1M of *Yersinia pestis*, a prototype member of a subfamily with characteristic

structural and functional features', *FEMS Immunology & Medical Microbiology*, 11(1), pp. 19-24.

Zavialov, A., Zav'yalova, G., Korpela, T. and Zav'yalov, V. (2007) 'FGL chaperone-assembled fimbrial polyadhesins: anti-immune armament of Gram-negative bacterial pathogens', *FEMS Microbiology Reviews*, 31(4), pp. 478-514.

Zavialov, A.V., Berglund, J., Pudney, A.F., Fooks, L.J., Ibrahim, T.M., Macintyre, S. and Knight, S.D. (2003) 'Structure and Biogenesis of the Capsular F1 Antigen from *Yersinia pestis*: Preserved Folding Energy Drives Fiber Formation', *Cell*, 113(5), pp. 587-596.

Zavialov, A.V., Kersley, J., Korpela, T., Zav'yalov, V.P., MacIntyre, S. and Knight, S.D. (2002) 'Donor strand complementation mechanism in the biogenesis of non-pilus systems', *Molecular Microbiology*, 45(4), pp. 983-995.

Zavialov, A.V. and Knight, S.D. (2007) 'A novel self-capping mechanism controls aggregation of periplasmic chaperone Caf1M', *Mol Microbiol*, 64(1), pp. 153-64.

Zavialov, Anton V., Tischenko, Vladimir M., Fooks, Laura J., Brandsdal, Bjørn O., Åqvist, J., Zav, yalov, Vladimir P., Macintyre, S. and Knight, Stefan D. (2005) 'Resolving the energy paradox of chaperone/usher-mediated fibre assembly', *Biochemical Journal*, 389(3), p. 685.

Zavyalov, V.P., Chernovskaya, T., Chapman, D., Karlyshev, A., Macintyre, S., Zavialov, A., Vasiliev, A., Denesyuk, A.I., Zavyalova, G., Dudich, I.V., Korpela, T. and Abramov, V. (1997) 'Influence of the conserved disulphide bond, exposed to the putative binding pocket, on the structure and function of the immunoglobulin-like molecular chaperone Caf1M of *Yersinia pestis*', *Biochemical Journal*, 324, pp. 571-578.

Zhang, Y. and Frohman, M.A. (1997) 'Using rapid amplification of cDNA ends (RACE) to obtain full-length cDNAs', *Methods in molecular biology (Clifton, N.J.)*, 69, pp. 61-87.

Appendix

Al-Jawdah et al. BMC Microbiology (2019) 19:68
https://doi.org/10.1186/s12866-019-1444-4

BMC Microbiology

RESEARCH ARTICLE Open Access

Induction of the immunoprotective coat of *Yersinia pestis* at body temperature is mediated by the Caf1R transcription factor

Abdulmajeed D. Al-Jawdah, Iglika G. Ivanova, Helen Waller, Neil D. Perkins, Jeremy H. Lakey and Daniel T. Peters*

Abstract

Background: Thermal regulation of gene expression occurs in many microorganisms, and is mediated via several typical mechanisms. *Yersinia pestis* is the causative agent of the plague and spreads by zoonotic transfer from fleas to mammalian blood with a concomitant rapid temperature change, from ambient to 37 °C, which induces the expression of capsular antigen (Caf1) that inhibits phagocytosis. Caf1 is formed into long polymeric fimbriae by a periplasmic chaperone (Caf1M) and outer membrane usher (Caf1A). All three are encoded on an operon regulated by an AraC-type transcription factor Caf1R. The aim of this study was to determine the role of Caf1R in the thermal control of *caf1* operon gene expression.

Results: PCR analysis of cDNA demonstrated that the genes of the operon are transcribed as a single polycistronic mRNA. Bioinformatic analysis, supported by deletion mutagenesis, then revealed a region containing the promoter of this polycistronic transcript that was critical for Caf1 protein expression. Caf1R was found to be essential for Caf1 protein production. Finally, RT-PCR analysis and western blot experiments showed large, Caf1R dependent increases in *caf1* operon transcripts upon a shift in temperature from 25 °C to 35 °C.

Conclusions: The results show that thermal control of Caf1 polymer production is established at the transcriptional level, in a Caf1R dependent manner. This gives us new insights into how a virulent pathogen evades destruction by the immune system by detecting and responding to environmental changes.

Keywords: Gene expression regulation, *Yersinia pestis*, AraC transcription factor, Temperature, Fimbriae, bacterial

Background

Temperature is an important stimulus for pathogens infecting mammalian and avian hosts, allowing them to differentiate between host and non-host environments. Microorganisms can employ a suite of different mechanisms to sense and respond to changes in temperature by altering their patterns of gene expression. These mechanisms can act at either the transcriptional or translational levels and can be mediated by protein, RNA or DNA [1]. Typical mechanisms include changes in the degree of supercoiling in DNA, as well as elements of local structure such as promoter curvature, which can affect transcription efficiency, when temperature increases or falls [1]. Another common mechanism involves RNA "thermometers" that form temperature dependent mRNA secondary structures that can inhibit translation below a critical temperature [1, 2]. In addition, some proteins can undergo conformational changes in response to alterations in temperature, resulting in the post-translational modification of downstream transcription factors [3, 4], inhibition of DNA binding [5, 6], or protein degradation [7], with subsequent changes in gene expression.

The bacterium *Yersinia pestis* is the etiologic agent of the plague, responsible for approximately 200 million deaths across the course of three major epidemics throughout history [8], with modern outbreaks still occurring, most recently in Madagascar [9]. One of the factors responsible for the virulence of this organism is its ability to avoid phagocytosis by macrophages [10]. This ability is dependent upon a gel-like capsule, composed of polymers of the capsular antigen fraction 1

* Correspondence: daniel.peters@ncl.ac.uk
Institute for Cell and Molecular Biosciences, Medical School, Newcastle University, Newcastle upon Tyne, UK

(Caf1) protein, which coats the bacterium and is thought to act by preventing adhesin-receptor interactions. Deletion of the capsule causes an increase in the uptake of *Y. pestis* bacilli by macrophages [10] showing Caf1 to be an important protection against the host's immune system.

Caf1 forms very long, thin polymers consisting of repeating ~ 15 kDa subunits with polymer lengths of up to 1.5 μm having been observed [11]. The polymers are highly flexible, appearing at high magnifications like beads on a string. Their biogenesis proceeds via the chaperone-usher (CU) pathway, which is employed for the production of a range of pilus structures in Gram negative bacteria [12]. Briefly, nascent, unfolded Caf1 subunits are exported to the periplasm, where they are bound by the chaperone protein Caf1M. This partially stabilises the Ig-like fold of the Caf1 subunit, before it is delivered to the usher protein, Caf1A, which resides in the outer membrane. Caf1A assembles the Caf1 subunits into a polymer that it exports to the cell surface [12] where it appears as a diffuse "flocculent" layer above the pellet in centrifuged cultures [13, 14]. The molecular basis for the polymerisation is the donation of the N-terminal β strand of one subunit into the acceptor cleft of the preceding subunit in the polymer, displacing the chaperone and stably completing the subunit fold [15]. This process, known as donor-strand complementation, results in a highly stable, non-covalent polymer [16].

In *Y. pestis*, Caf1 is expressed from the *caf1* operon present on the pMT plasmid [17]. Upstream of the *caf1M*, *caf1A* and *caf1* genes of the *caf1* operon is a fourth gene, *caf1R*, which is encoded on the opposite DNA strand. The domain structure of the Caf1R protein is that of a Rob-like AraC family transcription factor [18], consisting of a recognisable N-terminal helix-turn-helix DNA binding domain but with a distinctive C-terminal domain of unknown function. Although a putative promoter site and regulatory elements have been predicted upstream of *caf1M* [19], their presence has yet to be empirically determined.

Y. pestis becomes vulnerable to macrophage attack when it is injected into the mammalian host via a flea bite. The bacterium senses this event through a range of molecular mechanisms that perceive the temperature of its environment and exploit the temperature difference between the flea (on the hosts surface ~ 25 °C) and blood (~ 37 °C) [8]. Caf1 protein expression is induced at 37 °C [8, 10, 20], allowing the bacteria to rapidly produce this protein when most at risk of being phagocytosed. Temperature dependent increases in *caf1* operon gene transcription have been reported [20, 21] but whether Caf1R plays a role in coordinating the correct temperature dependent expression is not clear.

A previous report has described the Caf1R protein as a positive transcriptional regulator [22]. There are three arguments in favour of this designation: first, a plasmid containing only the 81 N-terminal residue helix-turn-helix DNA binding domain of Caf1R caused an increase in Caf1 production [22, 23]; second, inducing frame-shift mutations into Caf1R prevented Caf1 biogenesis; and finally, Caf1R displays homology with the AraC family of positive transcriptional regulators [22]. However, a link between Caf1R and the thermal switch has not been demonstrated.

Here, we determine the role of Caf1R in the thermosensitive control of gene expression from the *caf1* operon. We show that the *caf1M*, *caf1A* and *caf1* genes are transcribed as a single polycistronic transcript and identify the promoter region that is critical for their transcription. Deletion of *caf1R* completely abrogated Caf1 polymer expression, which was restored by complementing this deletion with *caf1R* on a second plasmid. Finally, RT-PCR and western blot experiments showed *caf1* operon transcripts are substantially upregulated upon a switch from 25 °C to 35 °C in a Caf1R dependent manner. Therefore, we propose that the temperature induced expression of Caf1 polymers is governed at the transcriptional level in a Caf1R dependent manner, thus revealing Caf1R to be a key factor in the determination of *Y. pestis* virulence.

Results

The *caf1* operon is transcribed as a polycistronic transcript

In the *caf1* operon, the *caf1R* coding sequence is on the opposite strand to, and separated by a 328 nucleotide (nt) intergenic region I, from the remaining genes, which are arranged with a very short (24 nt) intergenic region II between *caf1M* and *caf1A* and a larger intergenic region III of 80 nt between *caf1A* and *caf1* (Fig. 1a). To characterise the expression of *caf1* genes within the *caf1* operon, PCR analysis of the RNA transcripts was employed. Cultures of *E. coli* transformed with a plasmid containing the *caf1* operon (**pC**af1**OP**eron, pCOP, Additional file 1: Table S1) or the operon preceded by an upstream T7 promoter from the vector (pT7-COP, Additional file 1: Table S1), were grown from single colonies for 16 h at 35 °C to stimulate Caf1 production, following which total RNA was extracted and used to synthesise cDNA. To assess whether the *caf1* genes are expressed as part of a polycistronic transcript, the cDNA was used as a template, with primers complementary to the coding regions either side of the intergenic regions (Additional file 1: Figure S1). Therefore, if the genes are expressed on a large

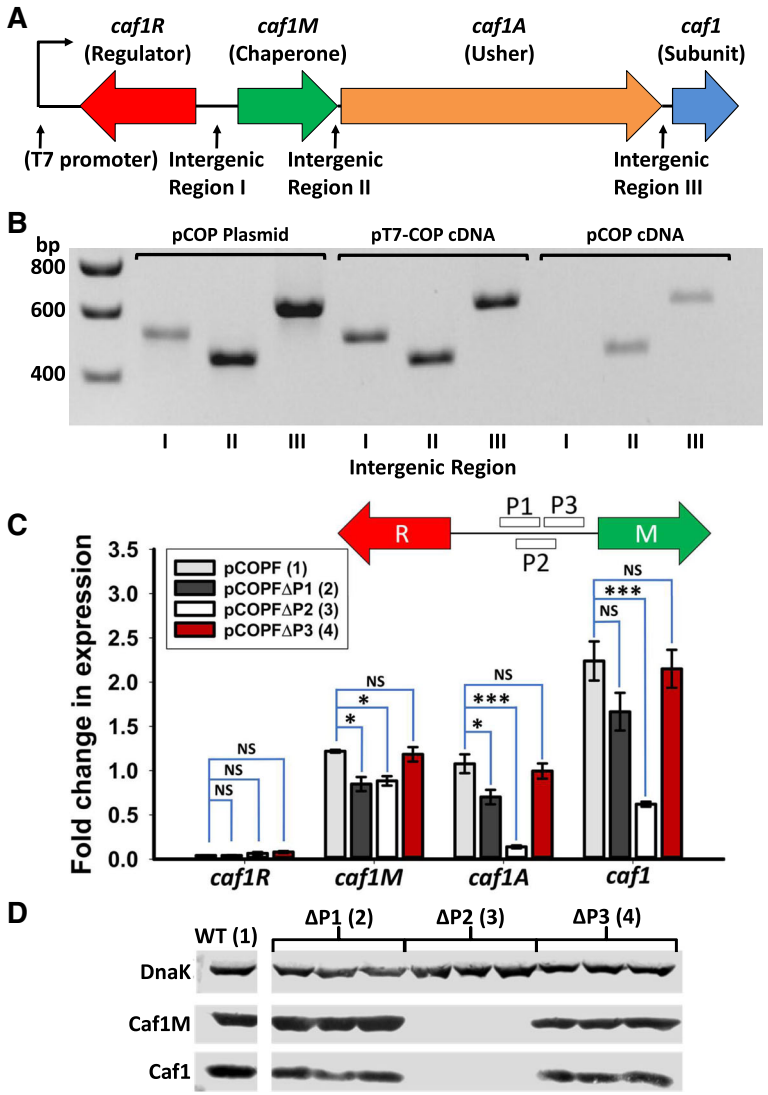

Fig. 1 Determination of *caf1* transcript size and promoter site. **a** Organisation of the *caf1* operon. The position of the T7 promoter, found in the artificial pT7-COP plasmid is highlighted. **b** Agarose gel showing the PCR amplification products corresponding to the three intergenic regions of the *caf1* operon generated from either the pCOP plasmid, or cDNA made using mRNA from cultures of *E. coli* transformed with pT7-COP or pCOP plasmids and grown at 35 °C for 16 h. Both plasmids contain the *caf1* operon as shown in (**a**). **c** Transcript levels of the *caf1* operon genes as determined by RT-PCR, from cultures grown for 16 h of *E. coli* transformed with either pCOPF (full *caf1* operon, Caf1R,M and A have FLAG tags), or pCOPF with the proposed promoter regions P1,2 or 3 deleted. Three cultures of each condition were grown, with RT-PCR reactions run in duplicate for each culture. Bar heights correspond to mean fold-change in expression relative to β-lactamase. Error bars represent standard error of the mean (S.E.M) from three biological replicates. Asterisks represent significant differences between groups (* - $P < 0.05$, ** - $P < 0.01$, *** - $P < 0.001$, NS – not significant, determined by ANOVA with Holm- Šidák *post-hoc* test). A diagram detailing the positions of the P1, 2 and 3 regions is shown in the top right of the graph. **d** Western blot of the above cultures showing the levels of Caf1M and Caf1 (detected by anti-FLAG tag and anti-Caf1 antibodies respectively), and using DnaK (detected by an anti-DnaK antibody) as a loading control

transcript, amplification of the intergenic regions should proceed when the cDNA is used as a template, whereas single transcripts will result in no amplification. Amplification products could be detected from intergenic regions II and III from pCOP, showing that *caf1M* and *caf1A*, as well as *caf1A* and *caf1*, are co-transcribed (Fig. 1b). The intergenic region between *caf1M* and *caf1A* is very small (24 nt), and unlikely to harbour a promoter. This suggests that in the natural system, *caf1M*, *caf1A* and *caf1* are all transcribed together on one mRNA of approx. 3900 nt. For pT7-COP, all three intergenic regions could be detected, suggesting that the T7 promoter, which is upstream of *caf1R*, drives transcription that proceeds through all four genes. This was of interest since we had previously found by chance that leaky expression from this T7 promoter was able to drive significant Caf1 polymer production [13].

Identification of the long *caf1* operon transcript promoter

To identify the promoter that drives transcription of this long mRNA, the DNA sequence between *caf1R* and *caf1M* (intergenic region I) was used as an input into the BPROM (Softberry Inc., Mount Kisco, NY, USA, http://www.softberry.com) [24] and Neural Network Promoter Prediction 2.2 (NNPP2.2) [25] webservers. These predictions led us to define three potential promoter regions of 41 bp (P1, P2 and P3, Additional file 1: Figure S2) upstream of *caf1M* that could potentially drive the transcription of the polycistronic *caf1* transcript. Interestingly, the P2 region contains the putative promoter previously predicted by Galyov et al. [23]. Each region was then separately deleted from pCOPF (in which Caf1R, M and A have added C-terminal FLAG tags, Additional file 1: Table S1) and the resulting plasmids used to transform BL21(DE3) cells. To assess the effect of the deletions, protein levels after 16 h of growth at 37 °C were determined by western blot using anti-FLAG antibodies and levels of each transcript were measured by RT-PCR (Fig. 1c,d). Deletion of the P3 region showed little effect on the levels of protein or transcript. For P1, deletion had little effect on protein expression, but did appear to cause a small drop in transcript level. However, the greatest effect was seen upon deletion of the P2 region, where synthesis of all the Caf1 proteins was completely abrogated, accompanied by a substantial decrease in *caf1A* and *caf1* transcript levels. The levels of *caf1M* transcript in the P1 and P2 deletions are both slightly reduced, but whereas Caf1M synthesis is unaffected by P1 deletion, no Caf1M protein was detected from the P2 delection. The deleted P2 region thus contains or overlaps the promoter responsible for the large, polycistronic transcript observed by PCR analysis.

Caf1R regulates the *caf1* operon

To further investigate the importance of Caf1R to Caf1 polymer production, the *caf1R* gene was deleted from the pCOPF plasmid (pCOPΔR, Additional file 1: Table S1). The pCOPF and pCOPΔR plasmids were transformed into BL21(DE3) *E. coli* cells and expression cultures were grown from single colonies at 35 °C. Expression of Caf1 polymers correlates with the appearance of a flocculent layer that sits on top of the cell pellet in centrifuged expression cultures [13, 14]. Expression cultures were assayed for flocculent layer production, which directly corresponded to the amount of Caf1 polymer secreted by the bacteria (Additional file 1: Figure S3). The *caf1R* deletion completely abolished Caf1 protein production from the pCOPΔR plasmid, with no flocculent layer observed and no Caf1 protein detectable in the pellet by western blot (Fig. 2), demonstrating that *caf1R* is necessary for Caf1 production.

pCOPFΔR was then co-transformed into BL21(DE3) cells with pBad-Caf1R (Additional file 1: Table S1), where *caf1R* is under the control of an arabinose inducible promoter. Complementing pCOPFΔR with pBad-Caf1R in the absence of arabinose rescued Caf1 protein expression (Fig. 2). Induction of Caf1R from the pBad Caf1R by the addition of arabinose also rescued Caf1 protein expression, although the height of the flocculent layer was lower in this case suggesting that only very low levels of Caf1R are needed to promote Caf1 protein expression, and that higher levels are inhibitory to Caf1 biogenesis. These results show that Caf1R is essential for Caf1 production.

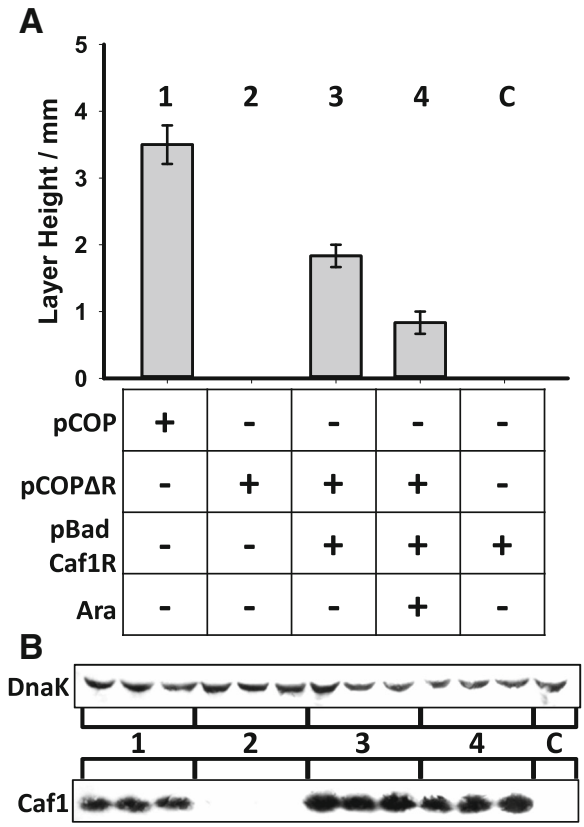

Fig. 2 Caf1 protein production is dependent upon low levels of Caf1R. **a** Graph of the flocculent layer height obtained from *E. coli* cultures containing either pCOP (full *caf1* operon, 1), pCOPΔR (*caf1* operon lacking Caf1R, 2), a co-transformation of pCOPΔR and pBad Caf1R (arabinose inducible Caf1R) without (3) and with (4) 0.005% *w/v* arabinose added to the culture, and pBad Caf1R only control (C). Cultures were grown for 16 h at 35 °C. Bar heights correspond to the mean flocculent layer height obtained from three separate cultures where error bars represent the standard error of the mean (S.E.M). **b** Western blot of three separate cell pellets from the above cultures, probed with an anti-Caf1 antibody. DnaK was used as a loading control, probed with an anti-DnaK antibody

Temperature induction of *caf1* operon expression occurs at the transcriptional level

It is useful for *Y. pestis* to control the formation of the Caf1 polymer using temperature in order to provide

immediate protection from macrophages upon transfer to a warm blooded host. The increase in *caf1* operon expression with temperature has been demonstrated previously [8, 20, 21], but the underlying molecular mechanism has not been investigated. Therefore, we tested the hypothesis that the Caf1R protein is the transcription factor that mediates the thermoregulation of Caf1 expression.

To assess this claim, *E. coli* were transformed with pCOPF or pCOPFΔR and grown for 16 h at 25 °C, such that *caf1* is not expressed. The stationary phase cultures were then diluted to 0.5 OD_{600} and then grown for one further hour at either 25 °C or 35 °C. Transcriptional levels of the *caf1* operon genes were then examined through RT-PCR. A large, significant increase in the relative transcript levels of *caf1M*, *caf1A* and *caf1* could be seen at 35 °C compared to 25 °C for cells transformed with pCOP, with fold increases of 23.3, 33.4 and 26.7 respectively (Fig. 3a). An 8.8 fold increase in *caf1R* transcript levels was also observed, but this was not statistically significantly different from the levels observed at 25 °C. Crucially, in the absence of Caf1R, no temperature dependent increase was seen (Fig. 3b), showing that the increase in transcription levels was not a general effect of temperature change. Additionally, the overall level of transcript produced from pCOPFΔR was significantly lower in all cases when compared to the system containing all four genes, again highlighting the influence of Caf1R.

The effect of the temperature change was then analysed at the translational level by performing western blot experiments on the same cultures used for RT-PCR. The levels of protein detected matched the RT-PCR data, where levels of Caf1M and Caf1 protein could be seen to increase upon incubation at the higher temperature relative to the lower temperature, while being completely undetectable in the absence of Caf1R (Fig. 3). Caf1A, a large membrane protein, was not visualised by western blot. Similarly, although Caf1R is soluble and of a similar size to the easily detected Caf1M, it was also not detected by anti-FLAG western blot, presumably in this case because its concentration was consistently below the limit of detection. This is not surprising, as transcription factors are often present in very low amounts in bacterial cells [26, 27], and our data show that even the low levels of expression from the uninduced, tightly regulated P_{BAD} promoter [28] are enough to rescue Caf1R deletion (Fig. 2).

When *E. coli* transformed with the pT7-COP plasmid were grown at 30 °C (where the natural Caf1R system is switched off and transcription is driven by the T7 promoter), high levels of Caf1 expression could still be detected (Fig. 4). The T7 promoter appears to be stronger than the natural *caf1R* promoter, since it generates a

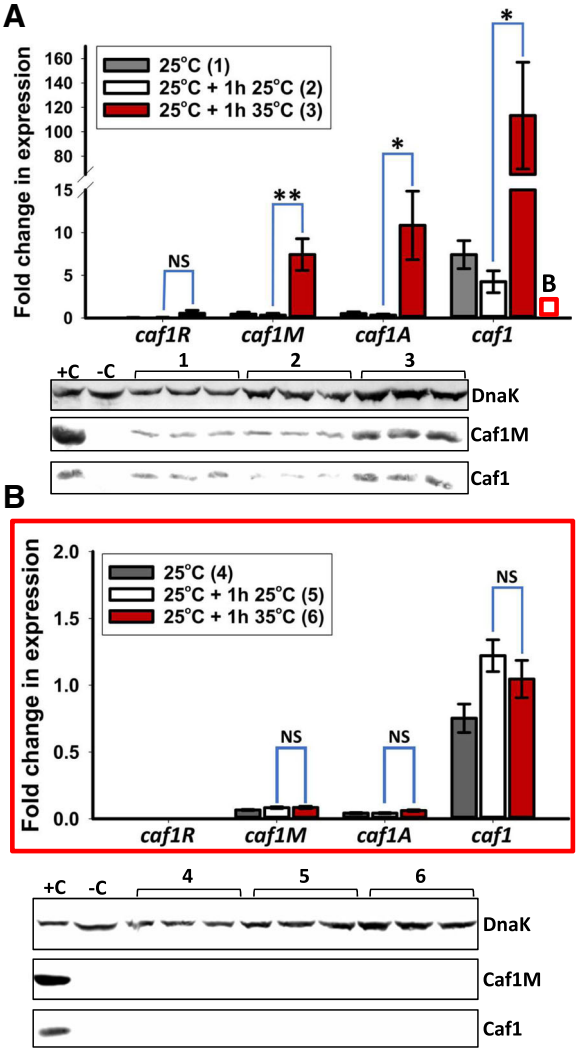

Fig. 3 *caf1* operon gene expression is temperature and Caf1R dependent. Transcript levels, determined by RT-PCR, of each *caf1* gene are shown for cultures of *E. coli* transformed with pCOPF (full *caf1* operon, (**a**)) and pCOPF**Δ**R (*caf1* operon, Caf1R deleted, (**b**)), grown at 25 °C overnight (~ 16 h), then either 25 °C or 35 °C for 1 further hour. The red box in **a** shows the approximate scale of the Y-axis in **b**. Three cultures of each condition were grown, with RT-PCR reactions run in duplicate for each culture. Bar heights correspond to mean fold-change in expression relative to β-lactamase. Error bars represent standard error of the mean (S.E.M) from three biological replicates. Asterisks represent significant differences between groups (* - $P < 0.05$, ** - $P < 0.01$, *** - $P < 0.001$, NS – not significant, determined by ANOVA with Holm- Šidák *post-hoc* test). Western blots showing the levels of Caf1M and Caf1 (detected by anti-FLAG tag and anti-Caf1 antibodies respectively) in the cell pellets of the expression cultures are displayed underneath each graph, using DnaK probed with an anti-DnaK antibody as a loading control. +C and –C represent the pellets of BL21(DE3) cells transformed with pCOPF and untransformed respectively, and grown for 16 h at 35 °C

greater amount of flocculent layer. Differences in flocculent height for pT7-COP containing cells are likely due to differences in the efficiency of T7 transcription between the two temperatures [29], and also the thermal activation of *caf1* operon transcription from the native promoter at 35 °C. The relative thermal independence of the T7 driven expression implies that the temperature regulation of the *caf1* operon takes place primarily at the transcriptional level, with post-transcriptional (translational and post-translational) regulation having very little, if any, role.

Discussion

caf1 operon genes are expressed from a polycistronic transcript

The results of this study show that the *caf1* operon is expressed as a long, polycistronic transcript from a promoter that resides in the P2 region in a Caf1R dependent manner. Previously, Galyov et al. predicted the presence of regulatory regions and a putative promoter that would drive *caf1M* transcription, and identified a Shine-Dalgarno sequence 7 bases upstream of the Caf1M start codon [23]. We used an independent software tool that identified potential promoter regions that could transcribe the polycistronic *caf1* operon transcript, before confirming their activity through deletion mutagenesis and RT-PCR. The putative promoter predicted by Galyov et al. overlaps with the P2 region we identify here as being essential for the expression of the operon. Therefore, we have independently identified and experimentally confirmed the presence of the promoter responsible for transcription of the *caf1* operon.

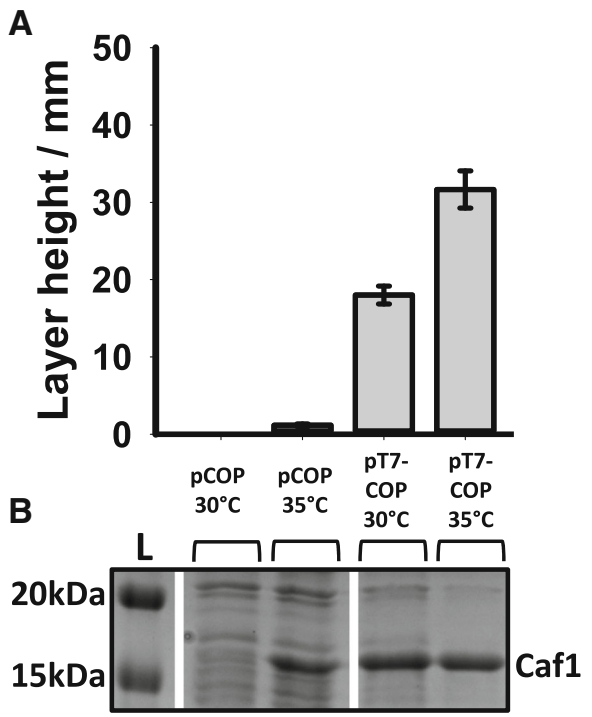

Fig. 4 Effect of a T7 promoter on Caf1 production at 30 °C and 35 °C. *E. coli* cells were transformed with either pCOP (full *caf1* operon) or pT7-COP (T7- full *caf1* operon) and grown at either 30 °C or 35 °C for 16 h, after which the flocculent layer height was measured (**a**) and analysed by SDS-PAGE (**b**). Error bars represent standard error of the mean (S.E.M) from three replicate cultures. The position of the band corresponding to Caf1 is highlighted next to the gel

Considering *caf1M*, *caf1A* and *caf1* are transcribed as a single polycistronic transcript, it was surprising that the RT-PCR results showed that each of the genes has a particular transcriptional level, where the levels of *caf1M* and *caf1A* transcripts were similar and *caf1* transcript much higher. There are two potential explanations of this: either the *caf1* gene can be transcribed by a second, Caf1R responsive promoter, or the translational units within the polycistronic RNA transcript have different levels of mRNA stability. We used 5′ RACE (**R**apid **A**mplification of **c**DNA **E**nds) to search for a second promoter upstream of the *caf1* gene. If a promoter exists in this region, the majority of transcript sequences should terminate inside intergenic region III at the transcriptional start site (Additional file 1: Figure S4), but 7 out of 8 sequences obtained continued through the intergenic region into *caf1A* (Additional file 1: Figure S5), thus corresponding to the observed polycistronic transcript. These sequences terminated at different locations, implying the termination is due to dissociation of the polymerase enzyme rather than a transcriptional start site within this region of *caf1A*. Moreover, deletion of the P2 region, which lies upstream of *caf1M* and is therefore responsible for transcription of the polycistronic transcript, abolished the expression of all *caf1* genes. If separate promoters were present for each gene, this deletion would have a substantially reduced effect on *caf1A* and *caf1* expression. A previous study by Galyov et al. [19] predicted the presence of a putative promoter upstream of *caf1*, at the 3′ end of the *caf1A* gene. However, in isolation this region did not enable *caf1* expression [19], suggesting that it does not contain a functional promoter, in concordance with our results.

As a second promoter could not be detected, the differential stability of the translational units within the mRNA, such as has been previously reported [30], is a more plausible explanation for the differing levels of transcript observed in the RT-PCR experiments. Upon a shift of the bacteria to 35 °C, such a system would allow Caf1R to activate the transcription of the *caf1* operon polycistronic mRNA, which would then be degraded to different degrees resulting in differential levels of transcript for *caf1M*, *caf1A* and *caf1*. To produce long Caf1 polymers, a large number of subunits is required, with fewer chaperones and ushers necessary. Therefore, this

system appears to allow cells to produce the optimal levels of each protein from a single transcript.

Thermosensitive transcriptional regulation of the *caf1* operon

The ability to regulate gene expression in response to changes in temperature is important for *Y. pestis*, which lives mainly in two environments with different ambient temperatures: the flea at ~ 25 °C [8], and mammals, with body temperatures of around 37 °C. This difference in temperature is therefore a simple signal that allows the bacterium to determine whether it is in the vector or the host where it behaves as a mammal pathogen, expressing factors related to virulence and immunoevasion. The results of the temperature switching experiment clearly show the rapid upregulation of the expression of *caf1M, caf1A* and *caf1* upon an increase in temperature. The fold increase in transcript levels with temperature, that we observe, correlates well with those previously described [20] for *caf1R* and *caf1M* (8.8 vs 8.9 fold and 23.3 vs 19.7 fold respectively), but we detect greater increases in transcription for *caf1A* and *caf1* (33.4 vs 9.7 fold and 26.7 vs 7.8 fold respectively). Crucially, this upregulation occurs at the transcriptional level, and is wholly Caf1R dependent.

The difference in transcriptional output is sufficient to describe the observed temperature sensitivity of Caf1 protein production. Not only did the levels of Caf1M and Caf1 observed by western blot correlate with the levels of transcript seen by RT-PCR but bypass of transcriptional regulation through the introduction of a T7 promoter permitted expression of Caf1 at 30 °C, when there was no expression from the native *caf1* operon under these conditions. If a method of translational or post-translational regulation was present (e.g. a riboswitch), then it would be expected that these mechanisms would regulate T7 driven *caf1* expression at lower temperatures. Therefore, these mechanisms are either not present, or have a very weak effect in the Caf1 system. Caf1R mediated transcriptional control is thus the main mechanism to express Caf1 polymers under the appropriate conditions.

Caf1R mediates thermal regulation

It is clear from our results that Caf1R is the element that transduces the temperature change into a transcriptional response. There are three models through which this could take place (Fig. 5). In model A (Fig. 5a), the Caf1R protein itself is the thermoresponsive element. At 25 °C the protein is either unable to bind DNA, or unable to stimulate transcription of the *caf1* operon. Upon a shift in temperature to 35 °C, the protein changes conformation, becoming active and inducing transcription of the *caf1* operon. In model B (Fig. 5b), the thermoresponsive

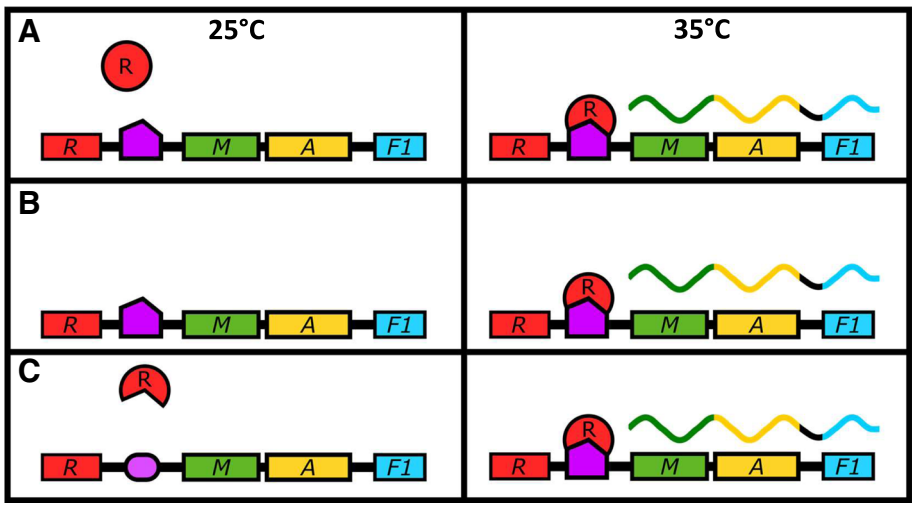

Fig. 5 Models of thermoresponsive Caf1R-dependent transcriptional control of gene expression. Three models that potentially describe the Caf1R dependent thermoresponsive increase in *caf1* operon transcription are shown, with their states at 25 °C and 35 °C depicted in the left and right hand side panels respectively. In model (**a**), the Caf1R protein is the thermoresponsive element. At 25 °C it is unable to activate transcription of the *caf1* operon, regardless of its DNA binding state. At 35 °C, a conformational change allows the protein to activate transcription. Inactive Caf1R is shown as a red circle and active Caf1R as a red arc. In (**b**), the Caf1R protein abundance is thermoresponsive. Transcription is not thermally regulated so either Caf1R is always transcribed but only translated at 35 °C or Caf1R is stabilised at 35 °C. The increased Caf1R protein level activates transcription of the operon. In (**c**), the DNA is the thermoresponsive element. At 25 °C, the DNA is not in an optimal conformation for Caf1R binding, and so transcription is not activated. At 35 °C, a change in the DNA facilitates optimal Caf1R binding, and so transcription of the operon is activated. Genes are shown as coloured rectangles (red, green, yellow and blue for *caf1R*, *caf1M*, *caf1A* and *caf1* respectively), with the Caf1R DNA binding site shown in purple. mRNA transcripts are represented as wavy lines coloured according to the gene they are transcribed from

element is the level of Caf1R, mediated either through temperature dependent increase in *caf1R* translation or decrease in Caf1R protein turnover. Finally, in model C (Fig. 5c), a temperature dependent structural change in a DNA regulatory region, e.g. supercoiling or promoter topology [1] is recognised by Caf1R, which binds and activates transcription.

Well studied examples found in *Yersinia* show how transcription factors can be thermally regulated in very different ways. In one case LcrF, which controls Type III secretion systems (TTSS) involved in phagocytosis evasion, undergoes translational control through the use of an RNA thermometer in which ribosome binding to the LcrF mRNA is inhibited by a temperature sensitive RNA secondary structure [2]. In contrast, the RovA transcription factor is itself a proteinaceous thermosensor. RovA functions as a "global regulator" of transcription in response to increases in temperature, resulting in the expression of the TTSS and Yop effector proteins in *Y. pestis*, and the internalization factor invasin in *Y. pseudotuberculosis* [31]. The mechanism through which RovA senses temperature involves a flexible loop situated between two alpha helices that have a role in dimer formation [7]. Upon temperature increase, the protein partially unfolds, lowering its affinity towards DNA and increasing its susceptibility to proteases, thus de-repressing the transcription of target genes.

The closest homologue of Caf1R is the Rob protein of *E. coli* [18], which consists of an N-terminal DNA binding domain coupled to a C-terminal GyrI-like domain [32]. Rob acts as a transcriptional activator that is induced by dipyridyl or decanoate [33]. In the uninduced state, the C-terminal domain causes sequestration of Rob into distinct cellular foci, preventing it from interacting with DNA and activating transcription. Addition of the inducer allows Rob to disperse through the cell and activate the transcription of its target genes [33] though direct interactions between its N-terminal domain and the σ^{70} subunit of RNA polymerase [34]. Truncated Rob proteins, where the GyrI domains were absent, were able to activate transcription in the absence of inducer [33, 35]. Additionally, MarA, which is homologous to the N-terminal DNA binding domain of Caf1R, contains no C-terminal domain, and so regulates transcription on the basis of its abundance rather than the presence of an inducer [33, 36]. A truncated Caf1R protein, containing only the first 82 amino acids of the DNA binding domain, was seen to allow expression of Caf1, whereas complete Caf1R knockout prevents Caf1 expression [22, 23]. The homology of Caf1R to these proteins therefore suggests that its N-terminal DNA binding domain is responsible for binding DNA and activating transcription, whilst the C-terminal GyrI-like domain is responsible for the characteristic Caf1R behaviour. Whether this is due to changes in Caf1R conformation (model A) or abundance (model B) we cannot resolve. Additionally, model C would also account for the results we describe here and Karlyshev et al. observed a series of repeated DNA sequences within intergenic region I which may interact with Caf1R and orient the DNA for optimal induction of gene expression [22]. It is also possible that a combination of these models is responsible. We made numerous attempts to purify sufficient amounts of Caf1R for such studies, but it appeared to be particularly unstable. Therefore, efforts to produce a soluble, functional construct will likely be necessary in order to answer these questions.

Thermosensitive transcriptional activators in Gram-negative bacterial pathogens

The function of Caf1R revealed in this study is of interest due to its role in the immunoevasive ability of a highly dangerous pathogen. However, it is possible that Caf1R also represents a class of transcription factors that has received little prior study. PSI-BLAST searches reveal that the most similar proteins to Caf1R are unannotated sequences from various Gram-negative bacteria such as *E. coli* and several *Salmonella* species. However, one similar protein sequence is LdaA from an enteropathogenic *E. coli* strain (EPEC). LdaA is a putative regulatory protein that exists as part of the *lda* locus, named locus of diffuse adherence for its apparent ability to attenuate the adherence of *E. coli* to HEp-2 cells [37]. The *lda* gene displays sequence similarity with other fimbrial proteins, particularly the K88 Fae proteins, but unlike these structured fimbriae, Lda fibres appear flexible when visualized by electron microscopy. These data, as well as the similarity of LdaE to known periplasmic chaperones [37], hint that the Lda proteins may assemble in a manner similar to Caf1, and form a similar structure. If this is the case, the sequence similarity of LdaA and Caf1R suggests it may potentially function in an analogous way. Additionally, homology between Caf1R and AfrR, a putative transcription factor in the AF/R1 operon, has been identified previously [38]. The AF/R1 pilus produced by the genes of this operon plays an important role in the adhesion and pathogenicity of a rabbit diarrheagenic *E. coli* (RDEC) strain. If these proteins indeed function in a similar way to Caf1R, it may be evidence of a conserved system of thermosensitive transcriptional activation present in Gram-negative pathogens.

Conclusion

Thermosensitive expression of the *caf1* operon is determined on the transcriptional level in a Caf1R

dependent manner. The *caf1* operon genes *caf1M*, *caf1A* and *caf1* are transcribed as a single polycistronic mRNA, from a promoter within the P2 region upstream of *caf1M*. The transcription of this mRNA is substantially upregulated upon a temperature shift from 25 °C to 35 °C. This is sufficient to allow protein expression and the subsequent formation of Caf1 polymers. As Caf1 has a key role in allowing *Y. pestis* bacteria to evade phagocytosis, these results reveal a simple but important mechanism through which a dangerous pathogen can sense and respond to its environment.

Methods

Plasmids and cell culture

Plasmid constructs (Additional file 1: Table S1) were derived from pGEM-T Caf1 and pBad33SD Caf1, described previously [13]. pCOP was constructed by substitution of the T7 promoter with a random sequence of the same length using a Q5 Site-directed mutagenesis kit (New England Biolabs). All other plasmids were constructed using an In-Fusion HD cloning kit (Clontech). Deletions were made by omitting the relevant region of the plasmid from the amplified region, then ligating the amplified region back together. Primers sequences are shown in Additional file 1: Table S2.

Transformations were performed using *E. coli* BL21 (DE3) (New England Biolabs). Three different colonies were selected from each plate and grown in LB media overnight at 37 °C, 180 rpm. Glycerol stocks were prepared for all transformed bacteria using 500 μl of bacterial culture and 500 μl of 60% (*v*/v) glycerol solution. Expression cultures were prepared using 5 ml TB media containing 100 μg ml^{-1} ampicillin and/or 20 μg ml^{-1} chloramphenicol and were inoculated using a stab from one of the glycerol stocks. To induce expression from pBad plasmids, L-arabinose was added at the concentrations described in the text at the same time as the cultures were inoculated. The flocculent layer heights were measured using a ruler after the centrifugation of capillary tubes containing Caf1 expression cultures at 2367 x g, 22 °C for 15 min.

For analysis of *caf1* operon transcript levels, bacterial cells were cultured in triplicate from glycerol stocks of *E. coli* BL21 (DE3) cells transformed with pCOPF (Additional file 1: Table S1) and grown at 25 °C overnight. The OD_{600} was measured for all cultures and then 2 ml were taken from each culture to be analysed by RT-PCR and western blot. The remaining cultures were diluted to an OD_{600} of 0.5 and then each culture was split into two cultures: one to be incubated at 25 °C and the other simultaneously at 35 °C. After one hour incubation the OD_{600} was measured and 2 ml samples taken from each culture for RT-PCR and western blot analysis.

RNA extraction and cDNA synthesis

RNA was isolated using an EZ-10 Total RNA Mini-Preps Kit (Bio Basic Inc.) by transferring 1 ml from each culture (~ 2×10^9 cells) separately to be centrifuged at 10,000 x g for 1 min. Supernatants were discarded and 100 μl of lysozyme solution (400 μg ml^{-1} lysozyme in RNase-free water) added to each sample pellet. The mixtures were resuspended thoroughly and RNA extracted according to the manufacturer's protocol. The concentration of RNA was measured using a Nanodrop UV spectrophotometer (Labtech). RNA samples were then used to perform cDNA synthesis using a Quanti-Tect Reverse Transcription Kit (Qiagen) according to the manufacturer's protocol. The resulting concentration of cDNA was measured using a Nanodrop UV spectrophotometer and the samples stored at -20 °C.

PCR and RT-PCR

The master reaction mix for RT-PCR experiments was prepared using GoTaq Flexi DNA Polymerase (Promega) by adding 4 μl nuclease-free water, 4 μl 5X Colorless GoTaq Flexi Buffer, 3 μl of 2 mM dNTPs, 3.2 μl of 25 mM $MgCl_2$, 0.2 μl SYBR Green (Sigma, S9430) (diluted 200 times with 100% DMSO), 0.2 μl GoTaq Flexi DNA Polymerase, 0.8 μl 10 μM primer mix and 5 μl of 50 ng μl^{-1} cDNA per reaction. Samples were loaded in a Rotor-Gene Q instrument (Qiagen) and critical threshold cycle values (CT values) were collected to measure the fold change in gene expression for each target gene, relative to β-lactamase transcription from the plasmid, with a threshold level of 0.5. Primer sequences are shown in Additional file 1: Table S2, with a schematic of the design shown in Additional file 1: Figure S1. Thermal cycling conditions are stated in Additional file 1: Table S3.

For transcript analysis, PCR was conducted using Phusion High-Fidelity DNA Polymerase kit from (New England Biolabs) by adding 10 μl of 5X Phusion GC buffer, 1 μl of 10 mM dNTPs, 2.5 μl each of 10 μM forward and reverse primers, 2 μl of 10 ng μl^{-1} template DNA, 1.5 μl of DMSO, 0.5 μl of Phusion DNA polymerase and nuclease-free water up to 50 μl. Thermal cycling conditions are stated in Additional file 1: Table S4.

Rapid amplification of cDNA ends (RACE)

5′ RACE experiments were performed as described previously [39] with minor modifications. Briefly, cDNA was prepared by performing the reverse transcriptase reaction using a 9 bp long gene specific primer (Additional file 1: Table S2) complementary to a 3′ region of the *caf1* gene. cDNA samples were diluted two times with nuclease-free water and purified using 10 kDa molecular weight cut-off Vivaspin centrifugal concentrators

(Sartorius) spun at 1000 x g for 30 min. This step was repeated to prepare the samples for poly-adenine tailing.

Poly-adenine tailing was performed by mixing 0.5 μl terminal transferase enzyme (20 Units / μl, New England Biolabs), 5 μl 10 x terminal transferase buffer, 5 μl of 2.5 mM $CoCl_2$, 0.5 μl of 10 mM dATP for tailing,, 5 pmol of cDNA to be tailed and nuclease-free water up to 50 μl. The mixture was incubated at 37° C for 30 min and then heated at 70° C for 10 min to inactivate the reaction.

Tailed cDNA was then amplified by PCR. The PCR reaction was prepared by mixing 0.5 μl of 2 U/μl Phusion DNA polymerase (New England Biolabs), 10 μl 5 x GC Phusion buffer, 1 μl of 10 mM dNTP mix, 1.2 μl of 10 μM (dT)17-adapter primer (Additional file 1: Table S2), 2.5 μl of 10 μM adapter primer (Additional file 1: Table S2), 2.5 μl of 10 μM gene specific primer (Additional file 1: Table S2) and 3% DMSO with 2 μl of tailed cDNA and nuclease-free water up to 50 μl. Cycling conditions are stated in Additional file 1: Table S5.

The PCR product was analysed by agarose gel electrophoresis. Amplification products were extracted from the gel using a Monarch gel extraction kit (New England Biolabs) according to the manufacturer's protocol. The concentration of pure DNA was measured using a Nanodrop UV spectrophotometer (Labtech) and sequenced by Eurofins Genomics.

Western blot

Bacterial cultures were centrifuged at 20,000 x g for 5 min. Supernatants and flocculent layers were discarded and cell pellets resuspended in 100 μl per OD_{600} of 100 mM DTT, 2% *w/v* SDS. The samples were heated at 95 ° C for 10 min and centrifuged at 20,000 x g for 10 min. 50 μl of this supernatant was mixed with 50 μl SDS loading buffer (2% w/v SDS, 0.1% w/v Bromophenol blue, 5 mM EDTA, 125 mM Tris pH 6.8, 15% *v*/v Glycerol and 1% v/v β-mercaptoethanol), boiled for 5 min and centrifuged at 20,000 x g for 5 min. The samples (10 μl) were resolved on 12% SDS-PAGE. Nitrocellulose membranes and blotting papers were soaked in CAPS buffer (10 mM CAPS pH 11) containing 20% methanol for 10 min. SDS-PAGE gels were soaked in the same buffer for 2 min. The blotting paper, nitrocellulose membrane and gel were assembled in a semi-dry blotter (Trans-Blot® SD semi-dry transfer cell, Biorad) and 18 V applied for 30 min. Blots were stained with Ponceau S solution (Sigma) for 10 min to view the efficiency of protein transfer and then rinsed with PBS buffer. The blots were blocked with TBS buffer (2.7 mM KCl, 38 mM Tris-HCl and 140 mM NaCl pH 8) containing 5% milk (from powder) at room temperature for 2 h and then rinsed with (1x) TBS buffer. The membranes were incubated for 4 h at room temperature or overnight at 4 °C with 2 ml 5% milk solution 1 μg ml^{-1} anti-FLAG antibody (Sigma) in order to detect the FLAG-tagged proteins and 6.9 μg ml^{-1} anti-Caf1 antibody (Abcam) for Caf1 protein detection. The blots were washed with (1x) TBS buffer two times for 5 min at room temperature then incubated for 4 h at room temperature or overnight at 4 °C with 2 ml TBS buffer containing 2.5–5 μg mL^{-1} anti-mouse antibody conjugated with horseradish peroxidase (Sigma). The blots were washed twice for 15 min with TBS buffer at room temperature, covered with developing solution and incubated at room temperature for a few minutes with shaking. Developing solution was prepared by dissolving 50 mg 4-chloronaphthol in 10 mL methanol and mixed with 50 ml developing buffer (20 mM Tris-HCl, 140 mM NaCl and 1 mM Na_2HPO_4 pH 7.2) containing 60 μl hydrogen peroxide. The blots were dried in air and the images were taken using a gel documentation system (Gel DocTM XR+, Biorad).

Bioinformatics

Promoter sequences were predicted using the BPROM (Softberry Inc., Mount Kisco, NY, USA, http://www.softberry.com) [24] and Neural Network Promoter Prediction 2.2 (NNPP2.2, 25] webservers. Sequence alignments were generated using the Clustal Omega [40] webserver and visualised using the ESPript [41] webserver (http://espript.ibcp.fr).

Statistical analysis

Statistical analyses were performed by one-way analysis of variance (ANOVA). Statistically significant differences between groups were then identified using a Holm-Šidák *post-hoc* test.

Additional file

Additional file 1: Table S1 List of constructs used in this study. **Table S2** List of primers used in this study. **Table S3** Thermal cycling conditions for RT-PCR. **Table S4** Thermal cycling conditions for PCR. **Table S5** Thermal cycling conditions for 5′ RACE. **Figure S1** Schematic of primer design. A diagram of the *caf1* operon present in the plasmids used in this study is shown, with arrows corresponding to the forward and reverse primers and placed in the approximate position where they bind. Forward primers are shown on top of the genes and reverse primers shown beneath. (A) Primers used for RT-PCR are shown in red for *caf1R*, green for *caf1M*, orange for *caf1A* and blue for *caf1*. (B) Primers used for detecting the presence of intergenic regions in cDNA are shown: purple for intergenic region I, olive for intergenic region II and dark blue for intergenic region III. **Figure S2** Schematic of the P1, P2 and P3 regions. The sequence of intergenic region I (INT1), located between *caf1R* and *caf1M* is shown, highlighted in green, with the P1, P2 and P3 regions aligned and highlighted in red, yellow and orange respectively. The ATG nucleotides corresponding to the start codon of *caf1M* are labelled with blue triangles. **Figure S3** Analysis of Caf1 content in the flocculent layer (A) Image of *E. coli* cultures containing the pT7-COP plasmid grown at 35 °C for the amounts of time stated, and centrifuged in capillary tubes to visualise the flocculent layer height. (B) SDS-PAGE analysis of the flocculent layers of the cultures from (A). (C) Graph of Caf1 band intensities in arbitrary units, obtained by densitometry of the gel shown in (B). **Figure S4** Diagram depicting 5′RACE experimental design. Part of the *caf1* operon is

shown, with *caf1A* in green, intergenic region III in orange and *caf1* in cyan. The location of the gene specific primer binding site at the 3′ end of *caf1* is highlighted. Using this primer, cDNA was synthesised and sequenced using the 5′RACE method. Predicted individual transcript sequences are depicted as red lines, where the length of the line represents the length of the sequence read. Only the promoter in the P2 region (responsible for transcription of the polycistronic mRNA) is present, and so the polymerase synthesises cDNA until it dissociates from the DNA. This means the majority of sequence reads continue through intergenic region III into *caf1A*, terminating at different positions. **Figure S5** Sequences obtained by 5′ RACE analysis of *caf1* transcripts. The partial coding sequence of *caf1A*, followed by intergenic region III and the coding sequence of *caf1* is shown (*caf1* operon, COP) aligned to sequence data obtained from 8 separate 5′ RACE reactions. Regions of similarity are bounded by blue boxes with red text, and regions of complete conservation highlighted in red with white text. The regions corresponding to *caf1A*, intergenic region III and *caf1* are underlined in green, orange and cyan respectively. (PDF 1619 kb)

Abbreviations

ANOVA: Analysis of variance; CU: Chaperone-usher; EPEC: Enteropathogenic *Escherichia coli*; PCR: Polymerase chain reaction; RACE: Rapid amplification of cDNA ends; RDEC: Rabbit diarrheagenic *Escherichia coli*; RT-PCR : Reverse transcription polymerase chain reaction; T3SS: Type 3 secretion system

Acknowledgements

Not applicable.

Funding

This project has been funded by the Industrial Biotechnology Catalyst (Innovate UK, BBSRC, EPSRC), award number BB/M018318/1 to support the translation, development and commercialisation of innovative Industrial Biotechnology processes. A.D.A was funded on a studentship from the Higher Committee for Education Development in Iraq. I. I is funded by CRUK program grant C1443-A22095. Equipment funded by Wellcome Trust grant 056232 was used in this study. The funding bodies had no role in the design of the study, the acquisition and analysis of data or the writing of the manuscript.

Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

ADA and DTP acquired the data; ADA, DTP and II analysed the data; ADA, DTP and JHL designed and conceptualized the study; DTP, HW, II, NDP and JHL were involved in supervision of the study; HW and NDP contributed materials towards the study; NDP and JHL acquired funding for the study; ADA, DTP and JHL wrote the manuscript, and II and HW assisted in the critical reading and revision of the manuscript. All authors have read and approved the final version of the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Received: 18 December 2018 Accepted: 25 March 2019
Published online: 29 March 2019

References

1. Shapiro RS, Cowen LE. Thermal Control of Microbial Development and Virulence: Molecular Mechanisms of Microbial Temperature Sensing. mBio. 2012;3(5):e00238–12.
2. Hoe NP, Goguen JD. Temperature sensing in Yersinia pestis: translation of the LcrF activator protein is thermally regulated. J Bacteriol. 1993;175(24):7901–9.
3. Chakraborty S, Li M, Chatterjee C, Sivaraman J, Leung KY, Mok Y-K. Temperature and mg (2+) sensing by a novel PhoP-PhoQ two-component system for regulation of virulence in Edwardsiella tarda. J Biol Chem. 2010;285(50):38876–88.
4. Albanesi D, Martín M, Trajtenberg F, Mansilla MC, Haouz A, Alzari PM, et al. Structural plasticity and catalysis regulation of a thermosensor histidine kinase. Proc Natl Acad Sci. 2009;106(38):16185.
5. Elsholz AKW, Michalik S, Zühlke D, Hecker M, Gerth U. CtsR, the gram-positive master regulator of protein quality control, feels the heat. EMBO J. 2010;29(21):3621.
6. Hurme R, Berndt KD, Normark SJ, Rhen M. A proteinaceous gene regulatory thermometer in Salmonella. Cell. 1997;90(1):55–64.
7. Quade N, Mendonca C, Herbst K, Heroven AK, Ritter C, Heinz DW, et al. Structural basis for intrinsic Thermosensing by the master virulence regulator RovA of Yersinia. J Biol Chem. 2012;287(43):35796–803.
8. Perry RD, Fetherston JD. Yersinia pestis--etiologic agent of plague. Clin Microbiol Rev. 1997;10(1):35.
9. Bertherate E. Plague around the world, 2010–2015. WHO Wkly Epidemiol Rec. 2016;91(8):89–93.
10. Du Y, Rosqvist R, Forsberg A. Role of fraction 1 antigen of Yersinia pestis in inhibition of phagocytosis. Infect Immun. 2002;70(3):1453–60.
11. Soliakov A, Harris JR, Watkinson A, Lakey JH. The structure of Yersinia pestis Caf1 polymer in free and adjuvant bound states. Vaccine. 2010;28(35):5746–54.
12. Sauer FG, Remaut H, Hultgren SJ, Waksman G. Fiber assembly by the chaperone-usher pathway. Biochim Biophys Acta. 2004;1694(1–3):259–67.
13. Roque AI, Soliakov A, Birch MA, Philips SR, Shah DSH, Lakey JH. Reversible Non-Stick Behaviour of a Bacterial Protein Polymer Provides a Tuneable Molecular Mimic for Cell and Tissue Engineering. Adv Mater (Deerfield Beach, Fla). 2014;26(17):2704–9.
14. Miller J, Williamson ED, Lakey JH, Pearce MJ, Jones SM, Titball RW. Macromolecular organisation of recombinant Yersinia pestis F1 antigen and the effect of structure on immunogenicity. FEMS Immunol Med Microbiol. 1998;21(3):213–21.
15. Zavialov AV, Kersley J, Korpela T, Zav'yalov VP, MacIntyre S, Knight SD. Donor strand complementation mechanism in the biogenesis of non-pilus systems. Mol Microbiol. 2002;45(4):983–95.
16. Ulusu Y, Dura G, Waller H, Benning MJ, Fulton DA, Lakey JH, et al. Thermal stability and rheological properties of the 'non-stick' Caf1 biomaterial. Biomed Mater. 2017;12(5):051001.
17. Lindler LE, Plano GV, Burland V, Mayhew GF, Blattner FR. Complete DNA sequence and detailed analysis of the Yersinia pestis KIM5 plasmid encoding murine toxin and capsular antigen. Infect Immun. 1998;66(12):5731.
18. Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos JL. Arac/XylS family of transcriptional regulators. Microbiol Mol Biol Rev. 1997;61(4):393–410.
19. Galyov EE, Smirnov O, Karlishev AV, Volkovoy KI, Denesyuk AI, Nazimov IV, et al. Nucleotide sequence of the Yersinia pestis gene encoding F1 antigen and the primary structure of the protein. Putative T and B cell epitopes. FEBS Lett. 1990;277(1–2):230–2.
20. Han Y, Zhou D, Pang X, Song Y, Zhang L, Bao J, et al. Microarray analysis of temperature-induced transcriptome of Yersinia pestis. Microbiol Immunol. 2004;48(11):791–805.
21. Motin VL, Georgescu AM, Fitch JP, Gu PP, Nelson DO, Mabery SL, et al. Temporal global changes in gene expression during temperature transition in Yersinia pestis. J Bacteriol. 2004;186(18):6298–305.
22. Karlyshev AV, Galyov EE, Abramov VM. Zav, Yalov VP. Caf1R gene and its role in the regulation of capsule formation of Y. Pestis. FEBS Lett. 1992;305(1):37–40.
23. Galyov EE, Karlishev AV, Chernovskaya TV, Dolgikh DA, Smirnov OY, Volkovoy KI, et al. Expression of the envelope antigen F1 of Yersinia pestis is

mediated by the product of caf1M gene having homology with the chaperone protein PapD of Escherichia coli. FEBS Lett. 1991;286(1):79–82.
24. Solovyev V, Salamov A. Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In: Li RW, Ed. Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies. Hauppauge, NY: Nova Science Publishers; 2011. p. 61–78.
25. Reese MG. Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. Comput Chem. 2001; 26(1):51–6.
26. Robison K, McGuire AM, Church GM. A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete Escherichia coli K-12 genome11Edited by R Ebright. J Mol Biol. 1998;284(2):241–54.
27. von Hippel PH, Revzin A, Gross CA, Wang AC. Non-specific DNA Binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects. Proc Natl Acad Sci U S A 1974;71(12):4808–12.
28. Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol. 1995;177(14):4121.
29. Chamberlin M, Ring J. Characterization of T7-specific ribonucleic acid polymerase: II. INHIBITORS OF THE ENZYME AND THEIR APPLICATION TO THE STUDY OF THE ENZYMATIC REACTION. J Biol Chem. 1973;248(6):2245–50.
30. Hui MP, Foley Pl Fau - Belasco JG, Belasco JG. Messenger RNA degradation in bacterial cells. (1545–2948 (Electronic)).
31. Cathelyn JS, Crosby SD, Lathem WW, Goldman WE, Miller VL. RovA, a global regulator of *Yersinia pestis*, specifically required for bubonic plague. Proc Natl Acad Sci. 2006;103(36):13514.
32. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Res. 2017;45(Database issue):D190–D9.
33. Griffith KL, Fitzpatrick MM, Keen EF, Wolf RE. Two functions of the C-terminal domain of Escherichia coli rob: mediating "sequestration-dispersal" as a novel off-on switch for regulating Rob's activity as a transcription activator and preventing degradation of rob by Lon protease. J Mol Biol. 2009;388(3):415–30.
34. Taliaferro LP, Keen Ef 3rd Fau - Sanchez-Alberola N, Sanchez-Alberola N Fau - Wolf RE, Jr., Wolf RE, Jr. Transcription activation by *Escherichia coli* Rob at class II promoters: protein-protein interactions between Rob's N-terminal domain and the σ (70) subunit of RNA polymerase. (1089–8638 (Electronic)).
35. Ariza RR, Li Z, Ringstad N, Demple B. Activation of multiple antibiotic resistance and binding of stress-inducible promoters by Escherichia coli rob protein. J Bacteriol. 1995;177(7):1655–61.
36. Barbosa TM, Levy SB. Differential Expression of over 60 Chromosomal Genes in *Escherichia coli* by Constitutive Expression of MarA. J Bacteriol. 2000;182(12):3467.
37. Scaletsky ICA, Michalski J, Torres AG, Dulguer MV, Kaper JB. Identification and characterization of the locus for diffuse adherence, which encodes a novel Afimbrial Adhesin found in atypical Enteropathogenic Escherichia coli. Infect Immun. 2005;73(8):4753–65.
38. Cantey JR, Blake RK, Williford JR, Moseley SL. Characterization of the Escherichia coliAF/R1 pilus operon: novel genes necessary for transcriptional regulation and for pilus-mediated adherence. Infect Immun. 1999;67(5): 2292–8.
39. Rapid amplification of 5′ complementary DNA ends (5′ RACE). Nature Methods. 2005;2:629.
40. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539 Available from: http://europepmc.org/abstract/MED/21988835.
41. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 2014;42(W1):W320–W4.