The Characterisation of the Interaction between PcrA and RNA Polymerase

Kate Harriott

BSc(Biotech)(Hons)

Doctor of Philosophy (PhD) Thesis

December 2011

Department of Biological Sciences
University of Newcastle, Australia

Statement of Originality

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or any other institution and, to the best of my knowledge and belief, contains no material previously published or written to any other person, except where due reference has been made in the text. I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying subject to the provisions of the Copyright Act 1968.

Kate Harriott

Acknowledgements

Many people deserve acknowledgement and a heartfelt thankyou for supporting me throughout my PhD.

Firstly, I would like to thank my supervisor Peter Lewis, for his guidance and support throughout my PhD. I am very grateful for your perseverance in making me see the positives, I will one day have a glass half full, I promise. I know that PcrA threw us for some loops over the years, but you made the problem solving a great experience.

To the members of the Lewis Lab past and present, and the newer members of the Grainge lab, thank you for making the lab such a great place to work. Thanks also for putting up with my tantrums when someone messed with my pipettes, bench or my awesome organisation of the lab. I really need to thank Xiao for all the help with the single particle analysis, without you I would be a results chapter down and probably still in the lab today.

To my family and friends, many of whom I have neglected while writing this thesis, thank you, thank you, thank you!! Particularly my Mum and Dad, who have always provided me with so much support. I love you all, through the good times and the bad. Finally to Ryan, who has dragged me kicking and screaming to write up this thesis with the promise of an awesome future, thank you. I have finally done it!

Contents

Abstract

RNA polymerase (RNAP) is the highly conserved multi-subunit enzyme that carries out transcription in all life forms. RNAP in bacterial species carries out all forms of RNA transcription (mRNA, rRNA and tRNA) and requires the interaction of additional factors to produce full length transcripts. The process of transcription is complex and new information is constantly emerging about the protein-protein interactions involved in the RNAP complex and the functional significance of these interactions.

This project aimed to characterise the interaction between the DNA helicase, PcrA and RNAP in the model Gram-positive organism *Bacillus subtilis*. This interaction was originally identified in a genomic library screen performed to identify novel transcription factors. The characterisation of the interaction provided information that allowed potential of roles for PcrA during transcription to be hypothesized. The PcrA-RNAP interaction was shown to be strong, resistant to high salt concentrations. The interaction was shown to be stable and could withstand salt concentrations which disrupted the interaction between RNAP and known transcription factor GreA. Furthermore, the interaction studies showed that there were multiple binding sites on both PcrA and RNAP. Two sites of interaction between PcrA and RNAP were identified in this study. Firstly, the PcrA CTD (amino acid 577-739) binds to the β subunit (amino acids 1-400) and secondly, the PcrA NTD (amino acids 85-310) interacts with the β' subunit (amino acids 1-102 and 228-310). The sites of interaction were investigated using a combination of techniques, including far-Western blotting, mutagenesis and single particle analysis. Extensive work was completed to identify the

regions of interaction in both PcrA and RNAP. This required the use of many different PcrA and RNAP protein fragments, which were created in this study. Additionally, mutagenic analysis was used to identify the specific amino acids involved in the interaction.

Excitingly, the characterisation of the PcrA-RNAP interaction in *B. subtilis* led to the investigation of this interaction in other bacteria. This work confirmed an interaction between PcrA and RNAP in closely related Gram-positive bacterium *Geobacillus stearothermophilus* and also between UvrD and RNAP in the Gram-negative bacterium *Escherichia coli*. The extension of this interaction into other bacterial species increases the significance of the interaction and indicates that the function of helicase binding to RNAP is conserved in both Gram-negative and Gram-positive bacteria.

The identified regions of interaction were used in conjunction with the PcrA-RNAP complex structure, to present a model of the interaction, which confirms PcrA binds to the upstream face of RNAP. The PcrA NTD-β' subunit interacting regions and the PcrA CTD-β subunit interacting regions were docked into the PcrA-RNAP complex structure. From the model of interaction it was hypothesised that PcrA was translocating upstream of RNAP in order to remove stalled RNAP complexes.

Chapter 1:

General Introduction

Chapter 1: General Introduction

1.1 RNA polymerase

1.1.1 RNAP structure

Prokaryotic RNAPs are multi-subunit enzymes, which are highly conserved. Bacterial core RNAPs comprise of 5 main subunits: in Gram-negative bacteria: $\alpha_2\beta\beta'\omega$ (Figure 1.1). In Gram-positive bacteria, such as *Bacillus subtilis* (*B. subtilis*), the core RNAP can include another subunit, δ (Helmann, 2003).
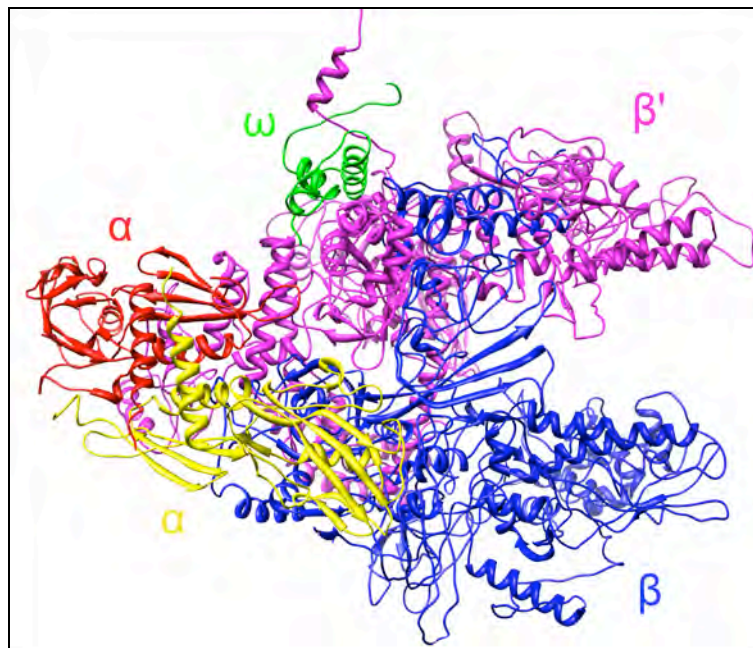
Figure 1.1: Side view of the homology model of core RNAP structure from *B. subtilis* (Johnston *et al*, 2009). The five subunits are: β in blue, β' in pink, the two α subunits in yellow and red, and ω in green. The cleft between the β and β' 'crab claw' is on the right while the α subunits anchor the complex on the left.

The X-ray crystallographic structure of RNAP core from the bacterium *Thermus aquaticus* has been solved (Zhang *et al*, 1999) and revealed that the largest subunits, β and β', assemble to form a positively charged clamp or 'crab claw' (Figure 1.2) which has an internal channel with a width of 25 Å, where the DNA enters to access the active

site (reviewed by Cramer, 2002; Murakami & Darst, 2003). The smaller α homodimer and ω subunits function to anchor the larger subunits, and α also has an additional role in promoter recognition (Borukhov & Nudler, 2003).
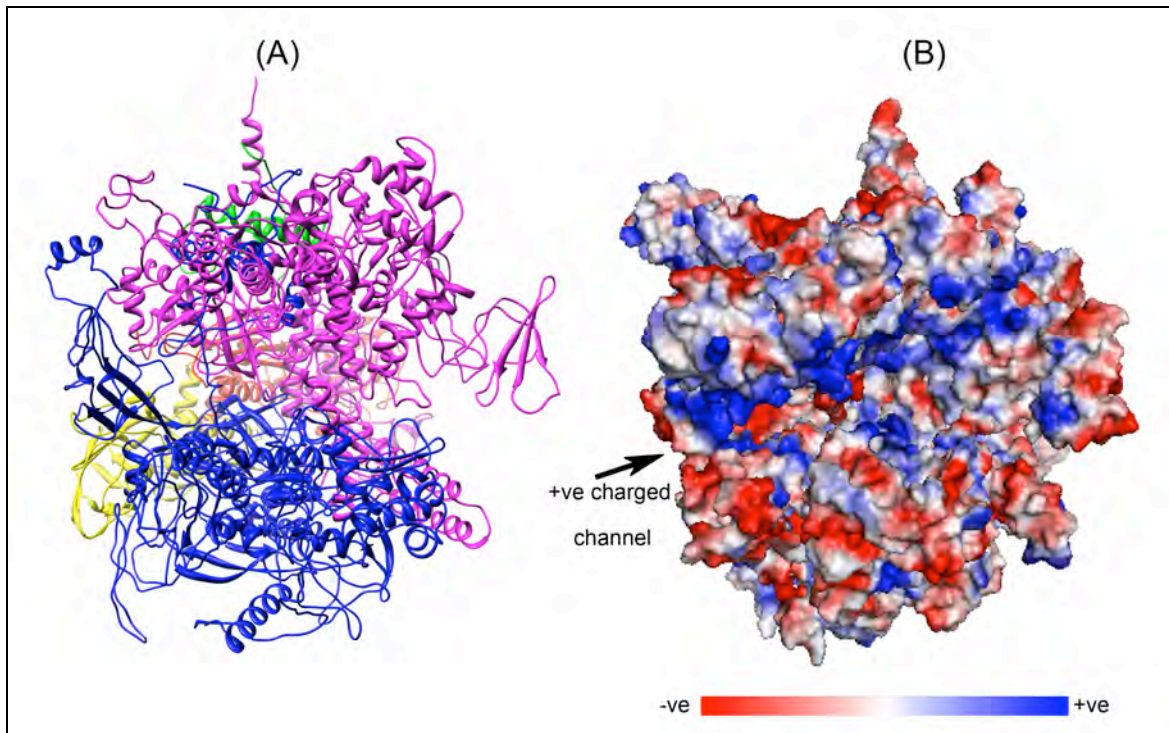
Figure 1.2: Face on view of the *B. subtilis* RNAP homology model that has been rotated 90° out of the page from Figure 1.1. (A) Homology model with the subunits coloured as in Figure 1.1. (B) Surface rendered model with vacuum electrostatics shown to indicate the surface charge. The positively charged cleft between β and β' subunits is shown in blue and indicated with the arrow.

The RNAP holoenzyme is formed after binding the sixth subunit σ, which is required for specific promoter recognition. The σ factor assists in loading the DNA into the entry channel by holding the β and β' 'crab claw' open (Cramer, 2002; Wigneshweraraj *et al*, 2005). Bacterial σ factors all share some degree of sequence similarity but the transcription of specific genes can require different σ factors (Murakami & Darst, 2003). The crystal structure of the holoenzyme from *Thermus thermophilus* can provide insights into the mechanism of transcription initiation (Vassylyev *et al*, 2002). The structure of the holoenzyme has shown the N-terminal region of σ forms a V-shape in

the upstream DNA binding channel of RNAP, while the σ C-terminus binds close to the RNA exit channel. The linker domain forms a hairpin structure that enters into the active site of RNAP (Vassylyev *et al*, 2002). The overall structure of the RNAP holoenzyme maintains the same core 'crab claw' structure and the interaction with the σ factor extends the β' pincer shape (Vassylyev *et al*, 2002).

RNAP consists of many different structural elements which are important during transcription and these structural elements are the contacts for interactions with DNA, RNA and transcription factors. Important elements within RNAP are highlighted on the crystal structure (Figure 1.3A) and in a simplified schematic (Figure 1.3B). These elements are often referred to when discussing RNAP structure.
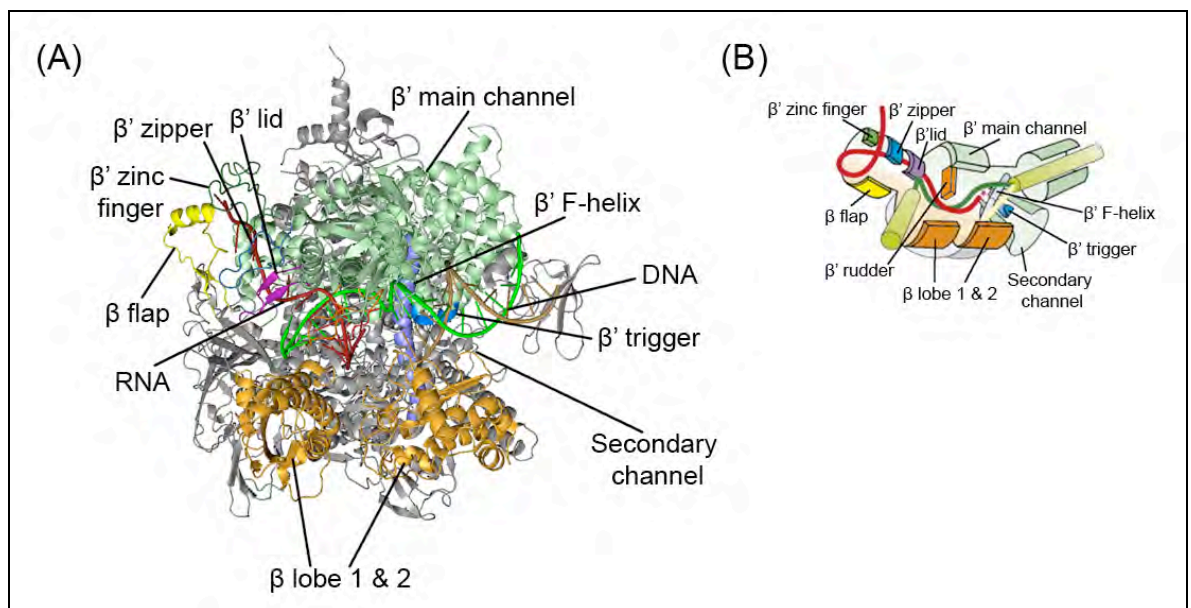
Figure 1.3: RNAP structural elements. (A) Crystal structure of the *Thermus thermophilus* elongation complex (adapted from Vassylyev *et al*, 2007) in the same orientation as Figure 1.2. (B) Schematic representation of the elongation complex (adapted from Borukhov & Nudler, 2008). β lobe – orange, β flap – yellow, β' zinc finger – dark green, β' zipper – blue, β' lid – purple, β' main channel – light green, β' F-helix – mauve, β' trigger – blue, secondary channel – grey, RNA – red, DNA green and brown.

1.1.2 The sigma (σ) factor

The σ sub-unit has four domains (Figure 1.4), separated by flexible linkers which bind across the upstream face of the RNAP crab claw (Murakami & Darst, 2003). The four domains are the N-terminal domain 1 (ND1), N-terminal domain 2 (ND2), the linker domain (LD) and the C-terminal domain (CD).
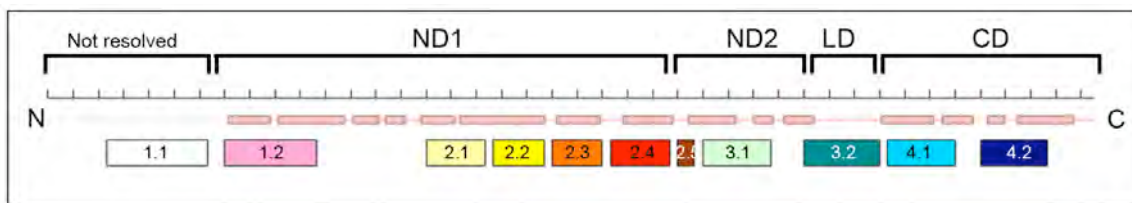
Figure 1.4: σ factor structural domains and conserved regions (Vassylyev *et al*, 2002). N-terminal domain 1 (ND1), N-terminal domain 2 (ND2), the linker domain (LD) and the C-terminal domain (CD)

The main interaction between σ and RNAP occurs *via* the σ ND1, which forms a U-shaped structure around the α-helical coiled coil region of β'. Other regions of σ also contact RNAP. ND2 bridges the ββ' pincers at the claw and the N-terminus of the linker domain contacts the β' subunit 'rudder' and 'lid' structures (Figure 1.3). The C-terminus of the linker domain becomes buried in the RNA exit channel and the CD has limited contacts with the β subunit flap tip helix and β' zinc finger (Figure 1.3) (Vassylyev *et al*, 2002).

Following the formation of the holoenzyme, σ regions 2 and 4 are spatially positioned to interact with the -10 and -35 promoter elements (Sevostyanova *et al*, 2007). Specifically, regions 2.4, 2.5 and 4.2 of the σ subunit are the regions that are required for the recognition of the promoter elements: -10, the extended -10 and -35 (Figure 1.5), respectively (Gross *et al*, 1998; Vassylyev *et al*, 2002).
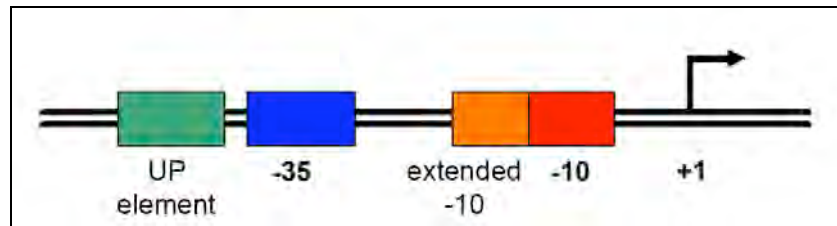
Figure 1.5: The typical upstream promoter region, with two conserved holoenzyme recognition sites at -10 (red) and -35 (blue) and the start of transcription at +1. The extended -10 box (orange) and UP element (green) are also shown.

1.1.3 Comparison of RNAP holoenzyme and elongation complex

More recently the crystal structure of the *Thermus thermophilus* elongation complex was resolved at 2.5 Å (Vassylyev *et al*, 2007). The crystal structure contains the core RNAP ($\alpha_2\beta\beta'\omega$) bound to a synthetic DNA, RNA/DNA hybrid and RNA transcript. This crystal structure can be compared with that of the holoenzyme, which was resolved at 2.6 Å (Vassylyev *et al*, 2002) to determine the structural differences in RNAP during initiation and elongation. The dissociation of the σ factor and the formation of the elongation complex produces some changes in the RNAP structure (Vassylyev *et al*, 2007). The holoenzyme structure already contains the important structural features, including the active site and the RNA exit channel, but due to the binding of σ they are reduced in size. During elongation and upon the release of σ the transcription bubble is increased in size. Finally, the RNA exit channel is constricted in the holoenzyme by the β flap (Figure 1.3) that was shifted by σ (Vassylyev *et al*, 2002) and with the release of σ the RNA exit channel is cleared of the β flap to allow the newly transcribed RNA to exit.

1.2 Prokaryotic transcription

Transcription is the synthesis of complementary ribonucleic acid (RNA) from template deoxyribonucleic acid (DNA) by RNA polymerase (RNAP) and is divided into three phases: initiation, elongation and termination (Figure 1.6). RNAPs in prokaryotes and eukaryotes are multiple subunit enzymes (Young *et al*, 2002), whilst mitochondrial and bacteriophage RNAPs are single subunit enzymes (reviewed by Bai *et al*, 2006). Prokaryote RNAPs have an evolutionarily conserved sequence and function throughout bacteria (Cramer, 2002; Young *et al*, 2002; Murakami & Darst, 2003).
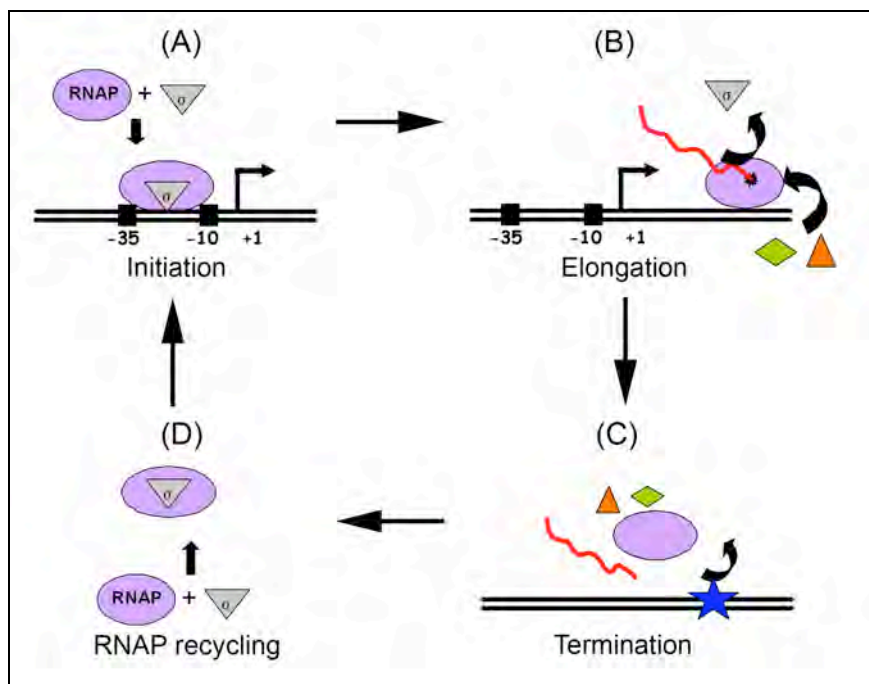
Figure 1.6: Transcription cycle. (A) Initiation. (B) Elongation. (C) Termination. (D) RNAP recycling. RNAP – purple oval, σ – grey triangle, RNA – red, elongation factors – green and orange.

1.2.1 Initiation

Core RNAP alone has transcriptional activity but requires the formation of the stable holoenzyme to correctly recognize the promoter region and initiate the transcription of genes.

The promoter region consists of conserved sequences upstream of the transcription start site (+1) at -10 and -35 base pairs (bp) (Figure 1.5, Hawley & McClure, 1983; Eick *et al*, 1994; Estrem *et al*, 1999). The RNAP holoenzyme contacts the promoter -35 and -10 regions through domains 2 and 4 of σ, respectively (Figure 1.4). During promoter recognition there are additional contacts between the RNAP α subunits C-terminal domains and the DNA UP element, upstream of the -35 promoter element (Caramori & Galizzi, 1998). Following RNAP holoenzyme binding to the promoter region, transcription of RNA begins from +1 (Eick *et al*, 1994).

Once the holoenzyme has identified the promoter region it initially binds to the DNA to form the closed promoter complex (Figure 1.7A). The closed promoter complex consists of the holoenzyme bound to the closed double stranded DNA (dsDNA). The RNAP holoenzyme can then undergo structural changes, the melting of the dsDNA promoter region and unwinding of downstream DNA (Figure 1.7B) to bind tightly to the DNA forming the open promoter complex (Figure 1.7C). The RNAP β, β' 'claw' in the closed complex is actually open to allow entry of the DNA double helix into the complex but the double stranded DNA is annealed (closed). Transcription requires a single stranded DNA (ssDNA) template to synthesise RNA and therefore the formation of the open promoter complex must occur before transcription begins. The open

promoter complex is formed by the closing of the RNAP β, β' claw and the melting (opening) of the dsDNA downstream of the start site, which completes the formation of the transcription bubble (reviewed by Young *et al*, 2002; Murakami & Darst, 2003). The template ssDNA enters the active site of the complex and the synthesis of the RNA chain begins (Figure 1.7C). NTPs required for RNA synthesis access the active site *via* entry through the secondary NTP channel (Figure 1.7C). During promoter melting and abortive initiation, the RNA exit channel is blocked by the σ 3.2 loop (Figure 1.4), but following the formation of the open promoter complex and the end of abortive initiation the σ 3.2 loop is removed from the RNA exit channel (Figure 1.7D). Several cycles of abortive initiation that release short RNA products of <11 nucleotides can occur before initiation is completed (Kapanidis *et al*, 2006). Promoter clearance occurs when the RNA transcript reaches ~13 nucleotides, core RNAP breaks away from the promoter region and releases σ (Figure 1.7E), to form the transcription elongation complex (EC, Figure 1.7F) (reviewed by von Hippel, 1998; Murakami & Darst, 2003).
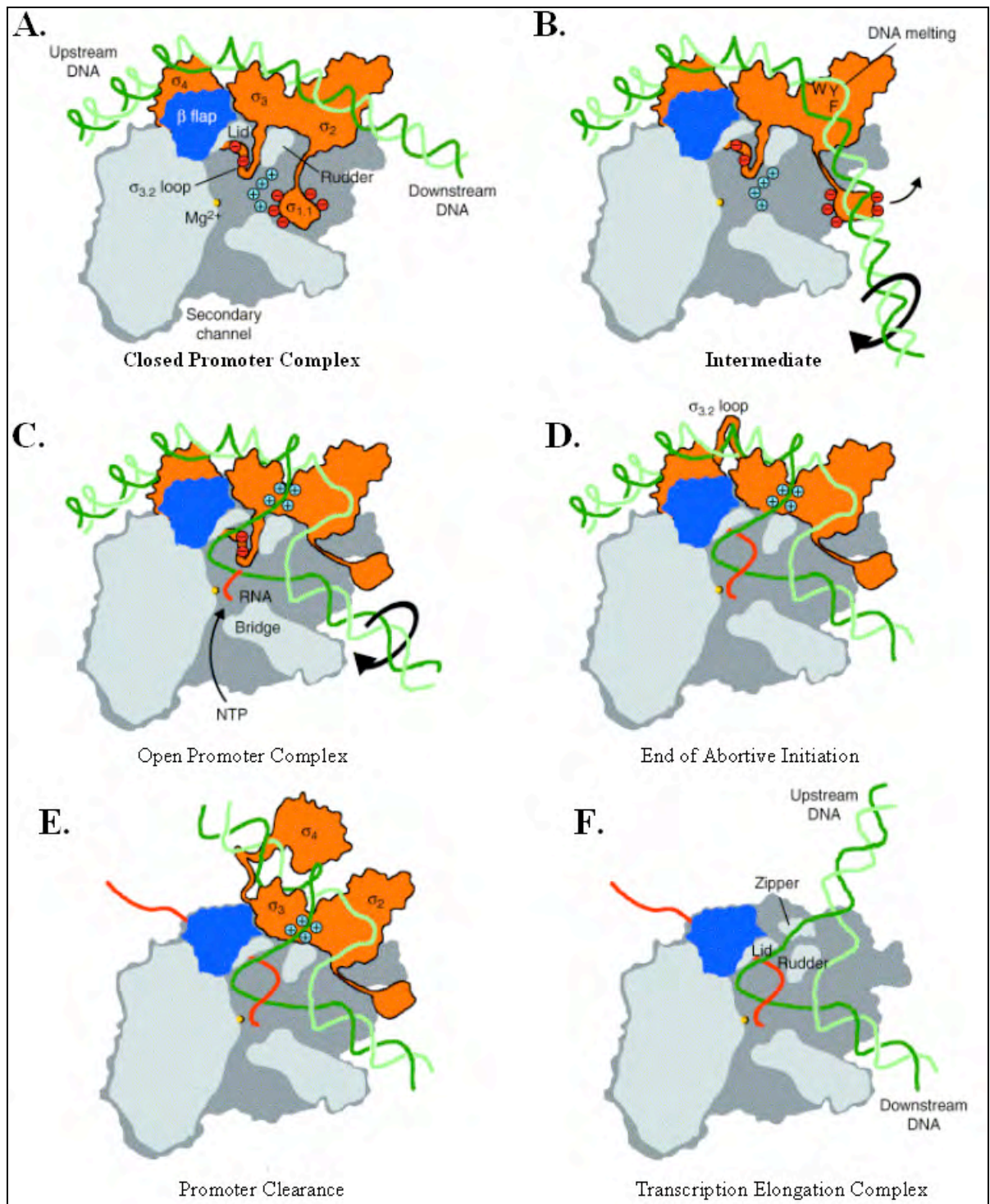
Figure 1.7: RNAP during transcription initiation. (A) Formation of the holoenzyme - σ binds to the core RNAP. Domains 2 and 4 of σ bind to the upstream promoter elements of dsDNA and form the closed promoter complex. (B) The formation of the open promoter complex begins with the melting of the dsDNA and unwinding of downstream DNA. (C) Open complex formation is complete when the melting of the dsDNA allows the ssDNA template strand to access the active site, NTPs enter the active site *via* the secondary NTP channel and transcription of RNA begins. (D) The σ 3.2 loop is moved from blocking the RNA exit channel when the RNA is transcribed, which is the end of abortive initiation. (E) When the transcript reaches ~13 nt promoter clearance occurs, breaking of the interaction between σ region 4 and β flap. (F) Promoter clearance and the dissociation of σ - formation of the elongation complex. β flap – blue, σ – orange, DNA – green and RNA – red. The view is looking down on top of the β subunit, but with most of β removed, revealing the inside of the RNAP active site channel (Murakami & Darst, 2003).

1.2.2 Elongation

The elongation phase of transcription begins with promoter clearance and the formation of the transcription elongation complex (EC). Promoter clearance requires the breaking of the promoter interactions and the dissociation of the σ sub-unit, after the RNA transcript reaches ~13 nucleotides long (Eick *et al*, 1994; Ebright, 2000; reviewed by Murakami & Darst, 2003), as previously mentioned. After the formation of the EC RNAP translocates along the DNA melting the DNA duplex and transcribing mRNA at ~35-60 nucleotides per second (Nudler, 1999). Transcription elongation does not always occur at a uniform rate and is determined by DNA sequence dependent pausing, DNA lesions, and interference by protein roadblocks (Smith & Savery, 2005; Bai *et al*, 2006). To overcome these types of events RNAP interacts with proteins, known as transcription elongation factors, which assist in transcription elongation and termination (Section 1.3).

1.2.3 Termination

1.2.3.1 Intrinsic termination

Termination of transcription is the release of the RNA and DNA from the elongation complex and can occur *via* intrinsic or extrinsic termination. Intrinsic termination is caused by a G-C rich inverted repeat DNA sequence, followed by a run of A residues, which results in a stable hairpin formation in the RNA, upstream from the site of termination (Epshtein *et al*, 2007). The precise mechanism with which the RNA hairpin interacts with the EC to cause termination is currently unresolved (Epshtein *et al*, 2007). To remove RNA from the terminated EC it was thought that forward

translocation of the complex was required (Figure 1.8A). The hairpin prevented

backtracking of the EC and therefore removed the transcript from the active site,

allowing it to leave *via* the exit channel (Brendel *et al*, 1986; Landick, 1999). However,

more recent studies into intrinsic termination have found another mechanism is

involved, in which the RNA hairpin causes conformational changes in the EC to cause

termination (Epshtein *et al*, 2007). Instead of forward translocation of the EC to remove

the RNA transcript, the allosteric model (Figure 1.8B) proposes that the hairpin enters

the exit channel to destablise the EC and release the RNA transcript (Gusarov & Nudler,

1999; Epshtein *et al*, 2007).
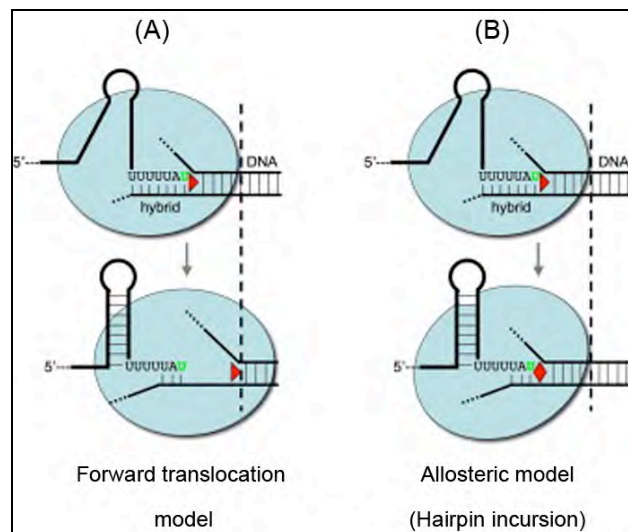
Figure 1.8: Alternative models of intrinsic termination. (A) Forward translocation model. The trapped RNAP moves forward along the DNA, releasing the RNA transcript. (B) Allosteric model. The RNAP doesn't move, instead the RNA hairpin enters the RNA exit channel, destabilizing the active centre and inactivating the EC (Epshtein *et al*, 2007).

Using the allosteric model of intrinsic termination the structural significance for RNAP

can be explained. The EC is paused after the transcription of the RNA hairpin and poly

U sequence, which allows the formation of the RNA hairpin (Figure 1.9A). The RNA

hairpin is nucleated *via* interactions with the exit channel, including the zinc finger and

the β flap domain (Figure 1.9B). The exit channel widens to allow for hairpin

incursion. As the hairpin stem increases to 4 to 5 bp it enters the primary channel of the EC. This begins the collapse of the transcription bubble and puts pressure on the β' rudder and the β lobe (Figure 1.9C). The primary channel widens, allowing the hairpin to move towards the secondary channel and when the RNA hairpin is fully formed, the head of the hairpin reaches the downstream site of the transcription bubble and the β' trigger loop (Figure 1.9D). Steric clashes between the trigger loop and the hairpin causes irreversible inactivation of the EC and the trigger loop movement also induces clamp opening, which releases the RNA and DNA from RNAP.
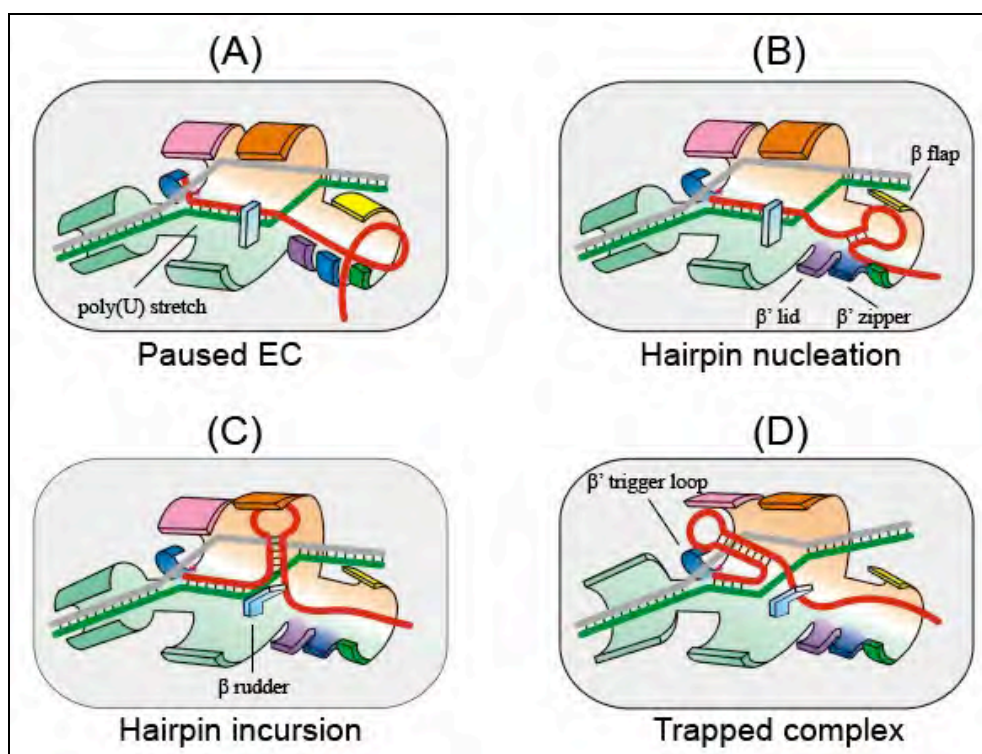
Figure 1.9: Allosteric model of intrinsic termination: (A) Paused EC, after the transcription of the RNA hairpin sequence. (B) Hairpin nucleation by the EC exit channel, allows entry of the newly forming hairpin into the EC. (C) Hairpin incursion, the hairpin enters the primary channel collapsing the transcription bubble. (D) Trapped complex, the fully formed RNA hairpin reaches the secondary channel and inactivates the EC and clashes with the trigger loop induce the release of the RNA and DNA (adapted from Epshtein *et al*, 2007)

1.2.3.2 Rho-dependant termination

An additional protein, Rho, is required for extrinsic termination. The Rho factor is a homohexameric protein complex that forms a ring shape, through which the RNA is pulled during termination. Rho monomer is 46.8 kDa, making the hexamer a large structure of 280.8 kDa. Rho is present in both Gram-positive and negative bacteria and has shown to be essential in Gram-negative bacteria. Rho is an RNA helicase with ATPase and translocase activity in the 5'-3' direction (Richardson, 2002; Epshtein *et al*, 2010).

The classical extrinsic termination model required Rho to bind to the RNA first, before contacting the elongating RNAP to cause termination (Banerjee *et al*, 2006). It was thought that Rho located a specific RNA sequence on the transcript, known as the rho-utilization (*rut*) site, and travelled along the nascent RNA towards the RNAP elongation complex (EC) (Richardson, 1996). It was thought that binding to the *rut* site activated the 5'-3' ATP dependent RNA-DNA helicase activity of Rho, allowing Rho to travel along the RNA towards the EC (reviewed by Ciampi, 2006). Once Rho reached the EC it disrupted the interaction of the EC and RNA, causing the release of the RNA transcript.

More recent data have disproved this model of Rho-dependent termination. It has been shown that Rho binds to RNAP in the absence of RNA, DNA or other proteins. Instead, Rho binds the EC early during transcription, rather than travelling along the nascent RNA, to contact RNAP (Mooney *et al*, 2009; Peters *et al*, 2009; Davis, 2011). Additionally it was shown that termination efficiency was increased if Rho was bound to an early EC and remained associated throughout elongation (Epshtein *et al*, 2010).

This updated model proposed that Rho binds the EC early in the transcription cycle (Figure 1.10A) and RNA is dragged through the middle of the hexameric complex by Rho, which causes an RNA loop to form and tighten as elongation continues (Figure 1.10B). The termination event is triggered by the EC inactivation, whereby the EC becomes trapped at termination sites and this inactivation is caused by Rho-induced conformational changes to the EC (Figure 1.10C). It is the site of the inactivation of the EC that determines the position of termination, as the dismantling of the EC that occurs after is a slower process (Figure 1.10D). The EC conformational changes that cause the inactivation occur in a region of the RNAP β' subunit known as the trigger-loop (TL) domain (Epshtein *et al*, 2010). The TL domain (Figure 1.9D and Figure 1.3, β' trigger) has been shown to have the ability to change conformation during transcription and to have a significant role throughout elongation and intrinsic termination (Epshtein *et al*, 2007) (Vassylyev *et al*, 2007).
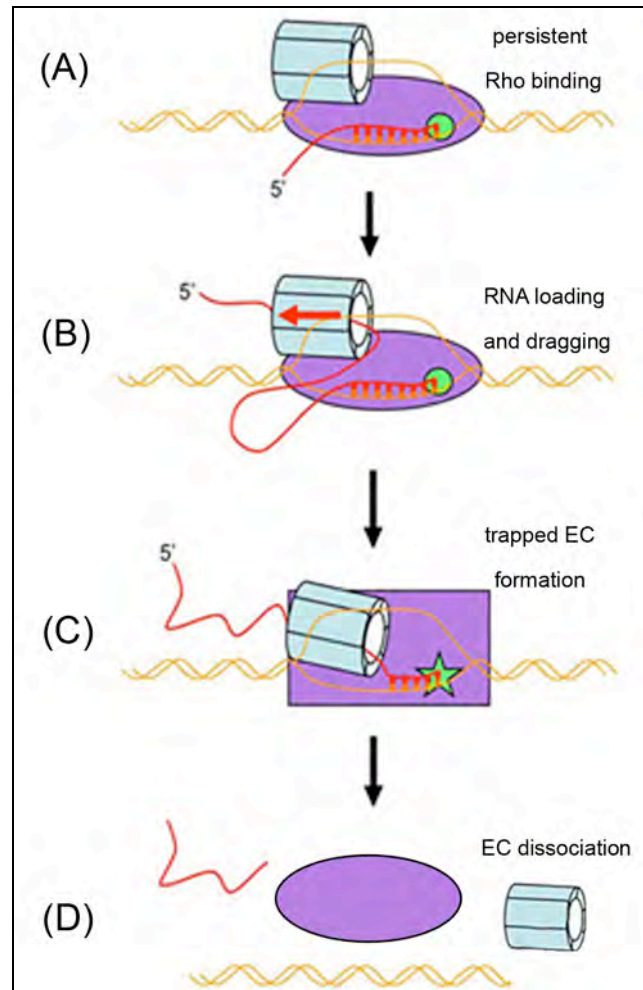
Figure 1.10: The molecular pathway to Rho-dependent termination. (A) Rho binds RNAP early in the transcription cycle. (B) Rho drags the RNA transcript through the hexamer's centre cavity, forming an RNA loop. (C) EC is inactivated (trapped). (D) EC dissociation, which releases the RNA transcript, RNAP, Rho and DNA. DNA – orange, RNA – red, Rho – blue cylinder, EC – purple oval (Epshtein *et al*, 2010).

It has also been shown that Rho dependent termination is not only required in mRNA transcription but can also be used in the resolution of extended RNA:DNA hybrids and may protect cells from the transcription of harmful foreign genes (Cardinale *et al*, 2008; Peters *et al*, 2009)

1.3 Transcription Factors

1.3.1 Gre factors

In *Escherichia coli* (*E. coli*) and other Gram-negative organisms GreA and GreB are transcription factors that have roles in proofreading, prevention of RNAP pausing and they can assist in the transition from initiation to elongation (Laptenko *et al*, 2003). Gre factors can reactivate backtracked or stalled RNAPs by cleaving the 3' end of the RNA transcript, which creates a new 3' terminus in the active site of the elongation complex (Borukhov *et al*, 1993; Komissarova & Kashlev, 1997). This cleavage of nascent RNA strands to resolve stalled complexes is conserved in all multi-subunit RNAPs (Fish & Kane, 2002). Although *E. coli* has two Gre factors (GreA and GreB) *B. subtilis* and other Gram-positive organisms have only a single GreA that performs functions of both *E. coli* factors (Kusuya *et al*, 2011).

1.3.2 Nus factors

The Nus transcription factors interact with RNAP and the RNA transcript (DeVito & Das, 1994) to modulate the rate of transcription and to increase or decrease transcription termination. NusA has roles in transcription pausing and can promote hairpin formation, which causes intrinsic termination (Richardson, 1996; Gusarov & Nudler, 2001; Davies *et al*, 2005). It was proposed that NusA could facilitate the formation of the hairpin by disrupting the interaction between the RNA transcript and RNAP at the upstream binding site (Gusarov & Nudler, 2001).

The role of NusA during antitermination has been characterised in the transcription of both ribosomal operons (*rrn*) and the bacteriophage λ (Richardson, 1996; Prasch *et al*, 2009). When NusA is bound to RNAP in complex with the other Nus factors (NusB, NusE and NusG) the complex allows transcription to continue through both intrinsic and extrinsic terminators (DeVito & Das, 1994; Borukhov *et al*, 2005). Phage λ anti-termination is mediated by an RNA recognition sequence (consisting of *boxA* and *boxB*) and at this sequence the Nus factors NusA, NusB, NusE and NusG can form the antiterminator complex (Figure 1.11). In the antitermination complex NusB and E bind to the *boxA* sequence, while NusA binds to the spacer RNA adjacent to *boxA*, RNAP α subunit and the λ N protein, while the λ N protein binds to *boxB* (Prasch *et al*, 2009). The λ N protein and *boxB* are not required in *rrn* antitermination (Morgan, 1986). The AR domains of NusA, shown in Figure 1.11 are only present in gamma proteobacteria and in most bacterial species NusA interacts with RNAP *via* the N-terminal domain (Ha *et al*, 2010).
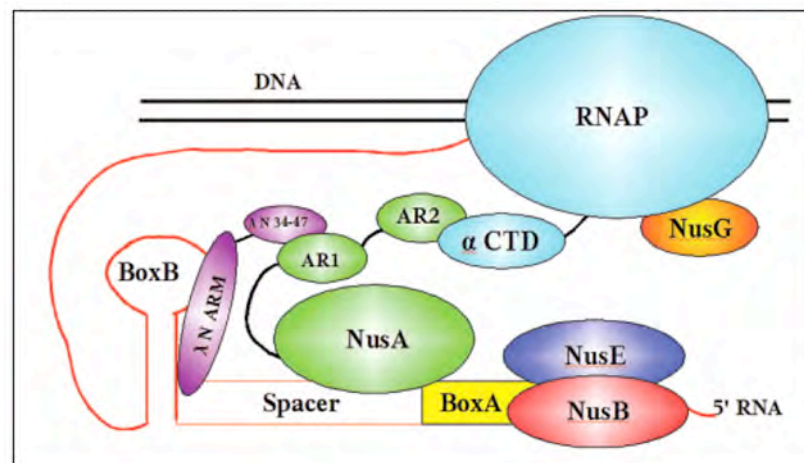
Figure 1.11: Model of phage λ antitermination. Nus factors (NusA, NusB, NusE and NusG) form the antiterminator complex during transcription. NusA (green) binds to RNAP (light blue), NusE (blue) and NusB bind to *boxA* (yellow) and λN (purple), which also binds *boxB*. NusG (yellow) binds RNAP in the complex (Prasch *et al*, 2009).

It has also been shown that the ratio of NusA:RNAP during transcription of mRNA and rRNA differs and could be a reason why NusA can act as both a terminator or anti-

terminator (Horwitz *et al*, 1987; Davies *et al*, 2005). Anti-terminator complexes in rRNA synthesis contain two NusA molecules, while only one NusA is involved in mRNA transcription (Davies *et al*, 2005).

1.3.3 Mfd

Mfd is the transcription-repair coupling factor identified in *E. coli* and *B. subtilis* (Selby & Sancar, 1993a; Selby & Sancar, 1993b; Ayora *et al*, 1996) and is conserved though many bacterial species. The roles of Mfd during transcription include removing stalled complexes from DNA lesions and recruiting nucleotide excision repair proteins to that site, the reactivation of backtracked complexes and the displacement of complexes stalled at protein roadblock sites (Selby & Sancar, 1993a; Selby & Sancar, 1993b; Ayora *et al*, 1996; Smith & Savery, 2005). Mfd binds to DNA upstream of the transcription bubble during elongation, when RNAP becomes stalled. Mfd cannot release holoenzyme complexes during initiation, which supports the findings that promoter complexes are resistant to transcription coupled repair (Park *et al*, 2002).

Mfd contains a conserved SF2 helicase motor, which is coupled to an RNAP interacting domain (Figure 1.12), but does not have helicase activity (Deaconescu *et al*, 2006; Singleton *et al*, 2007). This region, containing the helicase motifs has homology to the translocase motor portion of the RecG DNA repair protein (Roberts & Park, 2004). The N-terminal region of Mfd has a region of homology to UvrB, which recruits the excision repair proteins after removal of the stalled elongation complex (Roberts & Park, 2004).
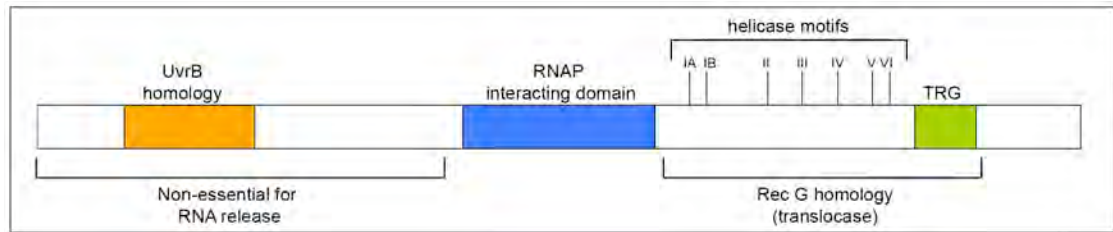
Figure 1.12: *E. coli* Mfd protein. The UvrB homology region (orange) is responsible for recruiting excision repair complex proteins, but is not required for the release of RNAP. The RNAP interaction domain binds to RNAP and the helicase motifs and the TRG (translocase in RecG) induce the translocation and release of RNAP (Roberts & Park, 2004).

1.3.4 DksA

DksA is a nutrient-responsive transcription factor identified in *E. coli* that has roles in the regulation of rRNA promoters (Paul *et al*, 2004). There is no known DksA homologue in *B. subtilis* (Krasny & Gourse, 2004; Paul *et al*, 2004). In *E. coli,* DksA binds to the secondary channel of RNAP and stabilizes the interaction with (p)ppGpp (guanosine (penta) tetraphosphate) (Paul *et al*, 2004). ppGpp is a molecule made as a response to nutritional stress and amino acid starvation, which binds to RNAP and increases the rate of open complex collapse at the promoter, preventing rRNA transcription (Barker & Gourse, 2001). The inhibition of rRNA transcription is important to ensure DNA replication completion occurs in times of nutritional stress or amino acid starvation. In the absence of DksA replication can be blocked by stalled transcription complexes at these times (Tehranchi *et al*, 2010).

1.3.5 RfaH

RfaH is an elongation factor in Gram-negative bacteria that can increase the expression of distal genes in large operons, by allowing RNAP to bypass roadblocks as it travels through the operon (Figure 1.13; Sevostyanova *et al*, 2008). RfaH binds to the β'

subunit clamp helices (CH), which is also the binding site of the σ factor. The binding of RfaH to RNAP occurs in two steps: firstly the N-terminal domain of RfaH binds to the *ops* DNA sequence (GGCCCTAGnnTG) on the nontemplate DNA strand (Artsimovitch & Landick, 2002). This causes the displacement of the RfaH C-terminal domain and the exposure of the RNAP binding site on the RfaH N-terminal domain and this site then interacts with the β' CH region (Sevostyanova *et al*, 2008).
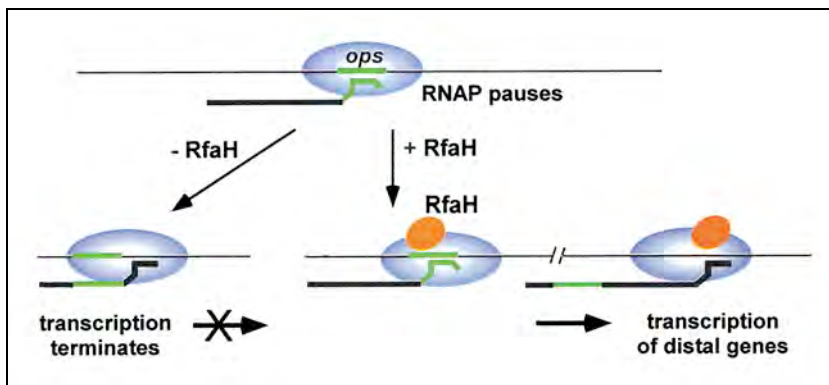
Figure 1.13: Recruitment of RfaH by RNAP (blue) paused at the *ops* DNA site (green). Without RfaH (-RfaH) transcription is terminated. With RfaH (+RfaH) (orange) bound, the EC is modified to a termination-resistant state and transcription of distal genes occurs (Artsimovitch & Landick, 2002).

1.3.6 δ subunit

The δ subunit is a non-essential subunit of RNAP in Gram-positive bacteria (Helmann, 2003), which co-purifies with RNAP and is abundant during cell growth (Lopez de Saro *et al*, 1999) . The tertiary structure of δ has an N-terminal region consisting of both α helices and β sheets, while the C-terminal region is unstructured in solution (Lopez de Saro *et al*, 1995). The highly acidic C-terminal region has been shown to displace RNA bound to RNAP (Lopez de Saro *et al*, 1995). Investigation of the function of δ has shown that δ can have differing effects on RNAP. The C-terminal region may be mimicking ssRNA and indicates δ may have a role in transcription termination or RNAP recycling (Lopez de Saro *et al*, 1995), which can stimulate overall transcription

rates (Achberger *et al*, 1982; Juang & Helmann, 1994). Conversely, δ can inhibit the open promoter complex formation by reducing the contacts between RNAP and the promoter DNA (Achberger *et al*, 1982), which reduces the transcription rate. The inhibitory or stimulatory effects of δ during transcription *in vitro* have shown to be influenced by both reaction conditions and importantly the DNA template (Juang & Helmann, 1994)

The factors outlined here are not the only factors required to ensure transcription occurs within bacteria. There are many additional factors that are specific to the transcription of certain genes. Additionally, there could be novel transcription factors, which have yet to be identified. Recent proteomic studies have shown that RNAP interacts with other protein in the cell, of particular interest was interactions with DNA helicases PcrA and HelD (Delumeau *et al*, 2011).

1.4 DNA Helicases

DNA helicases are motor proteins that separate duplex DNA through the energy release from ATP hydrolysis. Helicases have important roles in replication, transcription, recombination and repair of DNA, by unwinding double stranded DNA (Matson & Kaiser-Rogers, 1990; Marians, 1997). The biochemical processes which are shared by helicases include duplex unwinding, nucleic acid binding and ATP hydrolysis (Hall & Matson, 1999). The energy used by helicases to unwind duplex DNA is provided by nucleotide hydrolysis (Bird *et al*, 1998). Most helicases have the ability to catalyse the unwinding of DNA in either the 3' to 5' or 5' to 3' direction, but not both (Figure 1.14).
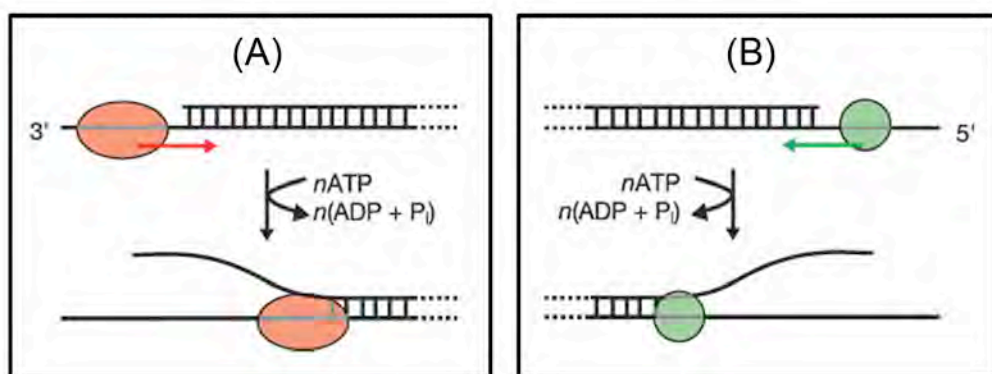
Figure 1.14: Helicases can translocate along DNA in different directions. (A) Helicase (red oval) binds to single strand of DNA and unwinds in a 3' to 5' direction. (B) Helicase (green circle) unwinds the DNA in a 5' to 3' direction (adapted from van Brabant *et al*, 2000). Bidirectional helicases can unwind the DNA in both directions.

Defining characteristics of helicases include the rate, directionalilty, processivity, step size and 'active' versus 'passive' mechanism of action (Singleton *et al*, 2007; Abdelhaleem, 2010). Helicases can be identified by a series of highly conserved motifs, which when structurally mapped, form a nucleotide-binding pocket and nucleic acid binding site (Hall & Matson, 1999). There are five separate families of helicases (Table 1.1) with each group being identified by a different set of conserved motifs (Hall & Matson, 1999; Abdelhaleem, 2010). There is an additional group of proteins, known as SWI/SNF proteins which also contain conserved helicase motifs (Section 1.4.1.1).

Table 1.1: Conserved helicase motifs present in the five different helicase families and closely related SWI/SNF proteins. Examples of helicases within each superfamily are also provided.

Helicase superfamily (SF)	Number of conserved motifs	Examples
SF1	7	PcrA, UvrD
SF2	7	UvrB, UL9
SF3	3	SV40 LTag
SF4	4	DnaB, Twinkle
SF5	4	Rho
SWI/SNF proteins	6	RapA/HARP

Superfamilies 1 (SF1) and 2 are closely related and contain helicases from prokaryotic, eukaryotic and viral organisms, all of which have seven conserved motifs (SF1; Figure 1.15), while members of SF3 have only three motifs and SF4 has four (Hall & Matson, 1999). SF1 and SF2 helicases usually act as monomers or dimers while SF3-5 helicases commonly form hexameric complexes (Abdelhaleem, 2010).
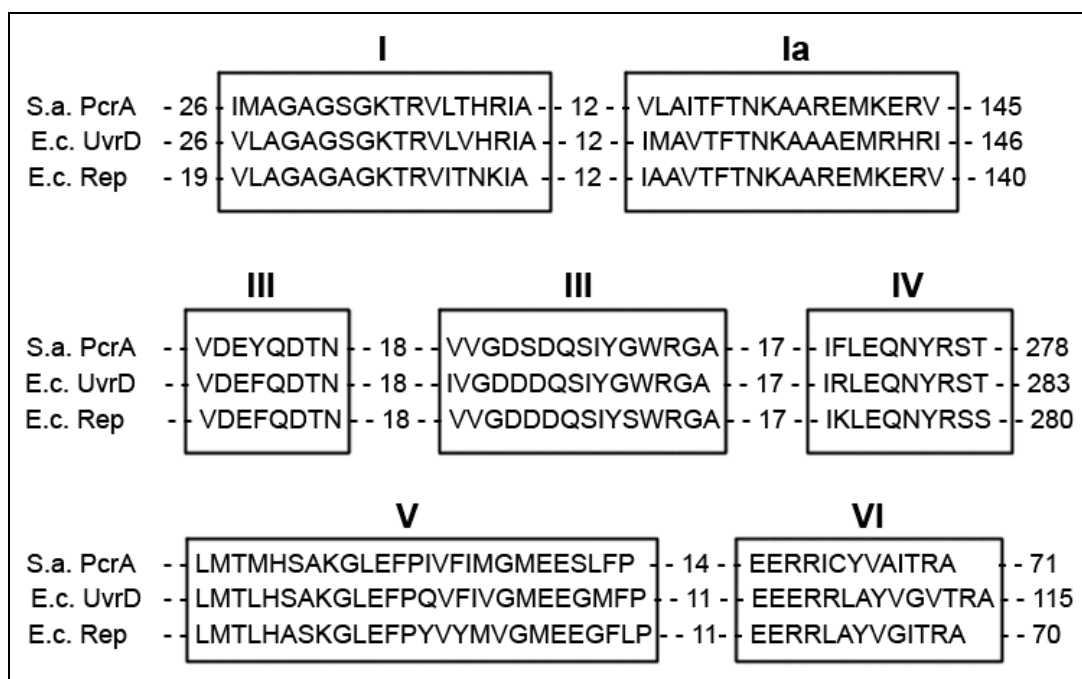
Figure 1.15: The seven conserved motifs of the DNA helicase superfamily I, shown here as an alignment of three different SF1 helicases: *S. aureus* PcrA, *E. coli* UvrD and *E. coli* Rep. The sequences within the boxes are conserved, the numbers between pairs of motifs is the number of amino acids that separate the motifs (Petit *et al*, 1998).

UvrB is a SF2 helicase that forms part of the UvrABC endonuclease, which is the enzyme responsible for nucleotide excision repair (NER) in bacteria (Truglio *et al*, 2006). SF2 helicase UL9 is a herpes simplex virus type-1 protein required for replication *in vivo* and has roles in the binding and unwinding of the origin of replication (Marintcheva & Weller, 2003). SF3 helicases were originally identified in genomes of small DNA and RNA viruses (Hickman & Dyda, 2005). SV40 LTag is a hexameric DNA helicase utilized in initiation of virus replication (Jankowsky & Fairman-Williams, 2010). *E. coli* DnaB is a SF4 helicase that is a part of the replisome

complex, which carries out DNA replication (Pomerantz & O'Donnell, 2007). SF4 helicases also include a mitochrondrial DNA helicase (Twinkle) essential for DNA maintenance (Matsushima *et al*, 2008). SF5 consists of the termination factor Rho (Section 1.2.3.2). Rho is an RNA helicase that is closely related to SF4 helicases, but was placed in its own separate family on sequence based differences (Abdelhaleem, 2010).

1.4.1 Helicases in Transcription

Protein-protein interactions between RNAP and DNA helicases have been identified in many studies. Investigation of the protein interaction network in *Escherichia coli* identified an interaction between RNAP and helicase RapA (Butland *et al*, 2005). Protein complex isolation experiments in *B. subtilis* have identified two helicases that interact with RNAP – PcrA and helicase IV (*helD*) (Delumeau *et al*, 2011). As previously mentioned (Section 1.1), bacterial RNAPs have an intrinsic helicase ability to melt the DNA during initiation and therefore do not require a DNA helicase to separate the dsDNA. Currently there is little known about the functional significance of helicases binding to RNAP during prokaryotic transcription. If the interaction is not required for promoter DNA melting, the role of the interaction between RNAP and DNA helicases is still to be determined.

1.4.1.1 SWI/SNF proteins – RapA and HARP

RapA (HepA) is a protein that has been identified as a subunit of the *E. coli* RNAP complex (Sukhodolets & Jin, 1998; McKinley & Sukhodolets, 2007). RapA shares similarity with the eukaryotic SWI/SNF proteins through six conserved helicase-like

motifs (Bork & Koonin, 1993). The functional role of the helicase-like domains in the SWI/SNF protein family remains unclear and none have shown helicase activity *in vitro* (McKinley & Sukhodolets, 2007). RapA has been shown to bind to RNAP core at the αβ' interface (Sukhodolets & Jin, 2000). RapA has a role in the recycling of the RNAP during *in vitro* transcription. Results showed that RapA activates transcription in multiple round assays, but not single round assays (Sukhodolets *et al*, 2001). It was proposed that the post-transcription/post-termination complex (PTC) was remodeled upon RapA binding allowing RNAP to enter a new cycle of transcription (Sukhodolets *et al*, 2001).

Another SWI/SNF family helicase is the eurkaryotic annealing helicase HARP (HepA-related protein) (Eisen *et al*, 1995; Yuan *et al*, 2009). HARP has no detectable helicase activity *in vitro*, but instead catalyzes the rewinding of Replication Protein A (RPA) coated ssDNA in a ATP dependent manner (Yusufzai & Kadonaga, 2008). HARP also functions to protect stalled replication forks, by binding to the ssDNA, which would otherwise be exposed to nucleases. Nuclease attack on ssDNA causes the collapse of replication forks and breaks in the DNA (Yuan *et al*, 2009). HARP can bind at the site of stalled replication forks to maintain the stability of the fork and allow for the resumption of replication without the requirement of homologous recombination (Driscoll & Cimprich, 2009). HARP can act as an annealing helicase on multiple DNA structures that arise when replication forks encounter DNA damage or lesions. Firstly it can counter-act helicase activity at the stalled fork, to prevent excessive unwinding (Figure 1.16I). Secondly it can prevent the unwinding of the leading and lagging strands following the dissociation of the DNA polymerase (Figure 1.16II & III) and finally HARP can prevent bubble formation behind the fork, which could potentially

cause incorrect processing of the DNA following replication (Driscoll & Cimprich, 2009).
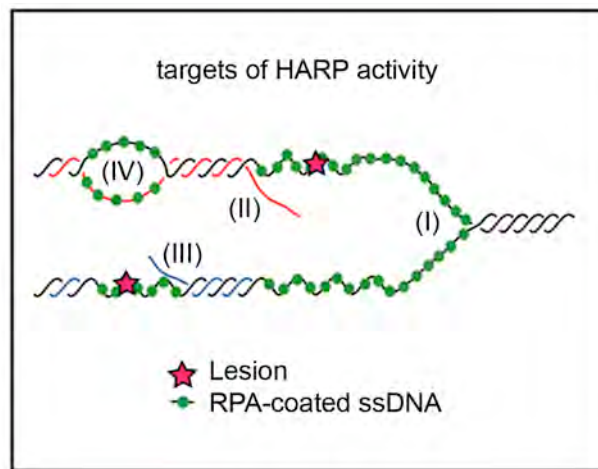
Figure 1.16: Targets of HARP activity at the replication fork. HARP can act to prevent events at stalled forks, which in turn stabilizes the forked structure. Such as: (I) limits helicase-mediated DNA unwinding. (II) prevents the unwinding of the leading strand as DNA polymerases become unstable and dissociate (III) and also prevents unwinding of the lagging strand and (IV) prevents bubble formation behind the fork in unstable regions. (Driscoll & Cimprich, 2009)

This role, of rewinding DNA in replication could also be a possible role for helicases interacting with RNAP. Replication can become stalled by collisions with head-on RNAP transcription complexes (Rudolph *et al*, 2007). The removal of RNAP requires additional factors, and Mfd (transcription-coupling repair factor, Section 1.3.3) has been shown to remove RNAP and promote the direct restart of the replication fork (Pomerantz & O'Donnell, 2010). It may be possible that the removal of stalled RNAP and restart of replication after these collisions could be performed by DNA helicases that interact with RNAP.

1.4.1.2 HelD

HelD (helicase IV, in *E. coli*) is a DNA helicase that unwinds DNA in the 3'-5' direction and has been implicated in the RecFOR pathway of recombination (Mendonca *et al*, 1993). HelD has also been identified as interacting with RNAP in *B. subtilis*

(Delumeau *et al*, 2011). Other helicases, PcrA and UvrD have also been implicated in the RecFOR pathway (Section 1.5.6.2).

1.4.1.3 TFIIH

Transcription factor IIH (TFIIH) is a eukaryotic basal transcription factor involved in the transcription of mRNA, rRNA and also in DNA repair (reviewed by Zurita & Merino, 2003). TFIIH consists of two sub-complexes, the core complex and the cyclin-activating kinase complex (CAK) (Zurita & Merino, 2003). Within the TFIIH complex there are two helicases, the ATP dependent DNA helicase *Xeroderma pigmentosum* complementation group B (XPB) in the core subunit and DNA-RNA helicase *Xeroderma pigmentosum* complementation group D (XPD) which links the two sub-complexes (Svejstrup *et al*, 1996; Zurita & Merino, 2003). XPB has 3-5' helicase activity required in both transcription and DNA repair, while XPD has 5-3' helicase activity utilized for DNA repair but not required during *in vitro* transcription (reviewed by Zurita & Merino, 2003).

Eukaryotic transcription of mRNA by RNA polymerase II (RNA Pol II) requires the essential function of the TFIIH helicase to cause DNA melting and the formation of the open promoter complex during transcription initiation (reviewed by Zurita & Merino, 2003) and promoter escape (Moreland *et al*, 1999). This is an important function of TFIIH because RNA Pol II does not have its own inherent helicase action as do RNA Pol I and III and bacterial RNA polymerases (Lin *et al*, 2005).

TFIIH DNA helicase mutational studies have shown that the ATP hydrolysis activity of TFIIH is required for DNA melting and that the helicase action is only required for

promoter clearance and the coupling of transcription with post transcriptional RNA

processing during elongation (Lin *et al*, 2005). Conversely the helicase mutants caused

the spontaneous melting of DNA without ATP hydrolysis, but acted to block promoter

clearance (Lin *et al*, 2005). If TFIIH was involved only during initiation then

transcription should occur without regulation. The results did not show this, and since

transcription was still defective this was attributed to non-functional open complex

formation and ineffective promoter clearance and elongation complex formation (Lin *et al*, 2005).

A role of TFIIH during elongation has also been implicated in the transcription of rRNA

genes by RNA polymerase I (RNA Pol I) (Iben *et al*, 2002). It was shown that TFIIH

has functions after initiation in RNA Pol I transcription and is not required for the

formation of the open promoter complex. To test if TFIIH was required during

initiation or elongation of rRNA by Pol I, abortive initiation assays were performed.

Results showed that complexes without TFIIH could undergo rounds of abortive

initiation. Full-length transcripts could not be produced in the absence of TFIIH. This

indicated that TFIIH had a role subsequent to transcription initiation, such as during

promoter clearance or elongation (Iben *et al*, 2002).

TFIIH complex also has roles in nucleotide excision repair (NER); the helicase subunits

of TFIIH (XPB) and (XPD) catalyse the unwinding of the dsDNA surrounding the DNA

lesion, targeted for repair (Weeda *et al*, 1990). TFIIH initially catalyses the opening of

the DNA, replication protein A (RPA) then binds to this region and ATP-driven

unwinding of the DNA lesion continues (Evans *et al*, 1997). This opening by TFIIH

generates a junction between dsDNA and the ssDNA lesion, which is targeted by

nuclease cleavage (Evans *et al*, 1997). The opening for the dsDNA surrounding a lesion targeted for repair by TFIIH is similar to the opening of the promoter complex by TFIIH during Pol II transcription (Weeda *et al*, 1990).

1.4.2 PcrA, UvrD and Rep

Rep, UvrD and PcrA are DNA helicases and members of the SF1 superfamily that have a DExx signature (Subramanya *et al*, 1996) and are considered to have a low processivity *in vitro*. Processivity is the number of base pairs unwound or translocated before the helicase dissociates from the DNA (Abdelhaleem, 2010). Low processivity means the helicases cannot unwind long strands of duplex DNA continuously, and *in vivo* they recruit other proteins to interact with and increase the activity of the helicase (Noirot-Gros *et al*, 2002b). PcrA and Rep can unwind short duplex regions *in vitro* (Yancey & Matson, 1991; Soultanas *et al*, 1999), while UvrD can unwind longer duplexes with decreased efficiency (Matson & George, 1987). The helicase activity can be increased by interaction with other proteins, for example although a PcrA monomer has a low processivity *in vitro* it can form a complex with an accessory protein RepD which increases its processivity (Soultanas *et al*, 1999; Zhang *et al*, 2007). The β-propeller protein YxaL is another protein shown to increase the processivity of PcrA helicase action *in vitro* (Noirot-Gros *et al*, 2002b), although the effects were less than that of RepD (Zhang *et al*, 2007).

1.5 PcrA

PcrA (plasmid copy number reduction) is a DNA helicase found to be essential for cell viability in Gram-positive bacteria (Petit *et al*, 1998; Chaudhuri *et al*, 2009). The *pcrA* gene has been identified in many Gram-positive bacteria, including but not limited to: *Bacillus anthracis, Geobacillus stearothermophilus, Bacillus subtilis, Lactococcus lactis, Staphylococcus aureus, Streptococcus pyogenes* and *Streptomyces coelicolor* (Petit *et al*, 1998; Chang *et al*, 2002; Naqvi *et al*, 2003; Anand *et al*, 2004). Although the *pcrA* gene has been identified in a number of Gram-positive bacteria the information available is still limited, particularly concerning its biological function. The alignment of the PcrA protein sequence (Appendix I) shows the protein is highly conserved through Gram-positive bacteria.

The crystal structure of *G. stearothermophilus* PcrA has been resolved in a complex with DNA (Figure 1.18A) (Velankar *et al*, 1999) and the biochemical characterisation of PcrA has been undertaken using the helicases from *S. aureus, G. stearothermophilis* and *B. anthracis* (Bird *et al*, 1998; Chang *et al*, 2002; Naqvi *et al*, 2003). The biochemical characterisation of PcrA has shown that it is a 3' – 5' DNA helicase in most bacterial species, but has been found to have bi-directional helicase activity in *S. aureus* and *B. anthracis* (Chang *et al*, 2002; Naqvi *et al*, 2003). This means *S. aureus* PcrA can unwind dsDNA in both the 3' to 5' and 5' to 3' directions, although studies showed it was stronger in the 3' to 5' direction, similar to other uni-directional PcrAs (Naqvi *et al*, 2003).

B. subtilis PcrA has been shown to have roles in plasmid rolling circle replication and the RecFOR pathway (see section 1.5.3; Petit *et al*, 1998; Petit & Ehrlich, 2002). PcrA is essential for cell viability and this has been attributed to its role in the RecFOR recombination repair pathway (Petit *et al*, 1998). *B. subtilis* PcrA has an open reading frame of 2220 nucleotides and a theoretical molecular weight of 83.5 kDa and was found to be 40 % identical to the Rep and UvrD helicases in *E. coli* and 61 % to *Staphylococcus aureus* PcrA (Petit *et al*, 1998).

1.5.1 PcrA structure

The crystal structure of *G. stearothermophilus* PcrA (Figure 1.18A) indicates that it is a monomeric protein with structural and sequence homology to Rep and UvrD (Subramanya *et al*, 1996; Velankar *et al*, 1999). PcrA has four domains, 1A, 1B, 2A and 2B (Figure 1.18B). Domains 1A and 2A are separated by a cleft, in which the ATP-binding and DNA-binding sites are located (Figure 1.18A; Subramanya *et al*, 1996; Velankar *et al*, 1999). The conserved SF1 helicase motifs are located in domains 1A and 2A (Figure 1.18C). The structure of PcrA represents a 'crab claw' with a large and small pincer (NB. This is not related to the crab claw structure of RNAP). PcrA translocates along ssDNA towards the ssDNA/dsDNA junction in the 3'-5' direction and the ssDNA threads through the core at the interface of the 1A/2A and 1B/2B domains (Dittrich & Schulten, 2006). The seven conserved helicase motifs are located on the bottom and side of the cleft, forming the protein interaction domain (Korolev *et al*, 1997; Marians, 1997). The structure of the extreme C-terminal region of PcrA was not solved in any of the published PcrA structures (Figure 1.18D).
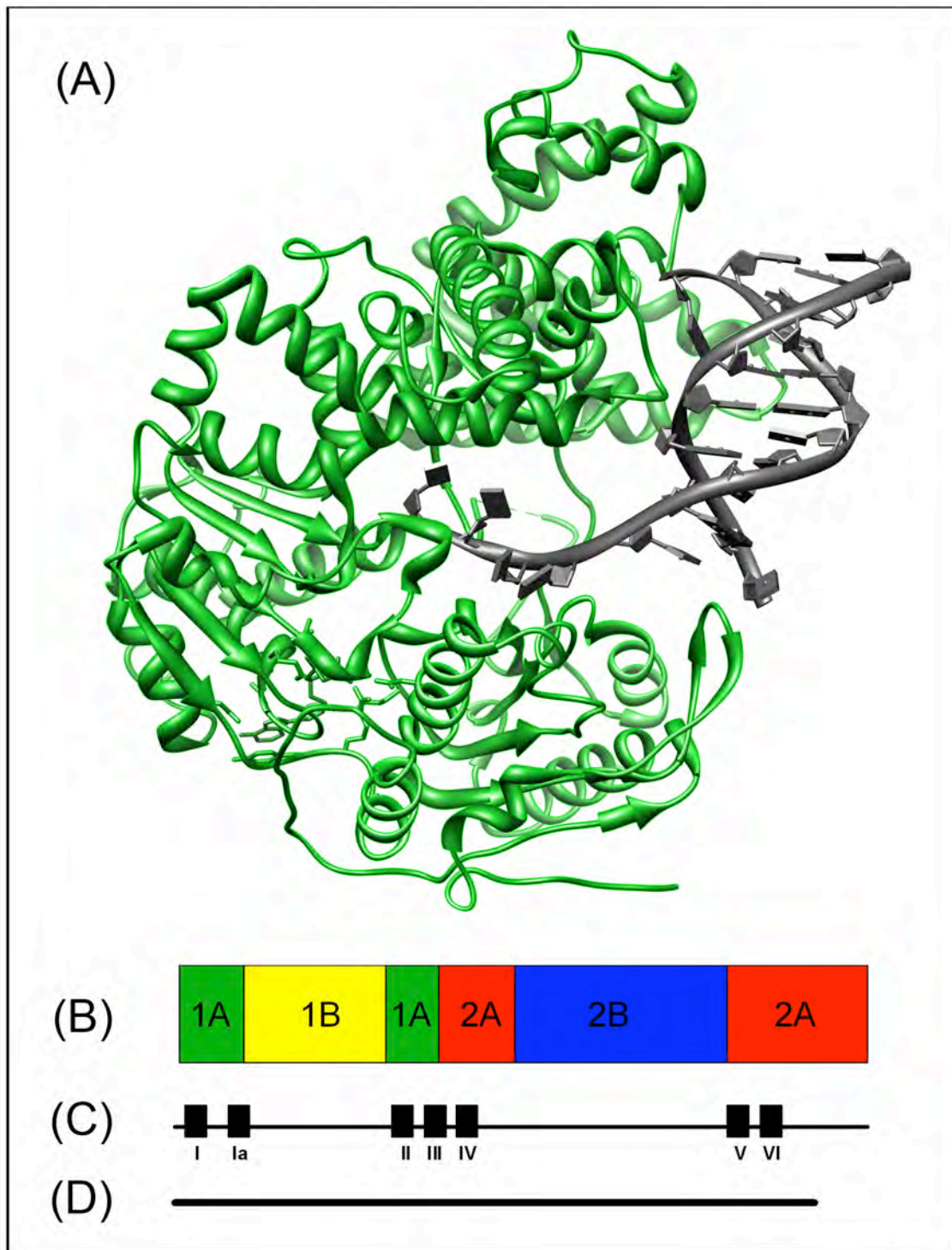
Figure 1.18: *G. stearothermophilus* PcrA. (A) The crystal structure of PcrA bound to DNA (3PRJ). (B) The linear domain structure, showing all four subunits, 1A and 2A form the cleft. (C) The location of the seven SF1 motifs in the PcrA domains (Subramanya *et al*, 1996). (D) The thick line is the first 652 aa of the PcrA structure that were resolved using X-ray crystallography.

PcrA, UvrD and Rep have similar structures, all containing 4 sub-domains (1A, 2A, 1B and 2B) and it has been shown that the 2B domain of a Rep monomer can rotate to change the conformation of the protein monomer (Rasnik *et al*, 2004). Studies into this 2B region of Rep showed that a mutated form with the 2B sub-domain deleted (RepΔ2B) has an increased monomeric helicase activity compared to the wildtype enzyme and that the 2B region could be auto-inhibitory (Cheng *et al*, 2002; Brendza *et al*, 2005). Following this it has been suggested that the 2B sub-domain may also have an auto-inhibitory role in PcrA and UvrD (Niedziela-Majka *et al*, 2007). This has yet to be investigated further, although earlier results from an investigation of PcrA activity showed point mutations in the 2B region caused a decrease in helicase activity (Soultanas *et al*, 2000), which does not support the 2B region having an auto-inhibitory role. In fact when two specific point mutations were introduced together the relative helicase activity was only 3 %, when compared to wild type enzyme (Soultanas *et al*, 2000).

1.5.2 DNA binding properties of PcrA

Studies have shown *S. aureus* PcrA binds with the greatest affinity *in vitro* to DNA hairpins that are followed by a single strand overhang (Anand & Khan, 2004), which allows the entry of the single strand region into the DNA binding site of PcrA. The sequence of the hairpin can by altered, and/or the loop removed without adversely affecting the binding affinity of PcrA, which indicates that it is the structure, not the DNA sequence that is the recognition site for PcrA. Other forms of DNA structures, such as single stranded, duplexes (forked, bubble, blunt or tailed) have a lowered binding recognition for PcrA (Anand & Khan, 2004).

1.5.3 PcrA helicase activity

It was believed PcrA helicase bound as a monomer to DNA and could bind to both double and single stranded DNA (Chang *et al*, 2002). The monomeric nature of the crystallized PcrA could indicate that the mechanism of helicase unwinding activity may act via an 'inchworm' rather than a rolling model (Bird *et al*, 1998; Velankar *et al*, 1999).
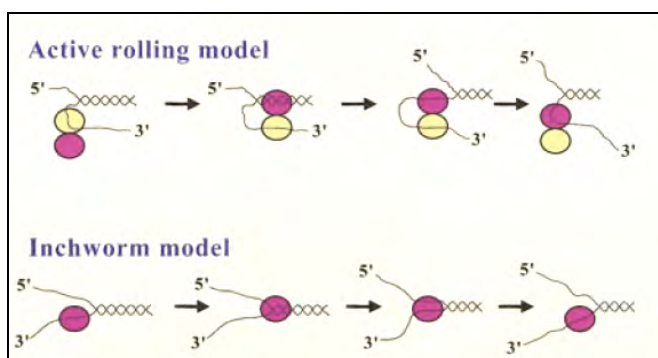
Figure 1.19: Active rolling and inchworm models of helicase activity. The active rolling model requires a protein dimer to bind to single and double stranded, whereas the protein monomer binds to both the single and double stranded DNA at same time in the inchworm model (Velankar *et al*, 1999).

The rolling model requires at least a dimeric protein because the protein monomer can only bind to single stranded or double stranded DNA, not both at the same time, in comparison the inchworm model allows the binding of both single and double stranded DNA by the protein monomer (Figure 1.19) (Velankar *et al*, 1999). The inchworm model proposed for PcrA activity has been modified from the original inchworm model (Yarranton & Gefter, 1979) and is divided into two separate processes; destabilization of DNA duplex ahead of the helicase and translocation along the single stranded DNA (Velankar *et al*, 1999). Further studies showed that these two processes, although coupled, occur separately, in different regions of PcrA (Soultanas *et al*, 2000; Yu *et al*, 2007). This also showed that PcrA helicase activity is an active mechanism, which means PcrA plays a direct role in destabilizing the dsDNA (Lohman & Bjornson, 1996). An active mechanism requires the helicase to bind to both ssDNA and dsDNA at the duplex junction and conformational

changes in the helicase, triggered by NTP binding and hydrolysis unwinds the dsDNA (Lohman & Bjornson, 1996). An alternative mechanism is the passive mechanism, which occurs when a DNA helicase unwinds DNA indirectly by binding to the ssDNA and translocating towards the duplex junction (Lohman & Bjornson, 1996)

The bidirectional helicase activity of PcrA in some bacteria may indicate PcrA has more than one function within the cell that requires helicase activity in either or both directions (Naqvi *et al*, 2003). Investigation of the bidirectional helicase activity has since indicated that PcrA is a stronger helicase in the 3' – 5' direction (Bird *et al*, 1998; Anand & Khan, 2004). More contemporary research into helicase activity using single-round experiments would be useful to confirm if PcrA possesses true bidirectional helicase activity or if that early result was an artifact of experimental design. This is because it was determined that *S. aureus* PcrA had bidirectional helicase activity using the multiple turnover helicase assays with an excess of PcrA (Naqvi *et al*, 2003; Anand & Khan, 2004). Additional research has also suggested that the bidirectionality of helicase activity *in vivo* in plasmid replication is unlikely (Zhang *et al*, 2007). Importantly *G. stearothermophilus* PcrA has been shown to be a 3' – 5' unidirectional helicase (Bird *et al*, 1998; Dillingham *et al*, 2000). The directionality of the *B. subtilis* PcrA has not yet been reported, although it thought to act in the 3' – 5' direction. This is supported by the characterisation of PcrA from different bacterial species, which have shown to have helicase activity dominant in the 3'-5' direction.

The bidirectional helicase activity of PcrA is unusual and is an area of study that needs further investigation. The PcrA protein sequences from different bacterial species have a high similarity (Appendix I) and the differences between the uni- and bi-directional species

may provide insight. Recent *in silico* studies on the translocase activity of DNA helicases have provided some possible amino acids that maybe important in determining the directionality of *B. subtilis* PcrA. The protein sequences of *G. stearothermophilus* and *S. aureus* PcrA were compared to identify possible regions that could control polarity (Yu *et al*, 2007). The comparison focused on the coevolutionary core regions of the protein, proposed to control the function of the protein. The study found 7 different amino acids between *G. stearothermophilus* and *S. aureus* translocation domains (1A and 2A), two of which are in helicase motif IV and could be the most important in determining translocation polarity (Yu *et al*, 2007). The residues are isoleucine at 596 (I596) and alanine at 600 (A600) in *S. aureus* and leucine at 603 (L603) and glycine at 607 (G607) in *G. stearothermophilus* (residues are shifted seven in *G. stearothermophilus*) (highlighted in Figure 1.20, Yu *et al*, 2007). When the *B. subtilis* PcrA is compared to the *G. stearothermophilus* and *S. aureus* sequence the two residues are a leucine and a glycine, as in *G. stearothermophilus* (Figure 1.20). This could be an initial indication that *B. subtilis* PcrA is a unidirectional translocase, although more experimental work must be performed to determine the polarity of *B. subtilis* PcrA.

```
gste 599  EERRLAYVGITRAEEELVLTSAQMRTLFGNIQMDPPSRFLNEIPAHLLETASR--------
bsub 600  EERRLAYVGITRAEQELYLTNAKMRTLFGRTNMNPESRFIAEIPDDLLENLNEKKETRATS
saur 592  EERRICYVAITRAEEVLYITHATSRMLFGRPQSNMPSRFLKEIPESLLENHSSGKRQTIQP
```

Figure 1.20: Alignment of *G. stearothermophilus* (gste)*, B. subtilis* (bsub) and *S. aureus* (saur) PcrA protein sequences from amino acids 599-651 (gste), 600-666 (bsub), 592-658 (saur). Highlighted in pink are the two amino acid residues thought to influence the translocation polarity (Yu *et al*, 2007).

This work provides a possible theory for helicase directionality, but when it was examined across a broader range of bacterial species it was found that it is unlikely that these two residues would influence polarity. For example *B. anthracis* PcrA was also implicated as possessing bidirectional activity (Naqvi *et al*, 2003) but it contains the highly conserved leucine and glycine the same as the unidirectional *G. stearothermophilus* PcrA.

1.5.4 PcrA translocation

The structure of *G. stearothermophilus* PcrA has been resolved using X-ray crystallography (PDB:1PJR) and additional structures show PcrA in complex with DNA (PDB:3PJR) and in complex with an ATP analogue, representing the "substrate" complex (PDB:2PJR). Characterisation studies have confirmed PcrA is a unidirectional helicase that moves in the 3' to 5' direction (Dillingham *et al*, 2000; Dillingham *et al*, 2002). Studies have also indicated PcrA uses an inchworm mechanism involving domains 1A and 2A (Figure 1.21) to translocate along ssDNA (Velankar *et al*, 1999; Yu *et al*, 2006; Yu *et al*, 2007). The rate of translocation is thought to occur at ~50 nucleotides per second, utilising one ATP molecule per nucleotide step (Dillingham *et al*, 2000).
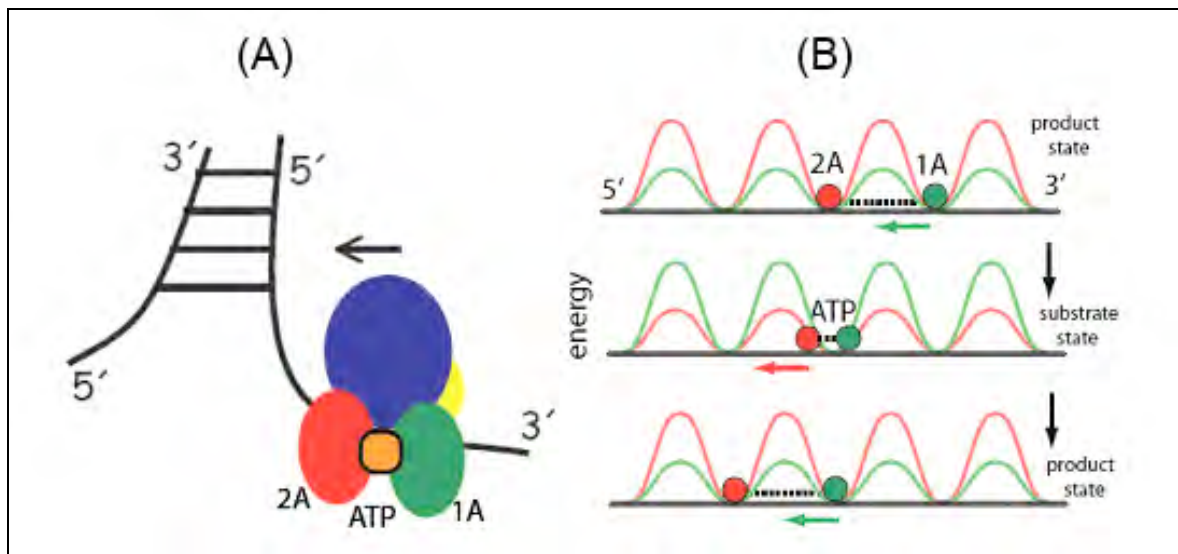
Figure 1.21: (A) PcrA translocating along ssDNA in the 3'-5' direction. The domains of PcrA are coloured: 1A green, 2A red, 1B yellow and 2B blue. ATP is represented by the orange square. (B) a simple physical model to explain PcrA translocation through alternating domain mobilities. In the product state (no ATP bound) 1A domain (green) has a low energy barrier and moves towards the 2A domain (red). In the substrate state ATP is bound, this lowers the energy barrier of the 2A domain, which moves forward on the DNA, releasing ADP (Yu *et al*, 2007).

Protein crystal structures have shown that domains 1A and 2A move closer together when ATP is bound (substrate state) and the domains move farther apart with the release of ADP (product state) (Yu *et al*, 2006; Yu *et al*, 2007). Domain 1A faces the 3' end of the

ssDNA, while 2A faces the 5' end (Figure 1.19A). When PcrA is in the product state 1A is more mobile and moves forward, toward 2A. Upon binding of ATP, PcrA is in the substrate state, where the mobility is reversed and the 2A domain can now move forward, away from 1A (Figure 1.21B) (Yu *et al*, 2007).

Further investigation using *G. stearothermophilus* PcrA has indicated that PcrA monomers are ssDNA translocases *in vitro* but are not processive helicases (Niedziela-Majka *et al*, 2007). This means PcrA can move along the ssDNA but does not unwind the dsDNA, as it would if it was acting as a helicase. This study utilised single turnover experiments, where the protein monomers were pre-bound to the DNA substrates before the reaction was initiated by the addition of ATP (Niedziela-Majka *et al*, 2007). Previously it had been assumed that PcrA acted a monomeric helicase, due to SF1 helicases being widely considered to act as monomers (Soultanas & Wigley, 2000; Soultanas & Wigley, 2001). This assumption was based on a number of results: Firstly, the crystal structure of *G. stearothermophilus* PcrA is in the monomeric form when bound to DNA (Subramanya *et al*, 1996). Secondly, characterisation studies carried out that showed only PcrA monomers were detected using gel filtration at high salt concentrations (Bird *et al*, 1998). Thirdly, it was shown that PcrA could translocate in the 3' – 5' direction as a monomer (Dillingham *et al*, 2000; Dillingham *et al*, 2002).

Until the study conducted by Niedziela-Majka *et al.,* 2007 the helicase assays performed were using multiple turnover conditions, which showed that an excess of PcrA effectively unwound duplex DNA (Soultanas *et al*, 2000). When single round conditions were utilised, the results indicated that although PcrA monomers could processively translocate along ssDNA, PcrA monomers could not unwind duplex DNA as short as 18 bp *in vitro*

(Niedziela-Majka *et al*, 2007). These types of experiments have also been performed using similar SF1 proteins *E. coli* Rep and UvrD (Cheng *et al*, 2001; Maluf *et al*, 2003) and have shown similar results. Both UvrD and Rep monomers have no helicase activity *in vitro* while still being processive translocases (Cheng *et al*, 2001; Maluf *et al*, 2003). These results suggest that the translocation and helicase activities of PcrA, UvrD and Rep are separate actions of these enzymes and furthermore, although translocation is necessary it is not sufficient to give helicase activity (Niedziela-Majka *et al*, 2007). It is possible the helicase activity of PcrA may instead occur through dimerisation.   PcrA monomers have been reported to have translocase activity but no helicase activity in single round experiments (Niedziela-Majka *et al*, 2007). Other studies have shown PcrA can unwind dsDNA but further investigation shows this activity is dependent on PcrA concentration and the length of the 3'-ssDNA tail in the DNA substrate *in vitro* (Yang *et al*, 2008b). Higher PcrA concentrations and longer 3'-ssDNA tail substrates are required for helicase activity, which may suggest that two molecules are bound on the 3'-ssDNA tail to begin dsDNA unwinding (Yang *et al*, 2008b).

1.5.5 Protein-protein interactions of PcrA

PcrA has been shown to be an abundant protein in Gram-positive bacteria, which has been shown to bind to a number of proteins within the cell (Figure 1.22). Firstly PcrA interacts with accessory proteins, known as processivity factors – YxaL and RepD. These proteins can increase the helicase activity of PcrA and aid in the loading of PcrA onto DNA (Noirot-Gros *et al*, 2002b; Zhang *et al*, 2007). PcrA also displaces RecA protein bound to DNA and inhibits RecA-mediated DNA strand exchange (Anand *et al*, 2007; Park *et al*, 2010). RepC interacts with PcrA during plasmid rolling circle replication (Section 1.5.6.1;

Anand *et al*, 2005). PcrA has also been shown to bind to UvrB and UvrC proteins during nucleotide excision repair (NER) and can displace the UvrC and excised DNA from a post-incision NER complex (Manelyte *et al*, 2009). Finally yeast two hybrid genomic screening identified interactions between PcrA and the β subunit of RNAP and additional unknown proteins – YerB and YwhK (Noirot-Gros *et al*, 2002a; Noirot-Gros *et al*, 2002b).
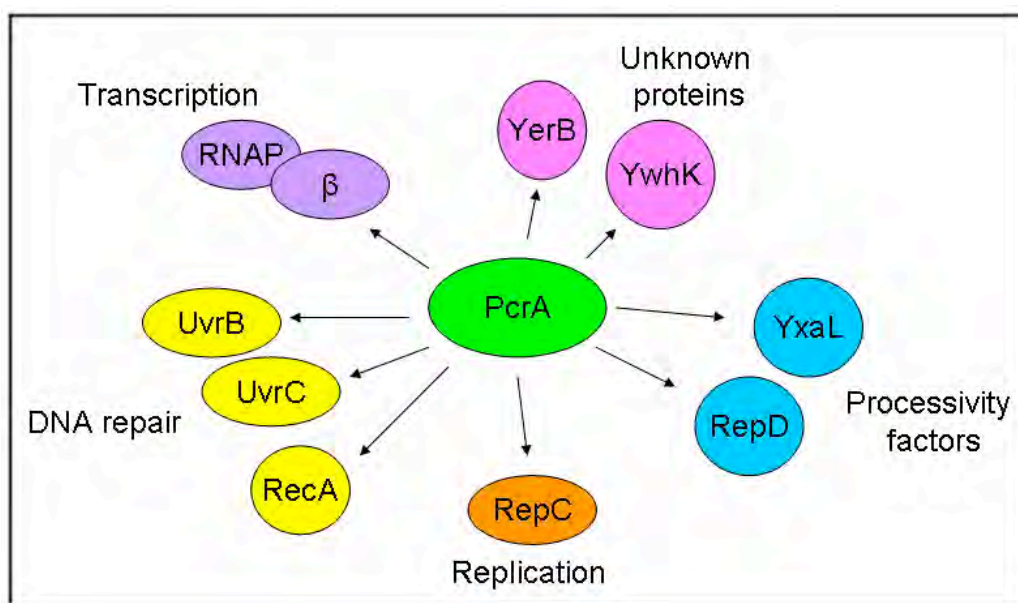
Figure 1.22: Known protein interactions with PcrA in Gram-positive bacteria. PcrA interacts with proteins that have roles in a variety of essential cell processes.

1.5.6 Functions of PcrA

Although PcrA has been found to be an essential helicase in the Gram-positive bacteria (Petit *et al*, 1998; Chaudhuri *et al*, 2009) the exact role of PcrA within the cell is not known (Naqvi *et al*, 2003).

1.5.6.1 Plasmid rolling circle replication

PcrA has been shown to have a role in plasmid rolling circle replication (Petit *et al*, 1998) and an essential role in homologous recombination repair (Petit & Ehrlich, 2002). Other possible roles of the PcrA helicase have been drawn from its similarity to other SFI

helicases, such as Rep, UvrD and RecB(CD) (Petit *et al*, 1998; Chang *et al*, 2002; Naqvi *et al*, 2003) and these roles may include DNA repair (Soultanas *et al*, 2000), removal of unwanted structures, such as RNA-DNA hybrids (R-loops) and nucleoprotein complexes (protein roadblocks) (Noirot-Gros *et al*, 2002b). Studies using PcrA mutants have also shown that PcrA does not play a major role in chromosome replication in *B. subtilis* (Petit *et al*, 1998). This may suggest PcrA has more similarities to UvrD than Rep, as Rep has been shown to have a role in replication. Additionally, PcrA and UvrD are abundant proteins within bacterial cells, 8000 and 5000 molecules/cell respectively, whereas Rep is present at lower levels, at only 50 molecules/cell (Scott & Kornberg, 1978; Arthur & Eastlake, 1983; Petit & Ehrlich, 2002).

PcrA was initially proven to have a function in rolling circle replication (RCR) in the pT181 plasmid (Iordanescu, 1993; Petit *et al*, 1998). The initiation of RCR in the Gram-positive bacterium *S. aureus* plasmid pT181 begins with a nick at the double stranded origin (dso) by the RepC initiator protein. PcrA interacts directly with the RepC protein (Chang *et al*, 2002; Anand *et al*, 2004) and begins to unwind the dsDNA, allowing replication of the plasmid to occur (Iordanescu, 1993; Anand *et al*, 2005). Mutational analysis showed the helicase activity of PcrA was required to replicate the plasmid DNA, whereas the DNA binding affinity was not affected (Anand *et al*, 2005).

1.5.6.2 RepD and PcrA during plasmid replication

Interaction between PcrA and the replication initiator protein RepD has between shown to increase the processivity of the PcrA helicase during plasmid replication (Soultanas *et al*, 1999; Zhang *et al*, 2007). RepD increases the processivity of PcrA by enhancing the affinity of PcrA for the *oriD* DNA and does not effect the ATPase cycle of PcrA during

plasmid replication (Toseland *et al*, 2009). PcrA is the host DNA helicase which is required for plasmid rolling circle replication (Iordanescu, 1993). Studies performed by Zhang *et al.,* (2007) have shown that RepD not only increases the helicase activity of *G. stearothermophilus* PcrA but can influence the loading of PcrA onto the DNA strands at the origin (*oriD*) during plasmid replication. Without an interaction with RepD, PcrA can bind to the nicked DNA and translocate in the 3'-5' direction but cannot bind to an intact origin (Zhang *et al*, 2007). RepD can load PcrA onto the other DNA strand, the continuous strand of the origin (Figure 1.23) and can also recruit PcrA to an intact origin (Zhang *et al*, 2007).
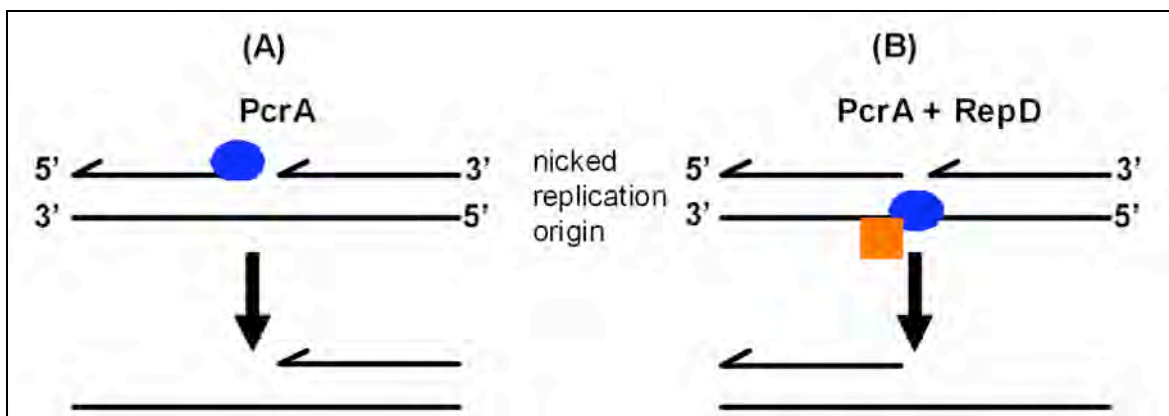
Figure 1.23: The *in vitro* results of plasmid replication origin loading of PcrA with and without processivity factor RepD. (A) Without RepD, PcrA (blue circle) can bind to the nicked strand and translocate towards the 5' end of the DNA. (B) RepD (orange rectangle) can load PcrA onto the continuous DNA strand, PcrA still translocates in the 3'-5' direction (adapted from Zhang *et al*, 2007).

The fact that PcrA can be loaded onto the nicked *oriD* and not the intact *oriD* in the absence of RepD suggests that PcrA has an affinity for nicked DNA substrates rather than a specific feature of *oriD*. UvrD does not possess affinity for nicked DNA and requires protein interactions to be loaded onto DNA in processes such as DNA repair (Dao & Modrich, 1998), similar to the loading of PcrA onto the continous strand by RepD in plasmid replication (Zhang *et al*, 2007).

1.5.6.3 Homologous recombination repair

Double stranded DNA breaks can occur in chromosomal DNA, caused by a variety of different events. One major source of dsDNA breaks is during replication of damaged DNA which causes the replication fork to collapse. There are two different approaches for the repair of the beaks: non-homologous end joining (NHEJ) and homologous recombination (HR) (reviewed by Yeeles & Dillingham, 2010). Homologous recombination is the process of converting parental DNA substrates into recombinant DNA via a series of enzymatically catalyzed reactions that can be divided into four major steps: 1 – Initiation, where the DNA substrate is generated, 2 – strand pairing and exchange, 3 – branch migration and finally 4 – branch resolution (Eisen & Hanawalt, 1999).

Homologous recombination is important for maintaining gene diversity and genome stability (Fernandez *et al*, 2000) and is essential for the repair of DNA lesions and breaks (Amundsen & Smith, 2003). Homologous recombination can process stalled or blocked replication forks by utilizing RecA protein to restart replication (Michel, 2000). Recombination reactions are subdivided into three stages: (I) formation of the ssDNA substrate with the RecA protein bound to form the nucleoprotein complex, (II) catalysis of the homologous pairing by the RecA-filament, and finally (III) the resolution of the DNA strands in homologous recombination (Mendonca *et al*, 1995). The recombination machinery usually requires three different types of enzymes: helicases, nucleases and synapsis proteins (Amundsen & Smith, 2003).

There are two major pathways of recombination in *E. coli*, the RecBCD pathway and the RecFOR pathway (Kuzminov & Stahl, 1999; Amundsen & Smith, 2003). The RecBCD (AddAB in *B. subtilis*) pathway is initiated by breaks in double stranded DNA and must

unwind the dsDNA, which then allows the RecA protein to coat the ssDNA and form a joint molecule intermediate, such as a Holliday junction (Bianco *et al*, 1998). RecFOR is an alternative pathway which does not act on broken dsDNA (Figure 1.24) and is instead used for RecA to bind to gapped ssDNA (Fernandez *et al*, 2000). RecF, with RecO and RecR proteins assist the binding of RecA to the single stranded DNA (Petit & Ehrlich, 2002). In *B. subtilis* a fourth gene *recL* is present but its function is still unknown (Alonso *et al*, 1988). After RecA binds to the RecFOR coated ssDNA recombination occurs, which creates the Holliday junction intermediate (Bianco *et al*, 1998). To process Holliday junctions and complete homologous recombination repair (Figure 1.24) RuvABC is required (Bianco *et al*, 1998). Unresolved Holliday junctions can accumulate and become toxic to the cell by preventing DNA replication and Okazaki fragment maturation.
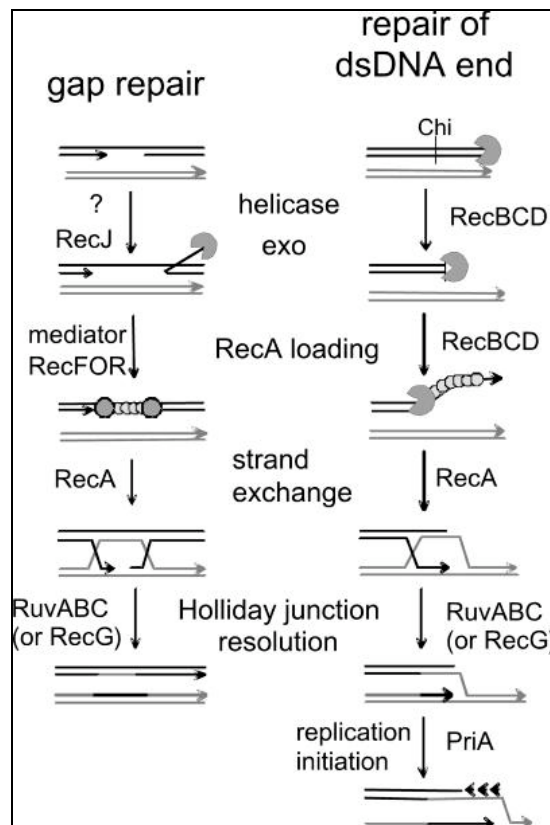
Figure 1.24: Models of the main homologous recombination pathways and the proteins involved. RecFOR is involved in gap repair and RecBCD in repairing dsDNA ends. Both pathways end in RecA mediated strand exchange and Holliday junction resolution (Rocha *et al*, 2005).

PcrA may also have a role resolving these Holliday junction intermediates in Gram-positive bacteria (Petit & Ehrlich, 2002). Studies have indicated PcrA has an essential function in homologous recombination repair. PcrA could play a role in reversing blocked recombination intermediates, Holliday junctions, or to counteract the RecFOR pathway and facilitate replication (Petit & Ehrlich, 2002). To show this, mutant PcrA proteins were constructed in *B. subtilis* and observed to be lethal. It was found that the lethality of PcrA mutations could be suppressed by mutations in the *recF, recO, recR* and *recL* genes (Petit & Ehrlich, 2002). Mutations introduced to other repair genes such as *rec*A and *add*A did not suppress the lethal PcrA mutations. This could indicate that the involvement of PcrA in homologous recombination is during the initial stages (Petit & Ehrlich, 2002), before RecA is recruited by the RecFORL proteins. Similar results were also obtained in *E. coli*, where *recF*, *recO* and *recR* mutations restored the viability of the *uvrD* mutants (Petit & Ehrlich, 2002). UvrD has been shown to reverse potentially toxic recombination intermediates *in vitro* (Flores *et al*, 2004; Flores *et al*, 2005).

1.5.6.4 Dismantling RecA filaments

Another function of PcrA is the removal of DNA bound proteins. It has been proposed that PcrA may have a role in the removal of RecA, bound at stalled replication forks. The role of RecA is to catalyse the pairing of homologous DNA during recombination that begins by the formation of RecA coated DNA filaments. Studies have shown that *G. stearothermophilus* PcrA can preferentially translocate along the lagging DNA strand, instead of unwinding the duplex DNA. PcrA binds to the duplex DNA (Figure 1.25A) and translocation begins in the 3' – 5' direction, which reels in the ssDNA, creating a loop structure (Figure 1.25B) in order to remove the RecA bound to the DNA (Park *et al*, 2010).

The ssDNA loop increases with the translocation of PcrA (Figure 1.24C), until PcrA reaches the end of the DNA and is released (Figure 1.25D).
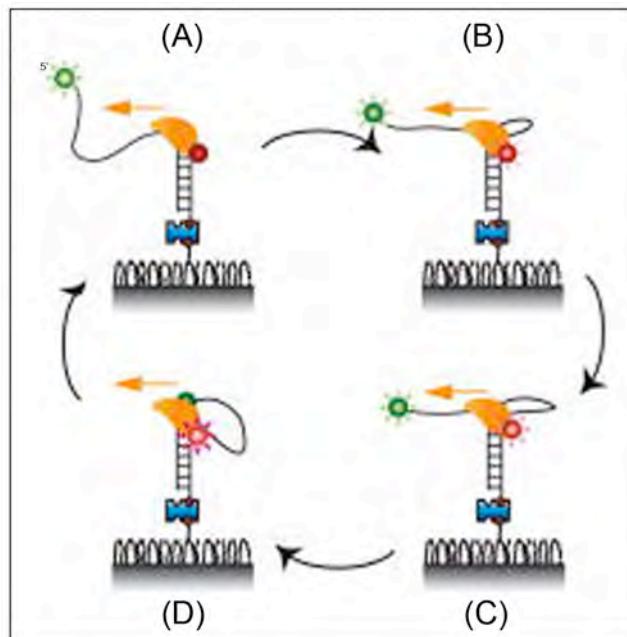
Figure 1.25: Reeling-in model of PcrA action to dismantle RecA filaments. RecA filaments form after replication has stalled to catalyse homologous recombination. (A) PcrA (in yellow) binds the DNA at the 5' duplex junction, where the ssDNA and dsDNA meet. (B) Translocation begins in the 3'-5' direction while PcrA maintains contact at the junction. (C) ssDNA loop formation, increases in size as PcrA translocation continues in the 3'-5' direction. (D) PcrA reaches the 5' end of the ssDNA and is released. Green and red circles are the FRET donor and acceptor molecules used to visualise the reeling in action of PcrA (Park *et al*, 2010).

UvrD in *E. coli* can also aid the restart of replication by the removal of the replication terminator protein (Tus) and can dismantle RecA bound recombination intermediates from a gap in the lagging strand of blocked replication forks (Veaute *et al*, 2005; Bidnenko *et al*, 2006). PcrA has been successfully substituted for UvrD to perform both these activities (Lestini & Michel, 2007; Park *et al*, 2010).

Overall PcrA has proven roles in plasmid replication and the RecFOR pathway of homologous recombination repair. It has been shown to be abundant with the cells and to interact with numerous other proteins, including RNAP. This suggests that PcrA could have additional housekeeping roles within Gram-positive bacteria.

1.6 Aims

The aim of this project was to characterise the interaction occurring between PcrA and RNAP. The interaction between PcrA and the β subunit of RNAP was initially identified when investigating the proteins of the replisome (Noirot-Gros *et al*, 2002a) and supported by additional yeast two-hybrid screening to identify new protein-protein interactions with RNAP.

The characterisation of the interaction initially involved *in vitro* binding studies, which were then extended to identify the specific regions of PcrA and RNAP that are involved in the interaction. PcrA is a large protein and RNAP is a multi subunit enzyme and therefore the exact position of the interaction could influence the possible role that PcrA could be carrying out during transcription. Investigation of the interaction was also extended to include a homologue of PcrA from Gram-negative bacteria – UvrD. The interaction was also examined *in vivo* in *B. subtilis,* using mutagenic analysis. PcrA is an essential protein and therefore point mutations, which disrupted the interaction with the β subunit, were identified and introduced into the chromosome. It was hoped that the successful characterisation of this interaction would provide insight into the possible roles PcrA may have during transcription.

Chapter 2:

Materials and Methods

Chapter 2: Materials and Methods

2.1 Media

All solutions used are outlined in Appendix II.

2.1.1 Solid Growth Media

E. coli strains were grown on nutrient agar (NA) (Appendix II) containing the antibiotic ampillicin (100 µg/ml) when required. Plates were incubated inverted, overnight at 37 °C.

B. subtilis strains were grown on NA containing appropriate antibiotic and supplements when required. Antibiotics included chloramphenicol (5 µg/ml), erythromycin (1 µg/mg) and supplements included 0.5 % (w/v) xylose and 0.5 mM IPTG. Plates were then incubated inverted, overnight at 37 °C.

2.1.2 Liquid Growth Media

E. coli cultures were grown in Luria-Bertani (LB) medium (Appendix II) containing ampicillin (100 µg/ml) with shaking for 6-16 h at 37 °C.

B. subtilis cultures were grown in LB containing required antibiotic and supplements overnight to reach stationary phase or to a known OD_{600} with shaking at 37 °C.

2.2 Bacterial Strains and plasmids

Table 2.1: Bacterial strains used and created in this work

Strain	Genotype	Source
E. coli		
DH5α	*F- endA1 hsdR17 supE44 thi-1 λ-recA1 gyrA96 relA1Δ(lacZYA-argF)U169φ80 dlacZ Δµ15*	Gibco BRL
BL21(DE3) pLysS	*F- ompT [lon] hsdSB (rB-mB-) λDE3 pLysS cmR*	(Studier & Moffatt, 1986)
C41 (DE3)	*F- ompT gal hsdSB (rB-mB-) dcm lon λDE3*	(Miroux & Walker, 1996)
B834 (DE3)	*F- ompT gal hsdSB (rB-mB-) met dcm lon λDE3*	Novagen
ER2566	*F- λ- fhuA2 [lon] ompT lacZ::T7 gene 1 gal sulA11 Δ(mcrC-mrr)114::IS10 R(mcr-73::miniTn10-TetS)2 R(zgb-210::Tn10)(TetS) endA1 [dcm]*	New England Biolabs
MCC217	*ER2566 and pSMG69*	(McGovern & Pollard, unpublished, 2006)
B. subtilis		
BS168	*trpC2*	C. Anagnostopoulous
HCR-9	*trpC2 uvrB9*	Okubo & Romig 1966
BS200	*trpC2 spo0A3 rpoC-6xHis spc*	Yang *et al.* 2009
EU299	*trpC2 chr::erm P_{wt} rpoB-SPA*	(Delumeau *et al*, 2011)Delumeau *et al.* 2011

BS494	*trpC2 chr:: cat P$_{wt}$ pcrA,P$_{xyl}$ 'pcrA*	This work: BS168 transformed with pNG515
BS495	*trpC2 chr:: cat P$_{wt}$ pcrA D692G, P$_{xyl}$ 'pcrA*	This work: BS168 transformed with pNG554
BS496	*trpC2 chr:: cat P$_{wt}$ pcrA E715G, P$_{xyl}$ 'pcrA*	This work: BS168 transformed with pNG763
BS497	*trpC2 chr:: cat P$_{wt}$ pcrA I718T, P$_{xyl}$ 'pcrA*	This work: BS168 transformed with pNG576
BS498	*trpC2 chr:: cat P$_{wt}$ pcrA A734V, P$_{xyl}$ 'pcrA*	This work: BS168 transformed with pNG724
BS500	*trpC2 chr:: cat P$_{wt}$ pcrA, P$_{xyl}$'pcrA, chr:: erm P$_{wt}$ rpoB-SPA*	This work: BS494 transformed with EU299 chromosomal DNA
BS501	*trpC2 chr:: cat P$_{wt}$ pcrA, P$_{xyl}$ 'pcrA D692G, chr:: erm P$_{wt}$ rpoB-SPA*	This work: BS495 transformed with EU299 chromosomal DNA
BS502	*trpC2 chr:: cat P$_{wt}$ pcrA E715G, P$_{xyl}$ 'pcrA, chr:: erm P$_{wt}$ rpoB-SPA*	This work: BS496 transformed with EU299 chromosomal DNA
BS503	*trpC2 chr:: cat P$_{wt}$ pcrA I718T, P$_{xyl}$ 'pcrA, chr:: erm P$_{wt}$ rpoB-SPA*	This work: BS497 transformed with EU299 chromosomal DNA
BS504	*trpC2 chr:: cat P$_{wt}$ pcrA A734V, P$_{xyl}$ 'pcrA, chr:: erm P$_{wt}$ rpoB-SPA*	This work: BS498 transformed with EU299 chromosomal DNA

bla, cat, erm, spc - ampicillin, chloramphenicol, erythromycin, spectinomycin resistance genes, respectively. *P$_{xyl}$* - xylose inducible promoter. *SPA* – 3xFlag and calmodulin binding peptide fusion tag for dual affinity purification.

Table 2.2: Plasmids used and created in this work

Plasmid	Genotype	Source
Parent cloning vectors		
pETMCSIII	*bla $P_{\varphi 10}$-6His'T_φ*	(Neylon *et al*, 2000)
pGEX 6P-1	*pSJ10ΔBam7stop7*	(GE Healthcare)
pSG1164	*bla cat P_{xyl}-gfpmut1*	(Lewis & Marston, 1999)
pNG209	*bla $P_{\varphi 10}$-'6HisT_φ*	(Yang & Lewis, 2008)
pNG698	*bla $P_{\varphi 10}$-'10His-StrepII-T_φ*	(Doherty, unpublished) 2009)
pNG612	*bla $P_{\varphi 10}$-6His'-bio - T_φ*	(Johnston, 2010)
pNG651	*bla $P_{\varphi 10}$-'3C-GST-T_φ*	(Yang, unpublished) 2007)
Far Western plasmids		
pNG490	*bla $P_{\varphi 10}$-6His-rpoB'(1-608 aa)-T_φ*	(Yang *et al*, 2009)
pNG479	*bla $P_{\varphi 10}$-6His-rpoB'(400-760 aa)-T_φ*	(Yang *et al*, 2009)
pNG480	*bla $P_{\varphi 10}$-6His-rpoB'(750-1040 aa)-T_φ*	(Yang *et al*, 2009)
pNG481	*bla $P_{\varphi 10}$-6His-rpo B'(950-1193 aa)-T_φ*	(Yang *et al*, 2009)
pNG482	*bla $P_{\varphi 10}$-6His-rpoC'(1-433 aa)-T_φ*	(Yang *et al*, 2009)
pNG483	*bla $P_{\varphi 10}$-6His-rpoC'(253-610 aa)-T_φ*	(Yang *et al*, 2009)
pNG484	*bla $P_{\varphi 10}$-6His-rpoC' (600-915 aa)-T_φ*	(Yang *et al*, 2009)
pNG485	*bla $P_{\varphi 10}$-6His-rpoC'(800-1199aa)-T_φ*	(Yang *et al*, 2009)
pNG750	*bla $P_{\varphi 10}$-6His-rpoC'(1-102aa)-T_φ*	(Kunalan, 2009)
pNG751	*bla $P_{\varphi 10}$-6His-rpoC'(103-228aa)-T_φ*	(Kunalan, 2009)
pNG752	*bla $P_{\varphi 10}$-6His-rpoC'(229-312aa)-T_φ*	(Kunalan, 2009)
pNG786	*bla $P_{\varphi 10}$- rpoC'(1-433aa)'3C-GST-T_φ*	(Yang, unpublished, 2010)
pNG841	*bla $P_{\varphi 10}$-rpoB'(1-109 aa)-T_φ*	This work: *rpoB* 1-109aa inserted into pNG209 via *Nde*I and *Acc*65I
pNG842	*bla $P_{\varphi 10}$-rpoB'(109-250 aa)-T_φ*	This work: *rpoB* 109-250aa inserted into pNG209 via *Nde*I and *Acc*65I

pNG857	*bla P$_{\varphi 10}$-rpoB'(250-427 aa)-T$_{\varphi}$*	This work: *rpoB* 250-427 aa inserted into pNG209 via *Nde*I and *Acc*65I
Protein overproduction		
pNG533	*bla P$_{\varphi 10}$-6His-TEV protease-T$_{\varphi}$*	
pSMG69	*bla lac^{q} P$_{\varphi 10}$-pcrA-T$_{\varphi}$*	(McGovern & Pollard, unpublished, 2006)
pNG511	*bla P$_{\varphi 10}$-6His-'pcrA(577-739aa)-T$_{\varphi}$*	(Harriott, 2006)
pNG512	*bla P$_{\varphi 10}$-6His pcrA'(1-214aa)-T$_{\varphi}$*	(Harriott, 2006)
pNG513	*pSJ10ΔBam7-'pcrA(577-739aa)-stop7*	(Harriott, 2006)
pNG514	*pSJ10ΔBam7-pcrA'(1-214aa)-stop7*	(Harriott, 2006)
pNG585	*bla P$_{\varphi 10}$-6His-E.coli uvrD-T$_{\varphi}$*	This work: UvrD inserted via *Xba*I and *Acc*65I
pNG587	*bla P$_{\varphi 10}$-6His-GSt pcrA-T$_{\varphi}$*	This work: GSt PcrA inserted via *Nde*I and *Acc*65I
pNG588	*bla P$_{\varphi 10}$-6His-GSt 'pcrA(567-739aa)-T$_{\varphi}$*	This work: GSt PcrA (567-724aa) inserted via *Nde*I and *Acc*65I
pNG822	*bla P$_{\varphi 10}$-6His-GSt pcrA'(1-312aa)-T$_{\varphi}$*	This work: GSt PcrA (1-312aa) inserted via *Nde*I and *Acc*65I
PcrA fragments		
pNG710	*bla P$_{\varphi 10}$-pcrA'(1-159aa)-6HisT$_{\varphi}$*	This work: PcrA 1-159aa inserted into pNG209 via *Nde*I and *Acc*65I
pNG711	*bla P$_{\varphi 10}$-pcrA'(1-159aa)-6His'-bio - T$_{\varphi}$*	This work: PcrA 1-159aa inserted into pNG612 via *Nde*I and *Acc*65I
pNG819	*bla P$_{\varphi 10}$- pcrA'(1-159aa)-'10His-StrepII-T$_{\varphi}$*	This work: PcrA 1-159aa inserted into pNG698 via *Nde*I and *Acc*65I
pNG712	*bla P$_{\varphi 10}$-pcrA'(1-258aa)-6HisT$_{\varphi}$*	This work: PcrA 1-258aa inserted into pNG209 via *Nde*I and *Acc*65I

pNG713	*bla P_{φ10}-pcrA'(1-258aa)-6His'-bio - T_φ*	This work: PcrA 1-258aa inserted into pNG612 via *Nde*I and *Acc*65I
pNG820	*bla P_{φ10}- pcrA'(1-258aa)-'10His-StrepII-T_φ*	This work: PcrA 1-258aa inserted into pNG698 via *Nde*I and *Acc*65I
pNG714	*bla P_{φ10}-pcrA'(1-258aa)-6HisT_φ*	This work: PcrA 1-258aa inserted into pNG209 via *Nde*I and *Acc*65I
pNG715	*bla P_{φ10}-pcrA'(1-258aa)-6His'-bio - T_φ*	This work: PcrA 1-258aa inserted into pNG612 via *Nde*I and *Acc*65I
pNG716	*bla P_{φ10}-pcrA'(1-498aa)-6HisT_φ*	This work: PcrA 1-498aa inserted into pNG209 via *Nde*I and *Acc*65I
pNG717	*bla P_{φ10}-pcrA'(1-498aa)-6His'-bio - T_φ*	This work: PcrA 1-498aa inserted into pNG612 via *Nde*I and *Acc*65I
pNG821	*bla P_{φ10}- pcrA'(1-498aa)-'10His-StrepII-T_φ*	This work: PcrA 1-498aa inserted into pNG698 via *Nde*I and *Acc*65I
pNG718	*bla P_{φ10}-pcrA'(1-590aa)-6HisT_φ*	This work: PcrA 1-590aa inserted into pNG209 via *Nde*I and *Acc*65I
pNG726	*bla P_{φ10}-pcrA'(1-590aa)-6His'-bio - T_φ*	This work: PcrA 1-590aa inserted into pNG612 via *Nde*I and *Acc*65I
pNG834	*bla P_{φ10}-pcrA'(1-85aa)'10His-StrepII-T_φ*	This work: PcrA 1-85aa inserted into pNG698 via *Nde*I and *Acc*65I
pNG835	*bla P_{φ10}-pcrA'(85-189aa)'10His-StrepII-T_φ*	This work: PcrA 85-189aa inserted into pNG698 via *Nde*I and *Acc*65I
pNG836	*bla P_{φ10}-pcrA'(189-310aa)'10His-StrepII-T_φ*	This work: PcrA 189-310aa inserted into pNG698 via *Nde*I and *Acc*65I

pNG837	*bla P$_{\varphi 10}$-pcrA'(310-498aa)'10His-StrepII-T$_{\varphi}$*	This work: PcrA 310-498aa inserted into pNG698 via *Nde*I and *Acc*65I
pNG838	*bla P$_{\varphi 10}$-pcrA'(1-85aa)'3C-GST-T$_{\varphi}$*	This work: PcrA 1-85aa inserted into pNG651 via *Nde*I and *Acc*65I
pNG839	*bla P$_{\varphi 10}$-pcrA'(189-310aa)'3C-GST-T$_{\varphi}$*	This work: PcrA 189-310aa inserted into pNG651 via *Nde*I and *Acc*65I
pNG840	*bla P$_{\varphi 10}$-pcrA'(310-498aa)'3C-GST-T$_{\varphi}$*	This work: PcrA 310-498aa inserted into pNG651 via *Nde*I and *Acc*65I

Mutagenesis

pNG515	*bla cat P$_{xyl}$-pcrA'(560-739aa)*	(Harriott, 2006)
pNG576	*bla cat P$_{xyl}$-pcrA'(560-739aa) I718T*	(Withers, 2007)
pNG554	*bla cat P$_{xyl}$-pcrA'(560-739aa)D692G*	This work: PcrA (aa) inserted into pSG1164 via *Acc*65I and *Spe*I. D692G mutation
pNG763	*bla cat P$_{xyl}$-pcrA'(560-739aa) E715G*	This work: PcrA (aa) inserted into pSG1164 via *Acc*65I and *Spe*I. E715G mutation
pNG724	*bla cat P$_{xyl}$-pcrA'(560-739aa) A734V*	This work: PcrA (aa) inserted into pSG1164 via *Acc*65I and *Spe*I. A734V mutation
pNG725	*bla cat P$_{xyl}$-pcrA'(560-739aa) I736T*	This work: PcrA (aa) inserted into pSG1164 via *Acc*65I and *Spe*I. I736T mutation
pNG574	*bla P$_{\varphi 10}$-'pcrA(577-739aa)6HisT$_{\varphi}$*	(Withers, 2007)
pNG575	*bla P$_{\varphi 10}$-'pcrA(577-739aa)6HisT$_{\varphi}$ I718T*	(Withers, 2007)
pNG684	*bla P$_{\varphi 10}$-'pcrA(577-739aa)6HisT$_{\varphi}$ D692G*	This work: PcrA (aa) inserted into pNG209 via *Nde*I and *Acc*65I. D692G mutation
pNG686	*bla P$_{\varphi 10}$-'pcrA(577-739aa)6HisT$_{\varphi}$ E715G*	This work: PcrA (aa) inserted into pNG209 via *Nde*I and *Acc*65I. E715G mutation

pNG720	*bla $P_{φ10}$-'pcrA(577-739aa)6HisT$_φ$ A734V*	This work: PcrA (aa) inserted into pNG209 via *Nde*I and *Acc*65I. A734V mutation
pNG721	*bla $P_{φ10}$-'pcrA(577-739aa)6HisT$_φ$ I736T*	This work: PcrA (aa) inserted into pNG209 via *Nde*I and *Acc*65I. I736T mutation
pNG832	*bla lac^q $P_{φ10}$-pcrA-T$_φ$ D692G*	This work: D692G mutation inserted into pSMG69
pNG824	*bla lacq $P_{φ10}$-pcrA-T$_φ$E715G*	This work: E715G mutation inserted into pSMG69

bla, cat - ampicillin, chloramphenicol resistance genes, respectively. $P_{φ10}$ - T7 promoter, P_{xyl} - xylose inducible promoter, 6His - 6 x histidine tag, *bio* - biotin tag, $T_φ$ - T7 transcription terminator

2.2.1 Storing bacterial stocks

2.2.1.1 E. coli

E. coli strains were grown overnight at 37 °C in LB with ampicillin. 500 µl of this

culture was mixed with 500 µl of sterile 50 % (v/v) glycerol and stored at -80 °C.

2.2.1.2 B. subtilis

B. subtilis strains were struck onto NA plates supplemented with appropriate antibiotic

(and 0.5 % (w/v) xylose if required) and grown at 37 °C overnight. 1 ml of sterile 50 %

(v/v) glycerol was inoculated, mixed and stored at -80 °C

2.3 DNA work

2.3.1 Polymerase Chain Reaction (PCR)

2.3.1.1 Oligonucleotide primers

The oligonucleotide primers (Sigma-Genosys) used in PCR amplification for cloning were designed to incorporate restriction enzyme recognition sites into the amplified DNA. This allowed for the ligation of the DNA and the plasmid within the appropriate restriction sites of the plasmid's multiple cloning site (MCS). Mutagenic and sequencing primers did not require the restriction enzyme recognition site.

Table 2.3: Oligonucleotide primers used in PCR amplification, including the sequence and the restriction site. Note: ( ) indicates the target vector.

Primer	Sequence 5' – 3'	Restriction site
PcrA		
PcrA 1 bp Fwd	G GCG GGT GAA CAA **CAT ATG** AAT TAT ATT AGC	*Nde*I
PcrA 489bp rev	GCC AGC AAC TTT AGA **GGT ACC** CTC CGG TTC	*Acc*65I
PcrA 790bp rev	GAT ATC CGC GCC **GGT ACC** TCT GTA GAT TGA	*Acc*65I
PcrA 1050bp rev	CTT CCG CTT GCC **GGT ACC** GTG AAG CTG ATG	*Acc*65I
PcrA 1490bp rev	TTC GAT CGA TTT **GGT ACC** CTT CAG CAT TTC	*Acc*65I
PcrA 1770bp rev	CAT GAG TGA ACG **GGT ACC** CGG GAA GAC GCC	*Acc*65I
PcrA 255bp Fwd	GGC GCG GAC **CAT ATG** TGG ATC TCC	*Nde*I
PcrA 255bp Rev	GGA GAT CCA **GGT ACC** GTC CGC GCC	*Acc*65I
PcrA 570bp Fwd	AAC CAG TGC **CAT ATG** TTC GAC GAT	*Nde*I
PcrA 570bp Rev	ATC GTC GAA **GGT ACC** GCA CTG GTT	*Acc*65I
PcrA 930bp Fwd	CCG AAA AAT **CAT ATG** ACG GAA AAT	*Nde*I

PcrA 930bp Rev	ATT TTC CGT **GGT ACC** ATT TTT CGG	*Acc*65I
PcrA 2217bp Rev	CC TTT CCT **GGT ACC** TTA CTG CTT TTC AAT AGG AGC	*Acc* 65I
PcrA CTD (pNG209) fwd	AGC CGT TCA **CAT ATG** GAG GAA GCG GAA	*Nde*I
PcrA CTD (pNG209) rev	TCC TTT CCT CTT **GGT ACC** CTG CTT TTC AAT	*Acc*65I
PcrA CTD (pSG1164) fwd	TTT CCG **GGT ACC** TTC TTG ATG GGG CTT	*Acc*65I
PcrA CTD (pSG1164) rev	GTC AGG CCA **ACT AGT** GCA TTA CTG CTT	*Spe*I

PcrA_{GSt}

PcrA_{Gst} fwd (pETMCSIII)	GTG GTG GAC **CAT ATG** AAT TTT TTA TCG GAA	*Nde*I
PcrA_{Gst} rev (pETMCSIII)	CCG TTC **GGT ACC** CTA TAC TTT CTC AAT CGC	*Acc*65I
PcrA_{Gst} NTD rev (pETMCSIII)	GAC GAT **GGT ACC** TTA TTC CGG GTT TTC CGT	*Acc*65I
PcrA_{Gst} CTD fwd (pETMCSIII)	ACG CTG CAC **CAT ATG** AAA GGG CTT GAG TTT	*Nde*I

UvrD

UvrD fwd (pETMCSII & pNG209)	5' TAC GCG GC**T CTA GA**A ATG GAC GTT TCT TAC	*Xba*I
UvrD rev (pETMCSIII)	TG CAT CCA **GGT ACC** TTA CAC CGA CTC CA	*Acc*65I
UvrD rev (pNG209)	A CTT ACG GCA **GGT ACC** CAC CGA CTC CAG	*Acc*65I

***rpoB* (β) subunit**

rpoB 1bp Fwd	GGG GTG AAT **CAT ATG** ACA GGT C	*Nde*I
rpoB 330bp Fwd	AAA GAC CAA **CAT ATG** TTC ATG GGT	*Nde*I
rpoB 330bp Rev	ACC CAT GAA **GGT ACC** TTA GTC TTT	*Acc*65I
rpoB 780bp Fwd	CGT TTC TTT **CAT ATG** AAA CGA TAC	*Nde*I
rpoB 780bp Rev	GTA TCG TTT **GGT ACC** TTA GAA ACG	*Acc*65I
rpoB 1300bp Fwd	GAG AGA ATG **CAT ATG** CAA GAT ACG	*Nde*I
rpoB 1300 Rev	CGT ATC TTG **GGT ACC** TTA TCT CTC	*Acc*65I
rpoB 1810 Rev	TAC ACG ATT **TCT AGA** AAC AAC G	*Xba*I

PcrA mutagenesis

D692G fwd	5' TGG GCA GTC GGA **GGT** AAG GCG GGC CAT AAA AAA TGG 3'	

D692G rev	5' CCA TTT TTT ATG GCC CGC CTT **ACC** TCC GAC TGC CCA 3'
W699R fwd	5' GCG GGC CAT AAA AAA **CGG** GGA ACA GGA ACT GTC 3'
W699R rev	5' GAC AGT TCC TGT TCC **CCG** TTT TTT ATG GCC CGC 3'
E715G fwd	5' GAA GGA GAA GGG ACA **GGG** CTT GAT ATT GCC TTC 3'
E715G rev	5' GAA GGC AAT ATC AAG **CCC** TGT CCC TTC TCC TTC 3'
I718T fwd	5' GGG ACA GAG CTT GAT **ACT** GCC TTC CCG AGC 3'
I718T rev	5' GCT CGG GAA GGC **AGT** ATC AAG CTC TGT CCC 3'
A734V fwd	5' CGC CTG TTA GCA GCA TTT **GTT** CCT ATT GAA AAG CAG 3'
A734V rev	5' CTG CTT TTC AAT AGG **AAC** AAA TGC TGC TAA CAG GCG 3'
I736T fwd	5' GC CTG TTA GCA GCA TTT GCT CCT **ACT** GAA AAG CAG 3'
I736T rev	5' CTG CTT TTC **AGT** AGG AGC AAA TGC TGC TAA CAG GC 3'

Sequencing

pcrA in vivo fwd	GAC AAG ACA TCG TTG CGT TCC CTG
pcrA in vivo rev	TGG GTA ACG CCA GGG TTT TCC
rpoB in vivo rev	CTC GCC TAC AGA ACG TAA ACG
pET3a fwd	CGA CTC ACT ATA GGG AGA CCA CAA C
pET3a rev	GGG TTA TGC TAG TTA TTG CT

2.3.1.2 Polymerase Chain Reactions

Polymerase chain reaction (PCR) was used to amplify the DNA in a Biometra T3 Thermocycler (Biometra, Germany). Reactions were performed in 0.5 ml PCR tubes containing:

2 x PCR Master Mix (Fermentas)	12.5 µl
Forward primer (0.1 µg/µl)	1 µl
Reverse primer (0.1 µg/µl)	1 µl
Template DNA	1 µl
Sterile MQW	9.5 µl
Total volume	25 µl

The reaction was started at 95 °C for 2 min, following this, the PCR cycle was:

	Temperature	Time
Denaturing	95 °C	30 s
Annealing	45-55 °C	30 s
Elongation	72 °C	60 s per 1 kilobase (kb)

This cycle was repeated for 35 cycles and then kept a 4 °C.

2.3.1.3 Mutagenic PCR

Site-directed PCR mutagenesis was performed as outlined in Ho *et al.* (1989).

2.3.1.3.1 Designing mutagenic oligonucleotide primers

Overlapping mutagenic primers were designed containing the mutation, representing a single amino acid change (Figure 2.1). The mutagenic forward primer (MFP) and the mutagenic reverse primer (MFP) were complementary (Table 2.3).
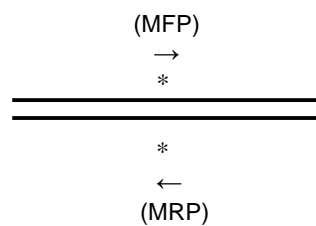
Figure 2.1: The * is the position of the mutation in the gene/gene fragment forward (MFP) and reverse (MRP) primers are designed with the mutation in the middle of the primer.

2.3.1.3.2 PCR reactions

Two separate PCR reactions were performed using the protocol in Section 2.3.1.2. PCR reaction 1 used the forward cloning primer (FWD) and the mutagenic reverse (MRP) and reaction 2 used the reverse cloning primer (REV) and the mutagenic forward primer (MFP). This created two PCR products with an overlapping DNA region, that contained the mutation (Figure 2.2). The two products were then gel purified (Section 2.3.3) and mixed to form the DNA template for the final PCR reaction, using the forward and reverse cloning primers.
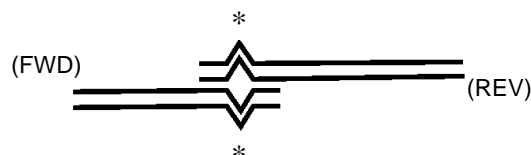
Figure 2.2: Overlapping mutagenic PCR fragments that were used to create the full length DNA containing the mutation. * - mutation, FWD – forward cloning primer, REV – reverse cloning primer.

2.3.2 DNA Agarose Gel Electrophoresis

Agarose gel electrophoresis was performed using the BioRad Mini Sub Cell GT system (Bio-Rad Laboratories). DNA grade agarose (Fermentas) was used to make 1 % (w/v) gels in 40 ml of 1 x TAE buffer (Appendix II), to which 1.5 µl of ethidium bromide (10 mg/ml) was added. The electrophoresis tank was filled with 1 x TAE buffer, until the gel was covered. Samples were mixed with 6 x loading buffer (Appendix II) before loading and the gel was run at 100 V for 20-30 min. DNA bands were visualized using the Kodak transilluminator TR-203. The size of the visualized band was determined using the molecular size marker, GeneRuler 100 base pair plus DNA ladder (Fermentas).

2.3.3 Gel Extraction and Purification

Following electrophoresis DNA bands were visualized under UV light, excised and purified using the Wizard® SV Gel and PCR clean-up system (Promega), using a microcentrifuge and following the manufacturer's instructions.

2.3.4 Restriction Digest

1 µl of each restriction enzyme (10 U/µl) was added, along with 2 µl of the appropriate restriction buffer to a 20 µl reaction. The amount of DNA added to the reaction varied (usually ~ 5-10 µl) and the remaining volume was made up with MilliQ water (MQW). The samples were incubated at 37 °C for 2 h and heat treated at 65 °C for 10 min. All restriction enzymes and buffers were supplied by Fermentas or Promega.

2.3.5 DNA ligation

DNA ligations were performed to covalently join the insert and vector. 1 Weiss unit (1 µl) of T4 DNA ligase (Fermentas) and 2 µl of supplied ligase buffer were added to a 20 µl reaction containing an approximate ratio of 3:1 insert to vector digested DNA. Up to 16 µl of insert could be added to the reaction but the typical volume used was 10 µl, with 2 µl of digested vector and the remaining volume made up with MilliQ water (MQW). Control ligations contained no insert and more MQW. Ligation reactions were incubated at room temperature for 2 h or at 4 °C overnight.

2.3.6 Small-scale Plasmid Preparations from *E. coli*

2.3.6.1 Alkaline lysis method

Single colonies were used to inoculate 2 ml of LB + amp (100 µg/ml) and incubated with shaking for 6-16 h at 37 °C. These cultures were mini-prepped to isolate the plasmid according to the method of Ish-Horowicz and Burke (1981).

2.3.6.2 Promega Kit Prep

Single colonies were used to inoculate 5 ml of LB + amp and incubated with shaking for 6-16 hs at 37 °C. The plasmid was isolated using the Wizard® Plus Minipreps DNA purification system (Promega), and a microfuge, following the manufacturer's instructions.

2.3.7 Plasmid DNA Sequencing

To sequence the plasmid, a kit prep of the plasmid DNA was used. The concentration of the plasmid was determined using either the OD_{260nm} value from the Nanodrop spectrophotometer (Thermo Fisher Scientific) or agarose gel electrophoresis against molecular size standards of known concentration, where dilutions of the plasmids were run on an agarose gel (Section 2.3.2) to determine the concentration required. 600-1500 ng of double stranded plasmid DNA was added to 6.4 pmol of sequencing primer (Table 2.3) in an Eppendorf tube (~12 µl), forward and reverse sequences were sent in separate Eppendorfs. The Eppendorfs were labeled, sealed with parafilm and sent express post to the Australian Genome Research Facility (AGRF) in Brisbane or Sydney (www.agrf.org.au). Once the plasmid was sequenced, the sequence was made available *via* the internet.

2.3.7.1 Sequencing of mutations

To confirm the *pcrA* mutation was correctly integrated into the *B. subtilis* chromosome after transformation, the chromosomal DNA was extracted (Section 2.3.8) and amplified using specially designed primers and the PCR reaction sent for sequencing (400-600 bp PCR product: 12-18 ng DNA sent).

2.3.8 Chloroform extraction of chromosomal DNA from *B. subtilis*

2-5 ml of LB (plus supplements) was inoculated and grown overnight at 37 °C. The cell pellet was collected by centrifugation and resuspended in 500 µl of TES (Appendix II). 25 µl of lysozyme (10 mg/ml stock) and 5 µl of RNase (10 mg/ml stock) were added.

Sample was mixed *via* vortex and incubated at 37 °C for 30 min. 50 µl of pronase (10 mg/ml stock) and 30 µl of 30 % (v/v) Sarkosyl were then added and the sample was mixed and incubated again at 37 °C for 30 min. 600 µl of chloroform was added, mixed and incubated on ice for 30 min. The sample was spun at 17 000 x g for 5 min to separate the layers. The aqueous layer (top) was removed to a clean 1.5 ml tube and the chloroform extraction was repeated. The aqueous sample was removed to a 2 ml tube and the DNA precipitated with 1.5 ml of 100 % ethanol. Sample was inverted several times to mix and spun to pellet the DNA. The supernatant was removed and the pellet washed with 70 % (v/v) ethanol. The pellet was air dried to remove all ethanol and the DNA was dissolved in 100 µl of MQW, heated at 65 °C for 10 min and stored at -20 °C.

2.4 Transformation protocols

2.4.1 Competent *E. coli* cells

Competent *E. coli* cells were prepared using the method outlined by Lederberg & Cohen (1974) and adapted by Kushner (1978).

2.4.1.1 E. coli DH5α

Ligation mixtures were transformed into competent *E. coli* DH5α. The ligation mixture (5-15 µl) was added to 50 µl of thawed competent cells and mixed gently by pipetting up and down. The samples were incubated on ice for 1 h and then heat shocked at 42 °C for two min. 150 µl of LB was added and the samples gently mixed and incubated at 37 °C for 1 h. The samples were spread on NA plates containing ampicillin (100 µg/ml)

and incubated overnight at 37 °C. When plasmid mini-prep DNA was used instead of the ligation mixture, only ~1 µl was required.

2.4.1.2 E. coli BL21 (DE3) pLysS and E. coli ER2566

Transformations using frozen competent *E. coli* BL21 (DE3) pLysS, B834 (DE3), C41 (DE3) and *E. coli* ER2566 cells followed the same procedure as the DH5α transformation (Section 2.4.1.1), except that plasmid mini-prep (1-4 µl) was used as these strains were used for protein overproduction.

2.4.2 *B. subtilis* transformation

Transformation of *B. subtilis* was performed using the method outlined by Kunst and Rapoport (1995). Briefly, *B. subtilis* strains were struck on an NA plate and incubated overnight at 37 °C. These cells were used to inoculate 10 ml of MD media (Appendix II) supplemented with 20 % (w/v) casamino acids (CAA). When the culture reached an OD_{600} between 1.0-1.5 another 10 ml of MD media was added and the culture incubated for 1 h. Then 800 µl of this culture was added to approximately 5 ng of DNA in pre-warmed test tubes and allowed to grow at 37 ºC whilst shaking. After 20 min, 25 µl of 20 % (w/v) CAA was added to each tube, which was incubated for 1 h before being plated onto NA plates containing the appropriate antibiotics and supplements.

2.5 Protein Work

2.5.1 Protein Overproduction

2.5.1.1 Protein induction

Colonies from a transformation (Section 2.4.1.2), using the appropriate plasmid were resuspended from plates in ~1 ml of LB + ampicillin (100 µg/ml) and used to inoculate 20 ml of LB + amp. The cultures were grown at 37 °C shaking until the OD_{600} reached ~ 0.6 - 0.7. This became T_0, when the exact OD_{600} was recorded and 1 ml taken, pelleted and frozen at -20 °C. Overproduction of the protein was then induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of between 0.05 mM to 0.5 mM. The culture was then grown with shaking, at 37 °C for a further 3 h, or at room temperature for 5 h. At T = 3/5 another OD_{600} reading was recorded and a 1 ml sample pelleted and frozen. The remaining culture was centrifuged at 5000 x g at 4 °C for 15 min and the pellet stored at -80 °C. Following the determination of the optimal conditions required for protein overproduction (Section 2.5.1.2) the induction was scaled up to 1 L and the culture pelleted at 8000 x g at 4 °C for 15 min and stored at -80 °C until protein purification was carried out.

2.5.1.2 Protein trial induction and solubility check

The 1 ml T_0 and $T_{3/5}$ samples were resuspended in TES (Appendix II) and lysed to release the protein before the induction results were checked using SDS-PAGE. The pellet was resuspended to an OD_{600} of 10, 0.8 of this volume was in 1 x TES and 0.2 was lysozyme (10 mg/ml in TES). The samples were incubated at 37 °C for 5-10 min and then sonicated 4 times of 12 short bursts with 1 min ice incubations in between. A

solubility check was performed on the $T_{3/5}$ sample. Firstly, 15 µl was removed as the whole cell sample (WCS). The remaining volume was spun in a microcentrifuge. The supernatant was the soluble fraction, which was removed and 15 µl kept along side the WCS and T_0 samples. The insoluble pellet was then washed three times in TES before being resuspended in the original volume of TES, and a 15 µl sample removed. All samples (T_0, WCS, sol, insol) were run on an SDS-PAGE gel (Section 2.5.2).

2.5.2 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

2.5.2.1 Gel cassette preparation & gel casting

To analyse protein samples SDS-PAGE gels were run using a Mini PROTEAN® 3 system, which was assembled as per manufacturer's instructions (Bio-Rad Laboratories). SDS-PAGE gels were made with a 10 % (w/v) resolving gel and a 6 % (w/v) stacking gel (Appendix II).

2.5.2.2 Module assembly and sample loading

The gel was removed from the casting frame and the electrophoresis module assembled as per instructions. Samples were added to equal amounts of SDS 2 x loading buffer (Appendix II) and denatured by heating at 97 °C for 2 min. Samples were loaded in wells using fine gel tips (~ 20 µl), 5 µl of PageRuler™ prestain protein ladder (Fermentas) was also loaded. Gels were then run at 200 V for 27 min, or until the marker was sufficiently separated.

2.5.2.3 Gel removal & staining

The gel was removed from the cassette and placed in Coomassie-blue stain (Appendix II). The gel was gently shaken while being stained for 1 h. The gel was then washed in distilled water and transferred to de-stain (Appendix II) overnight. The gels were then scanned.

2.5.3 Western blot

Protein samples were run on an SDS-PAGE gel (Section 2.5.2, without staining). The gel was then transferred to a nitrocellulose membrane using a Mini Trans-Blot Electrophoretic Transfer Cell System (Bio-Rad Laboratories). The gel transfer sandwich was assembled as per the manufacturer's instructions and run at 30 mA overnight.

2.5.3.1 Western blot

The membrane was removed and washed in ~10 ml of 1 x PBS three times for 5 min each time. The membrane was then blocked for 1 h in blocking solution (Appendix II) at room temperature on a shaking platform. The membrane was then washed three times in wash buffer (Appendix II) before a dilution of the primary antibody (Table 2.4) in blocking solution was added for 1 h, at room temperature, on a shaking platform. The membrane was rinsed three times and then washed three times for 5 min in wash buffer. A 1:3000 dilution of secondary antibody (Goat anti-Rabbit horse-radish peroxidase or Goat anti-Mouse horse-radish peroxidase, Bio-Rad) in blocking solution was added, for 1 h at room temperature on a shaking platform. After this the membrane was again subjected to three rinses and three 5 min washes in wash buffer.

The Opti-4CN^{TM} Substrate Kit (Bio-Rad Laboratories) was used, according to the manufacturer's instructions to detect the presence of antibody on the membrane. After detection the membrane was washed in MQW and scanned.

2.5.3.2 Far-Western blot

The far-Western blot was performed mostly as a normal Western blot. The gel was run (Section 2.5.2) and transferred overnight at 30 mA (Section 2.5.3). The membrane was removed and rinsed in 1 x PBS three times. The membrane was then washed in Tris far-Western buffer (TFWB, Appendix II) three times for 5 min each time. The membrane was then blocked for 1 h in refold buffer (TFWB + 5 % (w/v) skim milk, Appendix II) at room temperature on a shaking platform, which refolded the proteins on the membrane. The membrane was then probed with the protein of interest for at least 1 h. The membrane was rinsed and washed three times each in refold buffer before the addition of the primary antibody, which detected the binding of the probe protein. The antibody addition and detection was performed as a normal Western blot (Section 2.5.3.1).

2.5.3.3 Antibody dilutions used in the Western and far-Western blot

Table 2.4: Antibodies used in the Western and far-Western blots. The animal the antibody in was raised in and the dilution factor used

Antibody	Animal raised in	Dilution used
Anti-PcrA	Rabbit: polyclonal	1:5000
Anti-RpoB (β)	Rabbit: polyclonal	1:5000
Anti-RpoC (β')	Rabbit: polyclonal	1:5000
Anti-GreA	Rabbit: polyclonal	1:5000
Anti-His	Mouse: monoclonal	1:3000
Anti-GST	Rabbit: polyclonal	1:3000
Anti-SrepII	Rabbit: polyclonal	1:3000
Anti-rabbit-HRP	Goat	1:3000
Anti-mouse-HRP	Goat	1:3000

2.5.4 Protein purification

2.5.4.1 Purification of PcrA

2.5.4.1.1 Lysis

The overproduction pellet from 2 L of culture was thawed and resuspended in 20 ml of lysis buffer (Appendix II) and 5 ml of lysozyme (10 mg/ml in lysis buffer) was added. The sample was incubated at 37 °C until it became highly viscous, which indicated the cells were lysed. The sample was sonicated 10 x 1 min intervals, on ice, with 1 min ice incubations between each and centrifuged at 10 000 x g at 4 °C for 1 h to pellet the cell debris and insoluble protein. The pellet was discarded and the supernatant was used in the Polymin P precipitation.

2.5.4.1.2 Polymin P precipitation

The supernatant from the lysate was placed into a beaker at 4 °C and 2 ml of 10 % (v/v) Polymin P (Polyethylenimine) was added. The solution was left stirring for 10 min until the solution turned milky. The sample was centrifuged at 5000 x g at 4 °C for 10 min. The supernatant was discarded and the pellet resuspended in 25 ml of Polymin P wash buffer (Appendix II) and centrifuged again at 5000 x g for 10 min. The supernatant, which contained PcrA, was kept for the ammonium sulfate precipitation.

2.5.4.1.3 Ammonium sulphate precipitation

The supernatant from the Polymin P precipitation was placed in a flat bottom container on a stirrer at 4 °C. Ammonium sulphate ($(NH_4)_2SO_4$) which had been finely ground with a mortar and pestle was added incrementally, allowing each addition to completely dissolve to avoid localised areas of higher $(NH_4)_2SO_4$ concentration. PcrA was precipitated out of solution at a $(NH_4)_2SO_4$ saturation of 35 %, which is equal to 5.25 g per 25 ml of sample. Once all the $(NH_4)_2SO_4$ was dissolved, the sample was left stirring at 4 °C for 30 - 60 min. The sample was then centrifuged at 4 °C for 1 h at 14 000 x g. The supernatant was discarded and the pellet was stored at -20 °C until the next purification step.

2.5.4.1.4 Phenyl Sepharose – hydrophobicity chromatography

The ammonium sulfate precipitate was thawed and resuspended in 5 ml of buffer A with 0.2 M NaCl and after resuspension 2.1 ml of buffer B was added (to equal ~ 30 % $(NH_4)_2SO_4$ concentration, Appendix II). The sample was then centrifuged at 8000 x g for 1 h at 4 °C to remove particulate matter.

The phenyl Sepharose resin (GE Healthcare, Australia) was prepared as per manufacturers instructions, 5 ml of settled resin was poured in a econo-column (1.0 x 10 cm, 8 ml) and fitted with a flow adapter (Bio-Rad Laboratories). The column was equilibrated in starting buffer (30 % buffer B, Appendix II). All chromatography was carried out at 4 °C.

The sample was loaded onto the column *via* a peristaltic pump with a flow rate of 0.5 ml/min and the column was then washed with 2 column volumes (CV) of starting buffer. The gradient required to elute PcrA from the column was decreasing $(NH_4)_2SO_4$ concentration. The gradient was run over 3 CV, which was set up in a gradient mixer. The high to low salt gradient was applied to the column *via* the pump. The flow through and wash fractions were collected in 10 ml tubes. Once the gradient was applied at a flow rate of 0.5 ml/min, 50 x 0.5 ml fractions were collected using a fraction collector (Bio-Rad Laboratories). All fractions were run on SDS-PAGE gels and the fractions containing PcrA were pooled and applied to the heparin column.

2.5.4.1.5 Heparin affinity chromatography

The heparin affinity column was prepared as per manufacturers instructions; a 3.5 ml heparin-agarose column (Bio-Rad Laboratories) was used. The column was equilibrated in buffer A + 100 mM NaCl (Appendix II). The sample was then applied *via* the peristaltic pump with a flow rate of 0.5 ml/min and the column washed in 2 CV of buffer A. The gradient was then applied to the column, again using the gradient mixer. The low salt buffer A (0.25 M NaCl, 50 mM Tris, pH 8.0) to high salt buffer B (1.25 M NaCl, 50 mM Tris, pH 8.0) gradient was applied over 3 CV. The flow rate was 0.5 ml/min, the flow through and wash fractions were collected in 10 ml tubes and the

0.5 ml fractions from the gradient were collected in a fraction collector. The flow through, wash and elutions were run on SDS-PAGE gel and the fractions containing purified PcrA were dialysed overnight.

2.5.4.1.6 Dialysis

The elutions which contained the purified PcrA were pooled and dialysed overnight into dialysis buffer (Appendix II) at 4 °C. The dialysate was then aliquoted and stored at -80 °C.

2.5.4.2 Nickel affinity purification of His-tagged proteins

The cell pellet was resuspended in 4 ml/g wet weight of cell pellet in lysis buffer (Appendix II) + 0.01 % (v/v) Triton X-100 and 2 mg/ml lysozyme and incubated at 37 °C until the sample became highly viscous (~20 min), which indicated the cells were lysed. The sample was sonicated for 1 min intervals, on ice, with 1 min ice incubations between each and centrifuged at 10 000 x g at 4 °C for 1 h to pellet the cell debris and insoluble protein. Sample and buffers where passed through a 0.45 µm filter before purification. Purification was performed using 1 ml His-Trap HP columns (GE Healthcare) according to the manufacturer's instructions on an ÄKTA FPLC (GE Healthcare). The sample was loaded onto the column and the column was washed in 10 column volumes (CV) of wash buffer (45 mM imidazole). His-tagged proteins were then eluted from the column with 10 column volumes of elution buffer (500 mM imidazole, Appendix II). After purification the elutions were run on an SDS-PAGE gel to check purity, pooled and dialysed.

2.5.4.2.1 On-column refolding and nickel affinity purification of insoluble His-tagged

proteins

Cell pellets were resuspended in lysis buffer and lysed as in Section 2.5.4.2. After

centrifugation the soluble supernatant was removed and the insoluble pellet was then

resuspended in binding buffer containing 8 M urea (Appendix II) and rotated overnight

at 4 °C. The next morning the sample was centrifuged at 10 000 x g for 10 min to

remove any non solubilised protein then loaded onto the 1 ml His-Trap HP column (GE

Healthcare). Urea was removed over 10 CV by a gradient using binding buffer.

Purification was then performed as in Section 2.5.4.2 above.

2.5.4.2.2 Nickel affinity purification of insoluble His-tagged proteins

Purification of insoluble protein could also be carried out in 8 M urea. Purification

occurs as in Section 2.5.4.2.1, except the wash buffer and elution buffer also contained

8 M urea. The elutions containing the insoluble purified protein were pooled and

dialysed to refold the protein and remove the urea, in a step-wise fashion (8 M, 6 M, 4

M, 2 M 0 M urea) until no urea was present. Final sample was centrifuged to remove

any remaining insoluble protein and glycerol was added to the soluble protein to a final

concentration of 30 % (v/v) glycerol and stored at -80 °C.

2.5.5 Determination of protein concentration

The concentration of purified proteins was determined by measuring the absorbance at

280 nm (OD_{280}). The OD_{280} was used with the molar extinction coefficient ($M^{-1}cm^{-1}$) to

calculate the protein concentration. This concentration was then converted into µM

and/or mg/ml. To calculate the concentration in mg/ml the concentration (M) was

multiplied by the molecular weight (Da). The molar extinction coefficient was calculated on the ExPASy (**Ex**pert **P**rotein **A**nalysis **Sy**stem) proteomics server (www.expasy.org) using the amino acid sequence of the protein.

2.6 Affinity chromatography

2.6.1 Preparing the Affi-Gel column

The Affi-gel 15 resin (Bio-Rad laboratories) was mixed and 50-70 µl of gel slurry was placed into an Eppendorf tube and allowed to settle. The supernatant was removed and the resin washed with cold MQW (~200 µl) three times. The protein that was to be coupled to the Affi-gel resin was added and the sample was placed on a rotor at 4 °C for a minimum of 2 h. After 2 h, 50 µl of 1 M ethanolamine, pH 8 was added to block any free active esters remaining in the resin and the sample was again placed on the rotor at 4 °C for a minimum of 1 h. The resin was then allowed to settle and most of the supernatant removed. The remaining supernatant and resin was applied to the empty column, which was a gel loading pipette tip packed with glass wool. Wash buffer (50 mM phosphate buffer, pH 7.5, Appendix II) was added to settle the resin and sodium azide (0.2 % (v/v)) was applied and columns were stored at 4 °C. For the blank resin column 50 µl of ethanolamine was added directly to the washed resin and placed on the rotor for ~2 h at 4 °C.

2.6.2 Running the Affi-Gel column

The column was washed with at least 10 column volumes (CV) of wash buffer (Appendix II) before use to remove all the sodium azide. The protein sample which was applied to the column had been previously dialysed into wash buffer overnight at 4 °C. The protein sample was applied to the column, at room temperature. The flow through was collected and re-applied to the column continuously for 1 h to allow the interaction between the resin-coupled protein and protein sample to reach equilibrium. After 1 h the flow through was collected and the column was washed with 10 CV of wash buffer. NaCl (400 mM) was then applied to break the protein interaction and elute the sample protein, 6 x 50 µl elutions were collected. A higher concentration of NaCl (1 M) was then applied, to break stronger protein-protein interactions between the pair and 9 x 50 µl elutions were collected. The elutions were run on an SDS-PAGE gel (Section 2.5.2), with a positive control protein sample, the flow through and the wash. The gel was transferred to the Western blot membrane overnight and blotting was performed the following morning (Section 2.5.3.1). During some affinity chromatography experiments the elutions were pooled and acetone precipitated (Section 2.10.4) before loading onto the SDS-PAGE gel.

2.6.3 Cleaning the Affi-gel column

After the elutions were collected an additional 5-10 CV of 1 M NaCl was applied to the column to ensure any residual protein had been removed. The column was then re-equilibrated in wash buffer and stored in 0.2 % (v/v) sodium azide at 4 °C.

2.7 Growth curves

2.7.1 *B. subtilis* growth curves

The growth curves of the PcrA mutant *B. subtilis* strains were performed and compared to the wild type. Initially LB medium was inoculated with the strains previously struck onto solid media and grown overnight. This pre-culture was used to inoculate new LB medium (10 ml) to an OD_{600} of 0.05. The culture was incubated with shaking at 37 °C. At regular intervals (every 30-60 min) the OD_{600} was measured using the SmartSpec (Bio-rad) throughout the day.

2.7.2 Growth curves using the plate reader

Growth curves of multiple strains could be performed automatically in the FLUOstar Optima plate reader (BMG Labtech). The same pre-culture was used and diluted back to OD_{600} 0.05 in 6 replicates in a 96 well plate (Figure 2.3). The plate was incubated with shaking at 37 °C in the plate reader and the OD_{600} of each well was measured every 10 min. The experiment was normally run overnight and the following morning the data was exported to an Excel spreadsheet for data analysis.
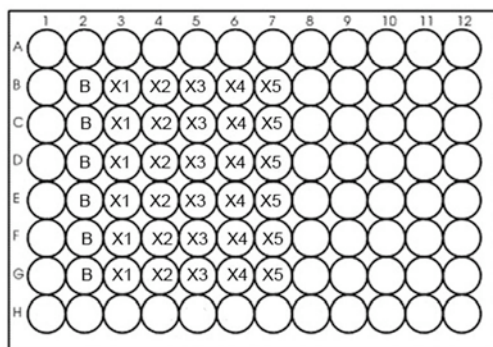
Figure 2.3: The layout of the 96 well plate used in the growth curves. B – blank, X1- 6 replicates of first strain, X2 – 6 replicates of second strain, X3 – 6 replicates of third strain, X4 – 6 replicates of fourth strain and X5 0 6 replicates of the fifth strain.

2.8 Microscopy

2.8.1 Liquid cultures for microscopy

Overnight cultures of *B. subtilis* strains were inoculated using a single colony from a freshly streaked plate and grown in LB supplemented with 5 µg/ml chloramphenicol and 0.5 % (w/v) xylose. For microscopy on mid-exponential phase cells, the overnight cultures were diluted back to an OD_{600} of 0.05 and grown to an OD_{600} of approximately 0.3-0.5. To visualise nucleoids 5 µl of DAPI (4',6-diamidino-2-phenylindole; ~30 ng/ml stock) was added to 5 µl of culture and immediately placed on the slide, on a 1.2 % (w/v) agarose pad.

2.8.2 Image processing and analysis

Images were acquired using a a Ziess Axioskop 2 epifluorescence microscope equipped with a Hamamatsu Orca-ER cooled charge-coupled camera. Exposure time for phase contrast images was 0.1 s while exposure time for DAPI images was 0.5 s. Image acquisition and cell length measurements were performed using MetaMorph v 7.1.3.0 (Universal Imaging Corps., USA). Cell length data was then exported to Excel for manipulation and graphical presentations.

2.9 Tandem affinity purification

2.9.1 Construction of strains

The *B. subtilis* strains containing the TAP-tagged *rpoB* and the *pcrA* mutations were constructed by first obtaining the chromosomal DNA of the TAP-tagged *rpoB* strain (Section 2.3.8). This chromosomal DNA was then transformed into the *B. subtilis* strains already containing the *pcrA* mutations (Section 2.4.2). The resultant transformants were checked using PCR and 0.5 mM IPTG/1 µg/ml erythromycin selection. The correct strains (Table 2.1) were grown in 500 ml of LB containing 5 µg/ml choloramphenicol, 0.5 % (w/v) xylose, 1 µg/ml erythromycin and 0.5 mM IPTG until they reached an OD_{600} of 0.6. Cells were harvested by centrifugation at 6000 x g for 5 min, washed in 10 ml of buffer A (Appendix II) and pelleted at 5200 x g for 7 min. Cell pellets were stored at -80 °C.

2.9.2 FLAG affinity purification

Cell pellets were resuspended in 2.5 ml Buffer B (Appendix II) per gram of cell weight, supplemented with 0.5 mg/ml lysozyme, 0.1 mg/ml RNase and 0.1 mg/ml DNase. The cell pellet was thoroughly resuspended and incubated at 37 °C for 15 min. The sample was then incubated on ice for a further 30 min. The sample was centrifuged for 30 min at 14500 x g.

Anti-FLAG M2-agarose (200 µl, Sigma-Aldrich) was equilibrated in 15 ml of cold buffer B twice. The resin was centrifuged for 1 min at 1000 x g and the buffer was

discarded. The cell extract supernatant was applied to the resin and the sample was rotated at 4 °C for 90 min. The resin was again centrifuged and the supernatant removed. The resin was transferred to a Bio-spin column (Bio-rad) and washed with 500 µl of cold buffer B twice. The resin was resuspended in 200 µl of TEV buffer (Appendix II) and 1 mg/ml TEV protease was added. The sample was rotated overnight at 4 °C.

2.9.3 Calmodulin-binding peptide affinity purification

Calmodulin-Sepharose 4B (200 µl, GE) was equilibrated with 15 ml of CBB buffer (Appendix II) twice and the resin transferred to a second Bio-spin column. The elution from the anti-FLAG resin was applied directly into the Calmodulin-Sepharose column. The anti-FLAG resin was then washed with 400 µl of CBB buffer and again applied directly onto the Calmodulin-Sepharose column. The Ca^{2+} concentration in the column was adjusted to 2 mM by the addition of 2 µl of 200 mM $CaCl_2$. The sample was then rotated at 4 °C for 2 h. The flow through was collected and the Calmodulin-Sepharose resin washed twice in 500 µl of cold CBB buffer. The resin was next washed with 200 µl of CWB (Appendix II) and finally samples were eluted in 2 x 500 µl of CEB (Appendix II).

2.9.4 Acetone precipitation

The protein elutions from the Calmodulin-Sepharose column were precipitated using the addition of 3 x volume of ice cold (-20 °C) acetone (1.5 ml). Samples were vortexed to mix and left at -20 °C overnight. Samples were centrifuged at 17 000 x g for 1 h at 4 °C

and the supernatant discarded, without disturbing the protein pellet. The samples were left to air dry and the pellet resolubilised in 20 µl of SDS loading buffer and heated at 65 °C for 15 min. The samples were run on a SDS-PAGE gel and transferred to nitrocellulose membrane for Western blotting (Sections 2.5.2 & 2.5.3).

<u>2.10.5 Overproduction and Nickel affinity purification of the TEV protease</u>

His-tagged TEV protease was overproduced by the transformation of pNG533 (Table 2.2) into *E. coli* BL21 (DE3) pLysS (Section 2.4.1.2). Protein induction occurred in auto-induction medium (Appendix II), grown for 24 h at room temperature (~22 °C). The his-tagged TEV was insoluble when overproduced in LB with the addition of 0.5 mM IPTG for 3 h at 37 °C. The soluble his-tagged TEV overproduced in auto-induction medium was purified using a HisTrap column (Section 2.5.4.2).

2.10 Single particle reconstruction of the PcrA-RNAP complex

<u>2.10.1 Image acquisition</u>

The negative stain dataset was created by forming the PcrA-RNAP complex *in vitro*. 1 µM RNAP purified from *B. subtilis* cells (BS200, Table 2.1) was mixed with purified PcrA (Section 2.5.4.1) at an approximate 1:1 molar ratio in complex buffer (Appendix II) at room temperature for 10 min, followed by 10 min on ice. The sample was diluted to 0.8 µM in ice cold complex buffer and 4 µl of the sample was applied to home made continuous carbon grids (X. Yang and R. Rothnagel, Institute for Molecular

Biosciences, Queensland) for 30 s before washing three times with complex buffer. The specimens were then stained with a 1 % (w/v) uranyl formate (UF) aqueous solution at room temperature for 30 s and imaged on a Tecnai F30 FEG TEM (field emission gun transmission electron microscope; FEI Company) operated at 300 keV at a magnification of 59,000 x at room temperature. Images were taking under low dose settings (~60 e-/Å^2) at a defocus of 600-900 nm and recorded on a GATAN 4k x 4k CCD camera (Gatan Inc.) with 2x binning resulting in a pixel size corresponding to 3.9 Å at the specimen level.

2.10.2 Particle picking

Particle picking was performed manually using *boxer* software within EMAN package v1.8 (Ludtke *et al.* 1999), using a 108 x 108 pixel box size. The data set was then manually pruned three times to remove the 'bad' picked particles, identified *via* visual assessment. The EMAN software package v1.8 (Ludkte *et al.* 1999) was used for image pre-processing and 3D reconstructions. Individual particles were appended (*proc2d*), center aligned (*centalignint*), and the box size reduced to 72 x 72-pixel images (*proc2d clip=72,72*).

2.10.3 Reconstructions

The refinement loop and initial models used to create the PcrA-RNAP complex model are listed in Appendix III.

Chapter 3:

PcrA and RNA polymerase

in vitro interaction studies

Chapter 3: PcrA and RNA polymerase *in vitro* interaction studies

3.1 Introduction

This chapter will cover the characterisation of the protein-protein interaction between PcrA and RNAP. The interaction was initially identified in a yeast two-hybrid screen performed investigating the replisome (Noirot-Gros *et al*, 2002a). It was confirmed in another yeast two-hybrid screen, which aimed to identify new transcription factors (P. Lewis, unpublished). The results indicated there was an interaction occurring between PcrA and the N-terminal region of the RNAP β subunit.

3.1.1 Yeast two-hybrid assay

The yeast two-hybrid assay is one method that can be used to identify protein-protein interactions, and genomic yeast two-hybrid screening can be used to identify novel protein-protein interactions. Yeast two-hybrid transcription separates the DNA binding domain from the activation domain of a transcription factor (Figure 3.1A), and an interaction between a second pair of proteins is required for the transcription of the reporter (Serebriiskii *et al*, 1999). The target protein is known as the 'bait' and is fused to a DNA binding domain (BD). Transcription is only activated when a second protein, the 'prey', which is fused to a transcription activator domain (AD), interacts with the bait (Figure 3.1B) (Serebriiskii *et al*, 1999). This is exploited in the genomic Y2H screen, where the 'bait' is exposed to many yeast plasmids separately, each containing a different protein fused to the activator domain.
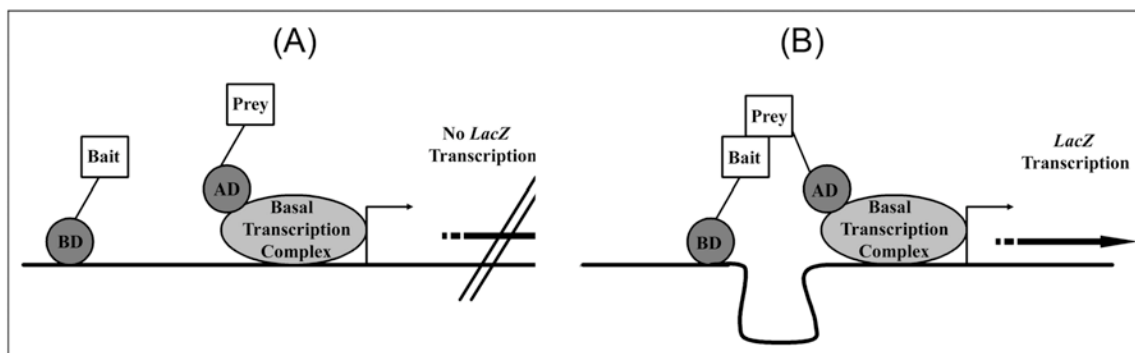
Figure 3.1: Yeast two-hybrid system. (A) Two-hybrid activation, where two proteins - bait and prey - have been fused to the binding and activation domains. If the two proteins do not interact, transcription does not occur. (B) The interaction between the bait and prey proteins leads to transcription of the reporter gene (*lacZ*).

A yeast two-hybrid screen was designed to identify novel protein-protein interactions occurring with RNAP, to discover novel transcription factors. The β and β' subunits, which form the catalytic core of RNAP were used as the bait proteins. To further specify the regions of interaction, the β and β' subunits were subdivided into smaller fragments and cloned as additional bait proteins. The bait proteins were then screened against a *B. subtilis* genomic library. The β subunit C-terminal terminal domain (CTD) fragments and the β' subunit bait proteins did not return many hits during the yeast two-hybrid screening and the most successful bait protein was the NTD (1-608 aa) of the β subunit, which received ~950 hits (P. Lewis, unpublished). This included 23 class A hits and 59 class C hits, including PcrA. Class A interactions are those which identify two or more overlapping clones which have differing fusion points in the library, while class C interactions are repeated clones with the same fusion point. The class A interaction between PcrA and RNAP was identified by a total of 16 hits using both the full length β and $\beta_{1-608aa}$ bait proteins (P. Lewis, unpublished). This result was supported in a previous yeast two-hybrid screen using PcrA as the bait protein (Noirot-Gros *et al*, 2002a). Detection of bi-directional interactions using the yeast two-hybrid is

rare (when both proteins pull out their interacting partner in reciprocal genomic screens), and is usually indicative of a strong interaction, commonly found in molecular machines.

3.1.2 Protein interaction studies

To further characterise this interaction, *in vitro* binding studies using purified PcrA and purified RNAP were performed. Affinity chromatography was initially used to confirm the interaction between RNAP and *B. subtilis* PcrA (PcrA$_{Bsub}$) (Harriott, 2006) and *G. stearothermophilus* PcrA (PcrA$_{Gst}$). Additionally, *in vitro* interaction studies using far-Western blots were performed to indentify specific regions involved in the interaction.

3.1.2.1 Affinity chromatography

Affinity chromatography involves the coupling of one protein to an affinity medium. The second target protein is then applied to the affinity column. If the two proteins interact the sample will remain bound to the protein in the column. Proteins that do not interact will flow through the column (Figure 3.2). The interaction between the coupled protein and the second protein must then be disrupted, most commonly by increasing the NaCl concentration.
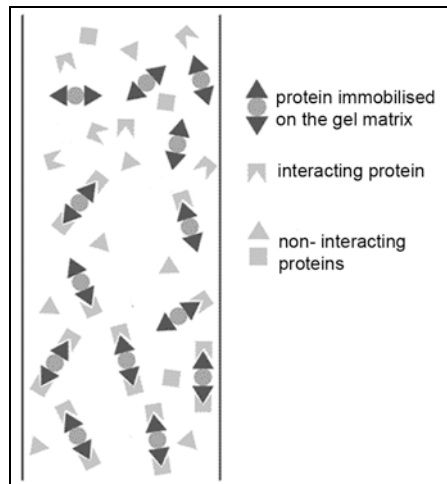
Figure 3.2: Affinity chromatography schematic. The first target protein sample (▲) is coupled to the chosen matrix (●). The second protein sample (⊓) is then applied to the affinity column. If the proteins interact it remains bound, if they don't it is passed through the column and collected as the flow through (adapted from Hiller-Sturmehofel *et al*, 2008).

This method can be used to achieve reciprocal results, as both target proteins can be bound in separate affinity columns. This allows for the appropriate control experiments to be performed and ensures that any *in vitro* interactions detected are specifically occurring between the two target proteins.

3.1.2.1 Far-Western blot

Far-Western blotting is an *in vitro* interaction assay that can be used to further define regions or even amino acids directly involved in protein-protein interactions (for example see Johnston *et al*, 2009). This was required in this study as RNAP is a large multiple subunit enzyme. Purified full length RNAP was used in the affinity chromatography and it was important to determine the position of PcrA binding on RNAP. This would allow for the hypothesis of possible roles PcrA could have during transcription, as a result of the interaction. Smaller RNAP fragments had been shown to be insoluble when overproduced. Attempts to refold and purify, while sometimes successful, showed the fragments were unstable (Doherty, 2007). The use of far-Western blots was employed to make use of the insoluble RNAP fragments. Far-

Western blotting uses a nitrocellulose membrane as the matrix to which the prey protein is bound. Immediate refolding of the proteins on the membrane was followed by the addition of the probe protein (bait). The membrane was then washed and detected using the antibody of the probe protein (Figure 3.3).
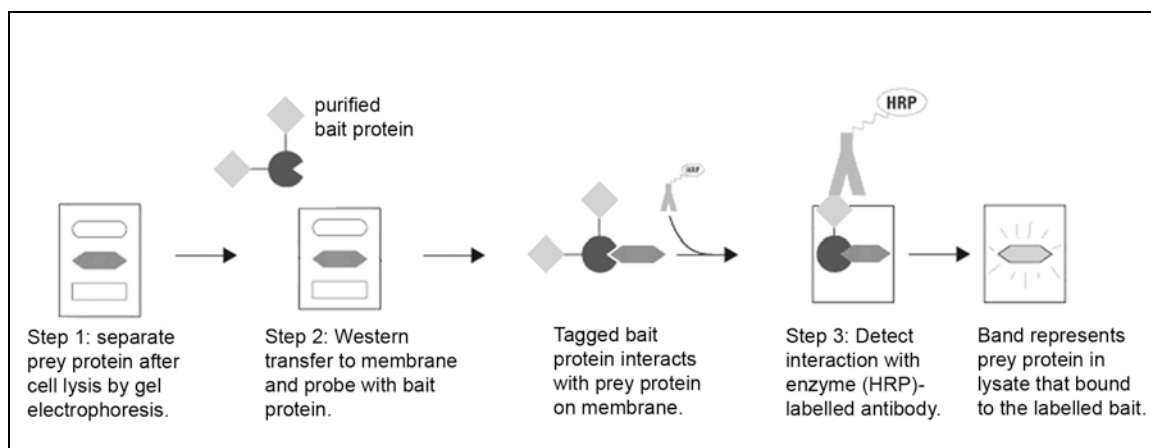
Figure 3.3: Far-Western blot. The proteins (prey) are separated using SDS-PAGE and transferred to nitrocellulose, the membrane is probed with the second protein of interest (bait) and if an interaction occurs between the two proteins then following detection a band is present which represents the bait protein bound to prey protein (adapted from Thermo Scientific Pierce, Protein Interaction Technical Handbook).

There were a number of benefits to using far-Western blotting to investigate the sites of interaction between PcrA and RNAP, including not having to purify the overproduced proteins. Also, the immediate refolding and probing of the prey proteins enabled the use of proteins fragments that may be unstable. The use of the overlapping fragments of the β and β' subunits of RNAP also allowed for more specific regions of protein interaction to be identified.

3.1.3 PcrA and UvrD

When investigating the interaction between PcrA and RNAP, PcrA from both *B. subtilis* (PcrA$_{Bsub}$) and *Geobacillus stearothermophilus* (PcrA$_{Gst}$) were used. PcrA from these two bacterial species share a high degree of sequence similarity and structural homology

and therefore either may be used when investigating the interaction. The identity and similarity between PcrA$_{Bsub}$, PcrA$_{Gst}$ and UvrD is shown in Table 3.1. The sequence ID between the two PcrA sequences was 70 % and the similarity was 84 %. When PcrA$_{Bsub}$ was compared to UvrD the identity was 41 % and similarity was 61 %. The identity and similarity between PcrA$_{Gst}$ and UvrD was slightly higher, with an identity of 43% and a similarity of 63% (Appendix IV).

Table 3.1: The sequence identity and similarity of PcrA$_{Gst}$ and UvrD proteins, when compared to PcrA$_{Bsub}$.

Protein	Identity (%)	Similarity (%)
PcrA$_{Gst}$	70	84
UvrD	41	61

The interaction between RNAP and another helicase, UvrD was also examined. UvrD is an SF1 helicase from the Gram-negative bacterium *E. coli*. There are two SFI helicases in *E. coli* – UvrD and Rep, the roles of which are thought to be encompassed by PcrA in Gram-positive bacteria. UvrD was chosen instead of Rep for a few reasons, such as the abundance of UvrD in *E. coli,* approximately 3000-5000 molecules/cell (Arthur & Eastlake, 1983; Petit *et al*, 1998). PcrA is abundant, 8000 molecules/cell in Gram-positive bacteria (Petit *et al*, 1998), while Rep is present at much lower levels, only 50 molecules/cell (Kornberg *et al*, 1978). Additionally Rep has roles in the replication of bacteriophages and viruses (McGlynn, 2011) and has also been implicated in PriC-directed replication restart (Heller & Marians, 2005). PcrA and UvrD have been shown to play no major role in chromosomal replication.

```
Bsub     1  MNYISNQLLSGLNPVQQEAVKTTDGPLLLMAGAGSGKTRVLTHRIAYLMAEKHVAPWNIL
Gst      1  MNFLSEQLLAHLNKEQQEAVRTTEGPLLLMAGAGSGKTRVLTHRIAYLMAEKHVAPWNIL
UvrD     1  --MDVSYLLDSLNDKQREAVAAPRSNLLVLAGAGSGKTRVLVHRIAWLMSVENCSPYSIM

Bsub    61  AITFTNKAAREMKERVESLLGPGADDIWISTFHSMCVRILRRDIDRIGINRNFSILDTAD
Gst     61  AITFTNKAAREMRERVQSLLGGAAEDVWISTFHSMCVRILRRDIDRIGINRNFSILDPTD
UvrD    59  AVTFTNKAAEMRHRIGQLMGTSQGGMWVGTFHGLAHRLLRAHHMDANLPQDFQILDSED

Bsub   121  QLSVIKGILKERNLDPKKFDPRSILGTISSAKNELTEPEEFSKVAGGYYDQVVSDVYADY
Gst    121  QLSVMKTILKEKNIDPKKFEPRTILGTISAAKNELLPPEQFAKRASTYYEKVVSDVYQEY
UvrD   119  QLRLLKRLIKAMNLDEKQWPPRQAMWYINSQKDEGLRPHHIQSYG-NPVEQTWQKVYQAY

Bsub   181  QKKLLKNQSLDFDDLIMTTIKLFDRVPEVLEFYQRKFQYIHVDEYQDTNRAQYMLVKQLA
Gst    181  QQRLLRNHSLDFDDLIMTTIQLFDRVPDVLHYYQYKFQYIHIDEYQDTNRAQYTLVKKLA
UvrD   178  QEACDRAGLVDFAELLLRAHELWLNKPHILQHYRERFTNILVDEFQDTNNIQYAWIRLLA

Bsub   241  ERFQNLCVVGDSDQSIYRWRGADITNILSFEKDYPNASVILLEQNYRSTKRILRAANEVI
Gst    241  ERFQNLCAVGDADQSIYRWRGADIQNILSFERDYPNAKVILLEQNYRSTKRILQAANEVI
UvrD   238  GDTGKVMIVGDDDQSIYGWRGAQVENIQRFLNDFPGAETIRLEQNYRSTSNILSAANALI

Bsub   301  KNNSNRKPKNLWTENDEGIKISYYRGDNEFGEGQFVAGKIHQLHSTGKRKLSDIAILYRT
Gst    301  EHNVNRKPKRIWTENPEGKPILYYEAMNEADEAQFVAGRIREAVERGERRYRDFAVLYRT
UvrD   298  ENNNGRLGKKLWTDGADGEPISLYCAFNELDEARFVVNRIKTWQDNGG-ALAECAILYRS

Bsub   361  NAQSRVIEETLLKAGINYNIVGGTKFYDRKEIKDILAYLRLVSNPDDDISFTRIVNVPKR
Gst    361  NAQSRVMEEMLLKANIPYQIVGGLKFYDRKEIKDILAYLRVIANPDDDLSLLRIINVPKR
UvrD   357  NAQSRVIEEALLQASVPYRIYGGMRFFERQEIKDALSYLRLIANRNDDAAFERVVNTPTR

Bsub   421  GVGATSLEKIASYAAINGLSFFQAIQQVDFIG-VSAKAANALDSFRQMIENLTNMQDYLS
Gst    421  GIGASTIDKLVRYAADHELSLFEALGELEMIG-LGAKAAGALAAFRSQLEQWTQLQEYVS
UvrD   417  GIGDRTLDVVRQTSRDRQLTLWQACREILQEKALAGRAASALQRFMELIDALAQETADMP

Bsub   480  ITELTEEILDKTEYREMLKAEKSIEAQSRLENIDEFLSVTKNFEQKSEDKILVAFLTDLA
Gst    480  VTELVEEVLDKSGYREMLKAERTIEAQSRLENLDEFLSVTKHFENVSDDKSLIAFLTDLA
UvrD   477  LHVQTDRVIKDSGLRTMYEQEKGEKGQTRIENLEELVTATRQFSYNEEDE---DLMPLQA

Bsub   540  LIADIDQLDQKEEESGGKDAITLMTLHAAKGLEFPVVFLMGLEEGVFPHSRSLMEEAEME
Gst    540  LISDIDELDGTEQAAEG-DAVMLMTLHAAKGLEFPVVFLIGMEEGIFPHNRSLEDDDEME
UvrD   534  FLSHAALEAGEGQADTWQDAVQLMTLHSAKGLEFPQVFIVGMEEGMFPSQMSLDEGGRLE

Bsub   600  EERRLAYVGITRAEQELYLTNAKMRTLFGRTNMNPESRFIAEIPDDLLENLNEKKETRAT
Gst    599  EERRLAYVGITRAEEELVLTSAQMRTLFGNIQMDPPSRFLNEIPAHLLETASR-------
UvrD   594  EERRLAYVGVTRAMQKLTLTYAETRRLYGKEVYHRPSRFIGELPEECVEEVR--------

Bsub   660  SARKMQPRRCPVSRPVSYASKTGGDTLN---WAVGDKAGHKWGTGTVVSVKGEGEGTEL
Gst    652  -------RQAGASRPAVSRPQASGAVGS---WKVGDRANHRKWGIGTVVSVRGGGDDQEL
UvrD   646  -------LRATVSRPVSHQRMGTPMVENDSGYKLGQRVRHAKFGEGTIVNVEGSGEHSRL

Bsub   717  DIAFPSPVGVKRLLAAFAPIEKQ
Gst    702  DIAFPSPIGIKRLLAKFAPIEKV
UvrD   699  QVAFQG-QGIKWLVAAVARLESV
```

Figure 3.4: Alignment of the protein sequence of PcrA from *B. subtilis* (Bsub) and *G. stearothermophilus* (Gst) and UvrD from *E. coli* (UvrD). Residues shaded in black are conserved, while grey shading indicates similar residues.

The protein sequence of both PcrA$_{Bsub}$ and PcrA$_{Gst}$ were aligned with UvrD from *E. coli* (Figure 3.4), which illustrates that the three proteins have a high sequence similarity. The protein structures of PcrA$_{Gst}$ and UvrD were then compared using the Dali database (Holm & Rosenstrom, 2010). This database used the Protein Data Bank (PDB) structures of PcrA and UvrD to compare the structural similarity of the proteins. The

results returned a sequence ID of 45%, a structural Z score of 40.2 and an RMSD of 2.6 Å. The Z score is a Dali measurement of structural similarity where a value above 2 is considered significant (Holm & Rosenstrom, 2010). RMSD is the root mean squared deviation of the superimposed Cα atoms from the two aligned structures and is another commonly used value in protein structure comparison, all atoms of the $PcrA_{Gst}$ and UvrD structures were used in the RMSD calculation. A value lower than 5 Å is considered similar (Xu & Zhang, 2010). When the protein structures are displayed in the same orientation, the similarity between them becomes apparent (Figure 3.5A & B). This is confirmed by overlaying the two protein structures (Figure 3.5C). The structures of PcrA (PDB:1PJR) and UvrD (PDB:3LFU) were aligned using UCSF Chimera (Pettersen *et al*, 2004).
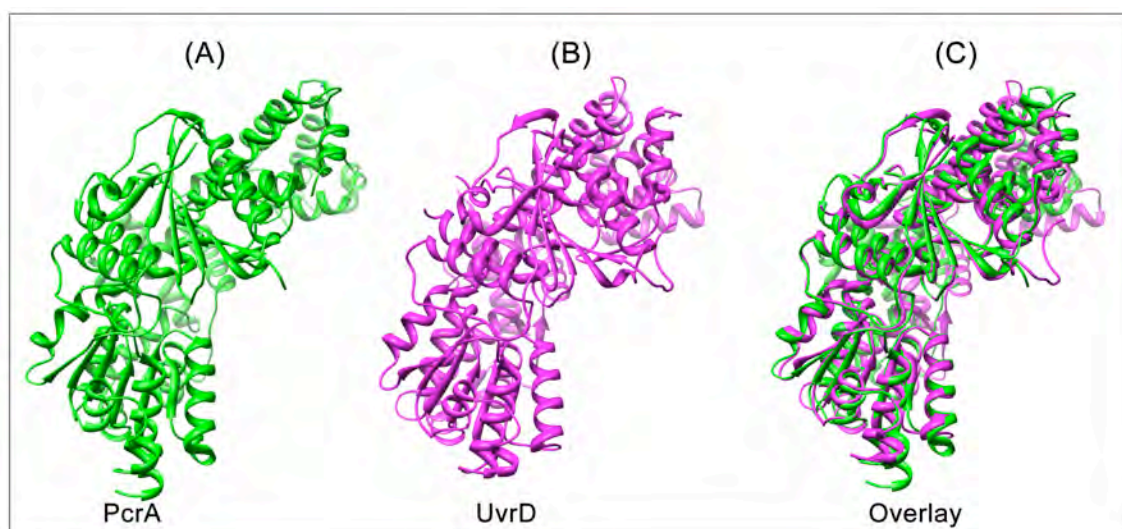

Figure 3.5: (A) Protein structure of $PcrA_{Gst}$, in green. (B) Protein structure of UvrD, in pink. (C) Overlay of the two structures.

3.1.4 Aims

The results presented in this chapter aimed to further characterise the interaction between RNAP and PcrA that was identified in the yeast two-hybrid assay, using a combination of affinity chromatography and far-Western blotting.

3.2 Results

3.2.1 Protein Purification

3.2.1.1 B. subtilis PcrA

Initial work to overproduce and purify PcrA$_{Bsub}$ utilizing an affinity tag was previously carried out. Originally this involved the creation of expression vectors containing *pcrA* with an affinity tag fusion (his$_6$ and GST), which could be used for purification by affinity chromatography. Initial attempts were unsuccessful. Despite exhaustive efforts with media, overproduction strains and conditions, the overproduction of PcrA could not be detected. Instead a plasmid containing un-tagged *pcrA* was received and subsequently used to overproduce and purify PcrA$_{Bsub}$. The plasmid pSMG69 (Figure 3.6) contained wild type *B. subtilis pcrA* under the control of the IPTG inducible T7 promoter (S. McGovern & P. Pollard, INRA, France). The plasmid was created from the parent vector pTYB1 from the IMPACT expression system (New England Biolabs).
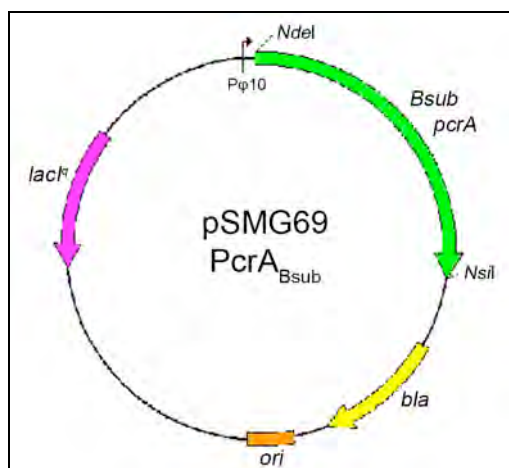

Figure 3.6: pSMG69, used for the overproduction and purification of PcrA. The *B. subtilis pcrA* gene was inserted between the *Nde*I and *Nsi*I restriction site. Also shown is the T7 promoter (P$_{\phi10}$), the origin of replication, the ampicillin resistance gene (*bla*) and the *lacI^q* gene, which encodes the *lac* repressor.

To overproduce PcrA$_{Bsub}$, pSMG69 was transformed into *E. coli* ER2566 (Section 2.4.1.2). PcrA$_{Bsub}$ overproduction was induced in this strain by the addition of 0.05 mM IPTG (Section 2.5.1.1) before the cell pellet was harvested.

PcrA$_{Bsub}$ purification began with two precipitation steps, following cell lysis and the isolation of the soluble supernatant. Firstly, Polymin P was used to precipitate nucleic acid and nucleic acid binding proteins, including PcrA$_{Bsub}$. The resulting pellet was washed with NaCl to release the proteins into solution. Ammonium sulphate was then used to precipitate the PcrA$_{Bsub}$ (Section 2.5.4.1).

Hydrophobic interaction chromatography allowed for the direct re-suspension of the ammonium sulphate pellet and application onto a phenyl Sepharose column. Proteins were eluted from the column using a decreasing salt gradient. The stronger the interaction, the lower the salt concentration had to be to break it. Elutions were run on an SDS-PAGE gel and PcrA$_{Bsub}$ was eluted at the very end of the gradient (Figure 3.7A, P1-P3).

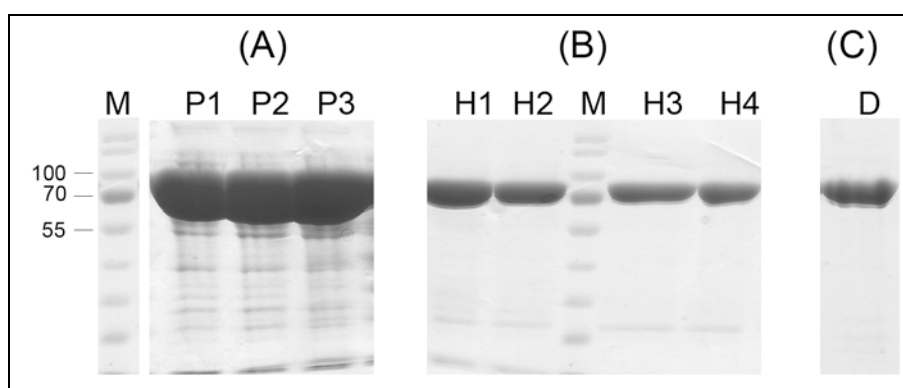Figure 3.7: Purification of PcrA$_{Bsub}$ from pSMG69. (A) Elutions from the phenyl Sepharose column containing PcrA. (B) Elutions from heparin column containing PcrA$_{Bsub}$. (C) Purified PcrA$_{Bsub}$ M – prestained protein ladder (Fermentas), P1-3 – phenyl Sepharose column elutions, H1-4 – heparin column elutions and D – Final PcrA dialysate for storage. Numbers on the left indicate marker sizes (kDa).

The phenyl Sepharose fractions containing PcrA$_{Bsub}$ were pooled and due to being eluted at the end of the gradient had a low NaCl concentration. This allowed the direct application onto the heparin column. The final step in the PcrA purification was the heparin column, which is commonly used in the purification of DNA binding proteins, as it mimics the negatively charged DNA. Figure 3.7B (H1-4) shows the elutions from the heparin column containing PcrA$_{Bsub}$. These samples were pooled and dialysed into storage buffer containing 30% glycerol and stored at -80 °C (Figure 3.7C).

3.2.1.2 G. stearothermophilus PcrA

PcrA from a similar bacterium *G. stearothermophilus* was also cloned, overproduced and purified. This was due to the problems with cloning and overproducing PcrA$_{Bsub}$ and it was thought PcrA$_{Gst}$ may have been easier to work with. Publications had shown that PcrA$_{Gst}$ had been overproduced, purified and successfully crystallised (Subramanya *et al*, 1996). A plasmid (pNG587; Table 2.2) was created using the parent vector pETMCSIII, which contained an N-terminal His$_6$ tag (Figure3.8).
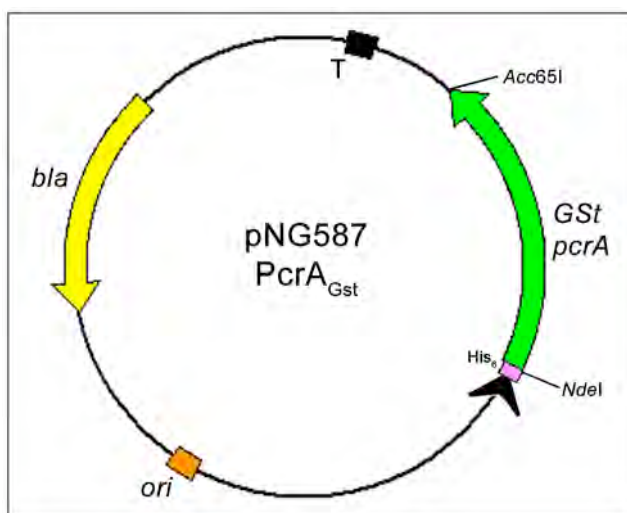
Figure 3.8: pNG587 – pETMCSIII plasmid containing *G. stearothermophilus pcrA* (PcrA$_{Gst}$). (A) Plasmid map of pNG587, with *pcrA* inserted between the *Nde*I and *Acc*65I restriction sites. There is an N-terminal His$_6$ tag before the *Nde*I restriction sites. Features shown are the T7 RNAP promoter (black arrow head) and terminator (T), ampicillin resistance gene (*bla)* and the plasmid origin of replication in *E. coli* (ori).

PcrA$_{Gst}$ overproduction was induced in *E. coli* BL21 (DE3) pLysS (Section 2.4.1.2). The cell pellet was harvested and purified using a HisTrap (GE) nickel column. As shown in Figure 3.9A, the overproduction and purification was successful. The PcrA seen in the elutions (Figure 3.9A E1-3) was concentrated and had a high level of purity. In comparison to PcrA$_{Bsub}$ the purification of PcrA$_{Gst}$ was quicker and simpler.
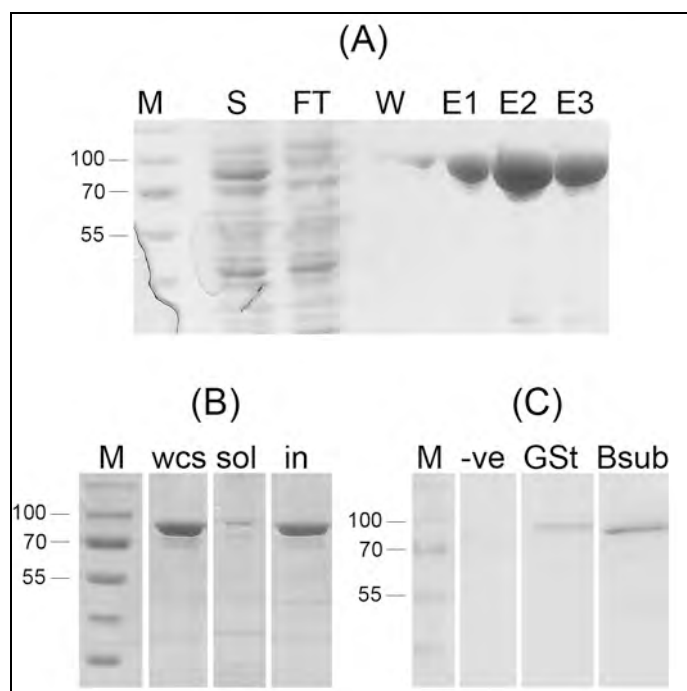
Figure 3.9: (A) Purification of PcrA$_{Gst}$. S - supernatant, FT - flow through, W - 45 mM imidazole wash and E1-5 – 500 mM imidazole elutions. (B) Purified PcrA$_{Gst}$ after dialysis into 200 mM NaCl buffer. WCS – whole cell sample, sol – soluble fraction, in – insoluble fraction. (C) Western blots to determine if the *B. subtilis* anti-PcrA antibody (1:5000) also detected the *G. stearothermophilus* protein. -ve – negative control, no PcrA (T$_0$), GSt - PcrA$_{Gst}$ and Sub – PcrA$_{Bsub}$. M – prestained protein ladder (Fermentas), numbers on the left indicate marker sizes (kDa).

Western blot analysis was performed on the purified PcrA$_{Gst}$ (Figure 3.9C) to ensure that the anti-PcrA antibody raised against the *B. subtilis* PcrA would react with the *G. stearothermophilus* protein. The Western blot showed that the antibody detected both purified PcrA$_{Bsub}$ and PcrA$_{Gst}$. This result was significant as the interaction studies required the use of Western blotting and the anti-PcrA antibody to detect PcrA binding to RNAP.

When PcrA$_{Gst}$ was dialysed into storage buffer, containing a low salt concentration (150 mM NaCl) the sample precipitated out of solution (Figure 3.9B). This meant that the soluble PcrA$_{Gst}$ could only be used in the interaction studies when the salt concentration remained high (500 mM). This was undesirable because the high salt concentration could potentially interfere with binding during the interaction studies that would be performed. To test the amount of NaCl required in the storage buffer for PcrA$_{Gst}$ to remain soluble, a series of dialyses were set up with varying NaCl concentrations. After dialysis the samples were centrifuged at 17 000 x g for 1 min to separate the soluble protein from the insoluble fraction. These samples were then run on an SDS-PAGE gel to assess the effect of lowering the NaCl concentration.
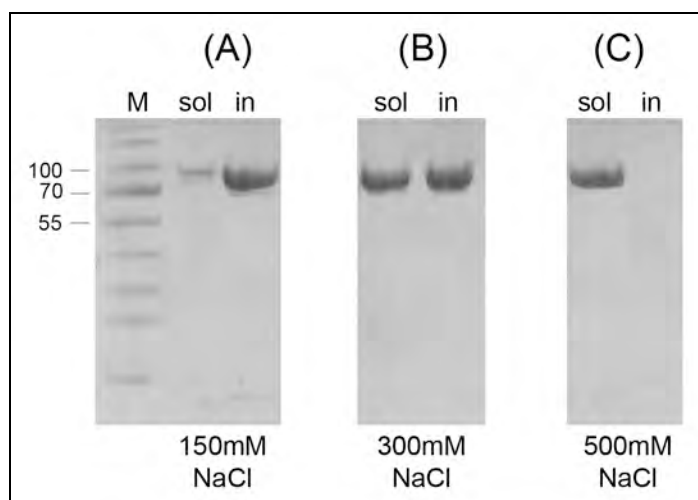
Figure 3.10: Coomassie-blue stained SDS PAGE gel check of PcrA$_{Gst}$ following dialysis. (A) Dialysis from 500 mM NaCl to 150mM NaCl. (B) Dialysis from 500 mM to 300 mM NaCl. (C) Control dialysis into 500 mM NaCl. M – prestained protein marker (Fermentas), sol – soluble fraction and in – insoluble fraction. Numbers on the left indicate marker sizes (kDa).

As expected, PcrA remained soluble in 500 mM NaCl (Figure 3.10C). Figure 3.10A showed that the majority of the PcrA was insoluble when the NaCl concentration was dropped to 150 mM but if the concentration was lowered to 300 mM NaCl,

approximately 50 % of the PcrA remained soluble (Figure 3.10B). This allowed for a

small decrease in the NaCl concentration if required in further assays.

3.2.1.3 E. coli UvrD

PcrA is only found in Gram-positive bacteria, but was found to be ~45 % identical to

the Rep and UvrD helicases in *E. coli* (Petit *et al*, 1998; Chang *et al*, 2002). Initially it

was hoped that UvrD could be obtained from external sources, Prof. N Dixon was

kindly going to provide purified UvrD, which unfortunately could not be located. A

plasmid request was also sent to Prof. S. Matson, for UvrD overproduction plasmid

pET11d-UvrD (Hall & Matson, 1997), which was never received. It was thought that

the cloning, overproduction and purification of UvrD from *E. coli* would be beneficial,

not only in interaction studies but as a control in the strand-displacement assays, when

assessing function. Therefore the *uvrD* gene was inserted into pETMCSIII using the

*Xba*I and *Acc*65I sites to give pNG585 (Figure 3.11).

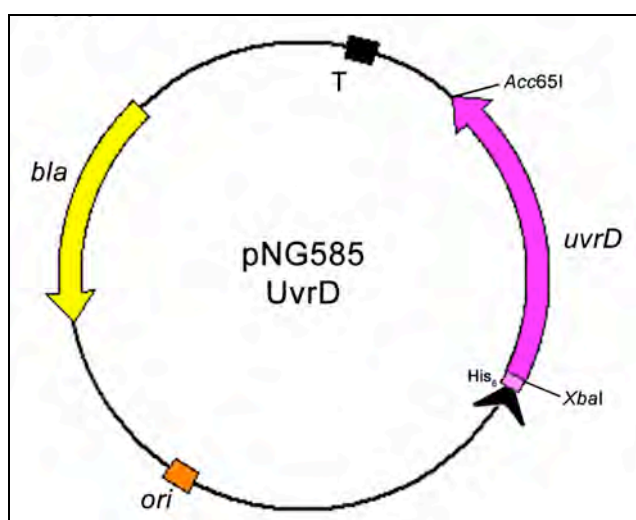Figure 3.11: pNG585 – UvrD overproduction plasmid. Plasmid map of pNG587, with *E. coli uvrD* inserted between the *Xba*I and *Acc*65I restriction sites, with an N-terminal His$_6$ tag. Features shown are the T7 RNAP promoter (black arrow head) and terminator (T), ampicillin resistance gene (*bla)* and the plasmid origin of replication in *E. coli* (*ori*).

UvrD overproduction was then induced in *E. coli* BL21 (DE3) pLysS (Section 2.4.1.2). Although the protein was overproduced it was completely insoluble (Figure 3.12A). *E. coli* BL21 (DE3) pLysS is a common *E. coli* strain used for the overproduction of recombinant proteins, but other strains can sometimes produce more soluble protein, and so trial inductions were performed with these additional *E. coli* strains. Figure 3.12B shows the results using *E. coli* C41 (DE3) and Figure 3.9C shows results using *E. coli* B834 (DE3). In all three strains the overproduced protein is seen only in the whole cell sample (wcs) and the insoluble fraction (in).
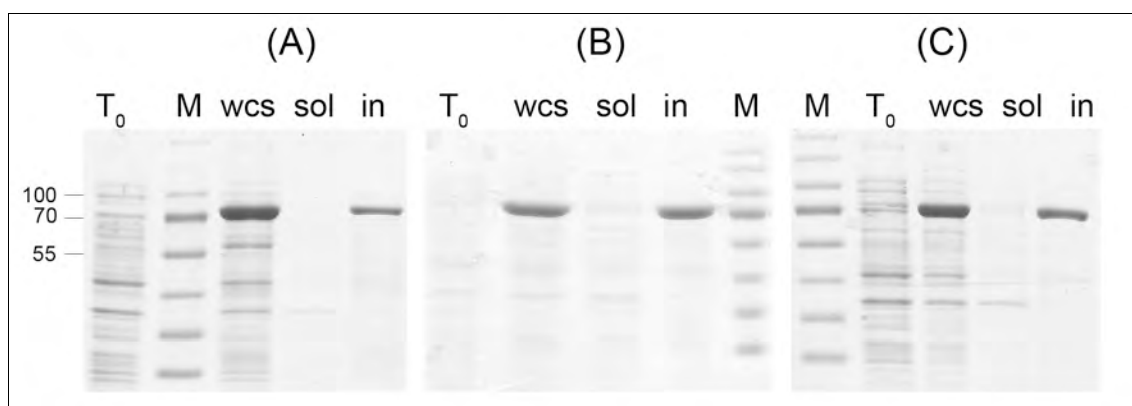
Figure 3.12: UvrD solubility checks, from His-tagged expression plasmid pNG585. (A) BL21 (DE3) pLysS, (B) C41 (DE3) and (C) B834 (DE3). M – protein marker – PAGE ruler (Fermentas), T_0 – before induction, WCS - T_5 whole cell sample, sol - T_5 soluble fraction, in - T_5 insoluble fraction. Numbers on the left indicate marker sizes (kDa).

Multiple attempts were made to overproduce soluble UvrD, by using a variety of induction conditions. The different induction conditions used are summarised in Table 3.2. These conditions were all trialed in the three *E. coli* strains, BL21 (DE3) pLysS, C41 (DE3) and B834 (DE3). Induction conditions that were altered were the concentration of IPTG, the temperature, the time and finally the media, where LB and IPTG were replaced with an auto-induction medium (Appendix I).

Table 3.2: Inductions conditions used in an attempt to overproduce soluble UvrD in all three *E. coli* overproduction strains – BL21 (DE3) pLysS, C41 (DE3) and B834 (DE3).

Induction method	Temperature (°C)	Incubation time (hr)
0.05mM IPTG	22	5
0.05mM IPTG	37	3
0.1 mM IPTG	22	5
0.1 mM IPTG	37	3
0.5 mM IPTG	22	5
0.5 mM IPTG	37	3
Auto-induction medium	22	24
Auto-induction medium	22	48
Auto-induction medium	37	24

A Western blot was performed on the insoluble fraction of the UvrD inductions (Figure 3.13) to determine if the anti-PcrA antibody raised against the *B. subtilis* PcrA would react with the UvrD protein, despite the proteins sharing only 41% identity (Figure 3.5). The Coomassie-blue stained SDS-PAGE gel showed there was more UvrD present on the membrane (Figure 3.13A), but the Western blot showed the anti-PcrA antibody reacted with the positive control (PcrA) only (Figure 3.13B). This was not a surprising result and meant that if UvrD was purified, any interaction studies performed would need an additional antibody, such as the anti-His tag antibody.
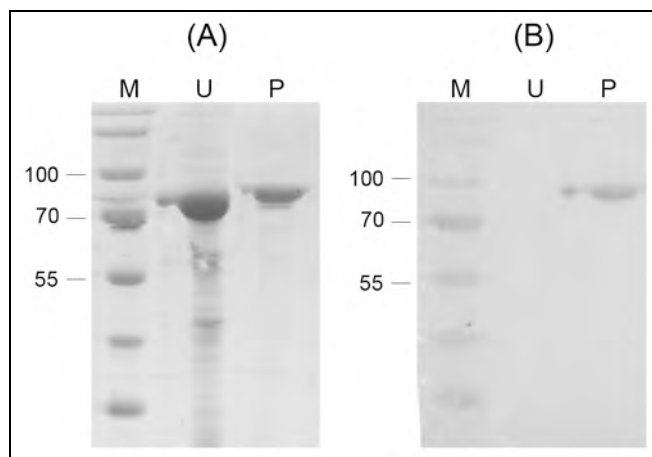
Figure 3.13: Western blot to test the anti-PcrA antibody against overproduced UvrD. (A) Coomassie-blue stained PAGE gel. (B) Western blot membrane. M – protein marker – PAGE ruler (Fermentas), U - UvrD overproduction, P - purified PcrA$_{Bsub}$. Numbers on the left indicate marker sizes (kDa).

Another attempt was made to overcome the UvrD insolubility problem by increasing the cell lysis volume. It was possible during the small trial inductions and solubility checks that the samples were not being re-suspended in enough lysis buffer, which may have caused the overproduced protein to aggregate. Therefore, identical 1 ml induction pellets were re-suspended in increasing volumes of lysis buffer and assessed for UvrD solubility. Increasing the volume from 250 µl to 500 µl and then 750 µl (Figure 3.14 B, C, and D respectively) seemed to affect the solubility slightly. Overproduced UvrD protein in the 250 µl and 500 µl samples were still only seen in the whole cell sample and the insoluble fraction, while the 750 µl soluble fraction did contain some UvrD.
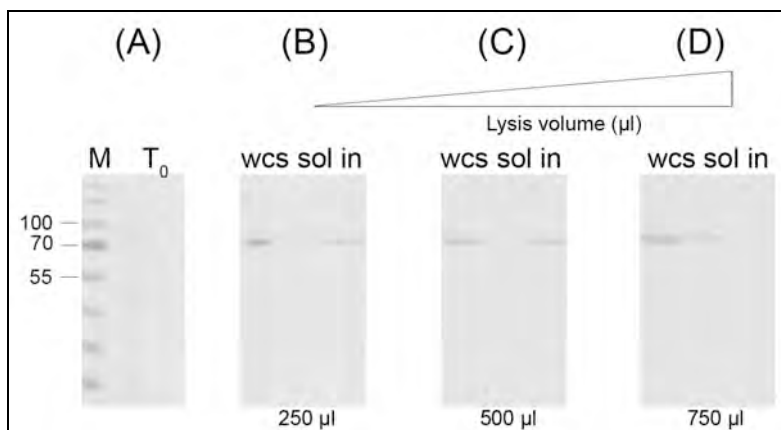
Figure 3.14: Coomassie-blue stained SDS PAGE gel check of UvrD solubility checks, 0.05 mM IPTG at room temperature for 5 h. (A) M – protein marker – PAGE ruler (Fermentas), T_0 – un-induced sample, (B) T_5 1 ml sample resuspended in 250 µl, (C) T_5 1 ml sample resuspended in 500 µl (D) T_5 1 ml sample resuspended in 750 µl . wcs – whole cell sample, sol – soluble fraction, in – insoluble fraction. Numbers on the left indicate marker sizes (kDa).

Despite increasing the lysis volume following a large scale induction of UvrD, multiple attempts were made to purify soluble UvrD. However these proved unsuccessful. Therefore, it was decided to purify UvrD from the insoluble fraction (Figure 3.15). Purification of insoluble UvrD required the resuspension of the insoluble fraction in 8 M urea (Section 2.5.4.2.2). The sample was then applied to a HisTrap (GE) column and protein purified as usual, except that the binding and elution buffers also contained 8 M urea. Column elutions were then pooled and resolubilised using stepwise dialysis was used to slowly remove the urea from the protein sample.
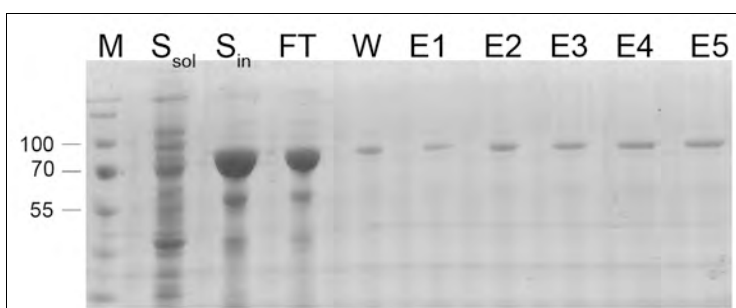
Figure 3.15: Insoluble purification of His-tagged *E. coli* UvrD using HisTrap column (GE). M – protein marker – PAGE ruler (Fermentas), S_{sol} - soluble supernatant, S_{in} – insoluble supernatant, in 8 M Urea, FT – flow through, W – 45 mM imidazole wash, E1-5 – 500 mM imidazole elutions. Numbers on the left indicate marker sizes (kDa).

The UvrD insoluble purification showed that although there was UvrD present in the insoluble 8 M urea sample (S_{in}) most of the UvrD did not bind to the column (Figure 3.15). Instead the majority of the UvrD passed directly through the HisTrap and was seen in the flow through (FT) sample. The small amount of UvrD that was eluted (E1-E5) was pooled and the urea removed by dialysis, but the UvrD precipitated out of solution as soon as the urea concentration was reduced. This protocol was tried using different nickel resins, in an attempt to increase the binding between the resin and UvrD in the presence of 8 M urea. The affinity between the resin and UvrD was never greatly increased and any purified protein continually precipitated when the urea was removed. At this point it was decided that no further effort would be spent on the purification of UvrD.

3.2.2 Establishing the PcrA-RNAP interaction

3.2.2.1 Affinity chromatography

Following the purification of the proteins the interaction between PcrA and RNAP was investigated using affinity chromatography. Some of this work was performed previously, using *B. subtilis* PcrA and RNAP and confirmed a specific interaction occurring between PcrA and RNAP (Harriott, 2006).

Three affinity columns were constructed: an RNAP column, a $PcrA_{Bsub}$ column and a blank negative control column. RNAP and $PcrA_{Bsub}$ were coupled to the resin, as described in Section 2.6.1. The use of these columns would illustrate the interaction detected was specifically occurring between the two proteins. Reciprocal interaction results would also be obtained by applying the protein to their respective protein partner

column: RNAP applied to the PcrA$_{Bsub}$ column, and the reverse PcrA$_{Bsub}$ applied to the RNAP column.

Following the application of the second protein to the column, the interactions occurring between the proteins were allowed to reach equilibrium and then NaCl was added to disrupt the protein-protein interactions (Section 2.6.2). Two NaCl concentrations were used to determine the strength of the interaction occurring.

The results of both the PcrA$_{Bsub}$ column (Figure 3.16A) and the RNAP column (Figure 3.16C) show there is strong interaction between the proteins that can withstand the addition of 0.4 M NaCl. The interaction between RNAP and PcrA$_{Bsub}$ is only broken with the addition of 1 M NaCl. Both RNAP (Figure 3.16B) and PcrA (Figure 3.16D) flowed straight through the blank column, which indicates the proteins were not interacting non-specifically with the affinity resin.
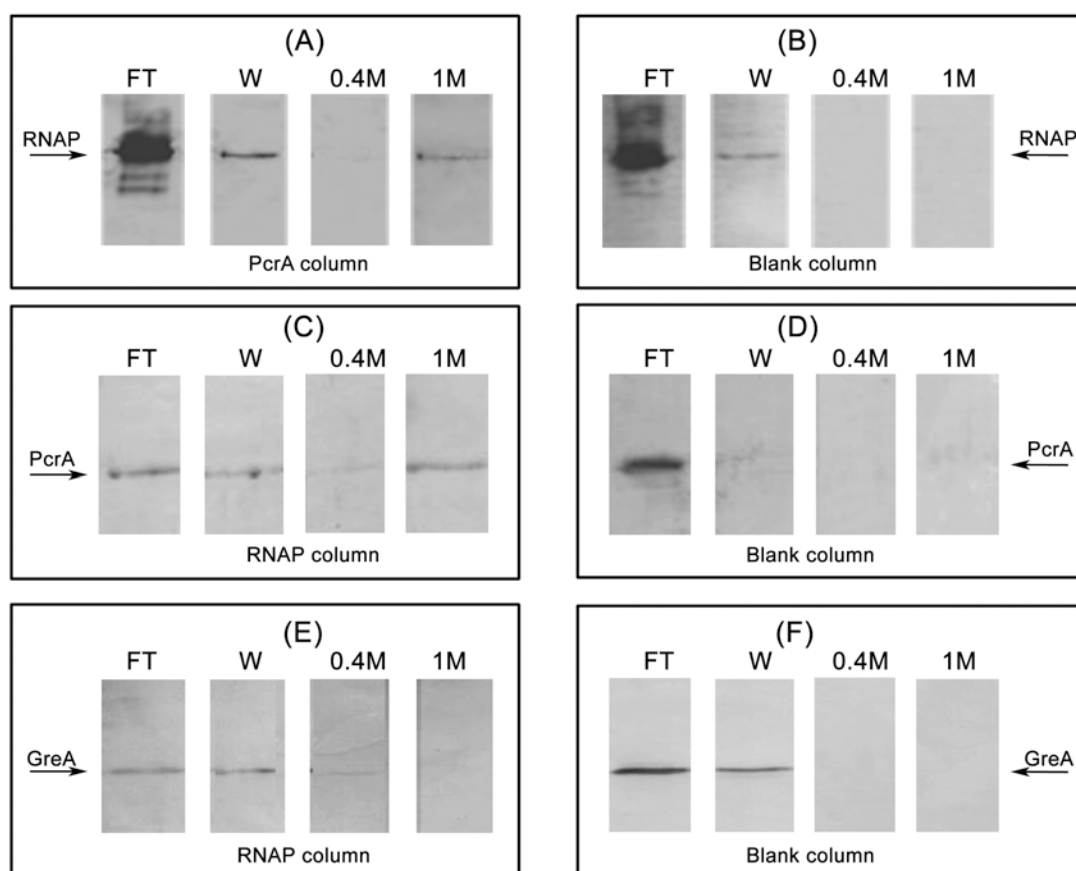
Figure 3.16: Western blot of the affinity chromatography confirming the PcrA$_{Bsub}$–RNAP interaction. (A) RNAP applied to the PcrA affi-gel column, detected using the anti-RpoC antibody. (B) RNAP applied to the blank affi-gel column. (C) PcrA$_{Bsub}$ applied to the RNAP affi-gel column, detected using the anti-PcrA antibody. (D) PcrA$_{Bsub}$ applied to the blank affi-gel column. (E) GreA applied to the RNAP affi-gel column, detected using the anti-GreA antibody. (F) GreA applied to the blank affi-gel column. FT – flow through after application of the protein, W – wash fraction, 0.4 M – 0.4 M NaCl elution and 1 M – 1 M NaCl elution.

As a positive control GreA was also applied to the RNAP and blank columns (Figure 3.16E & F, respectively). This showed that the interaction between PcrA$_{Bsub}$ and RNAP was stronger than the interaction between GreA and RNAP, as GreA was eluted from the RNAP column after the addition of 0.4 M NaCl (Figure 3.16E). The affinity columns showed that the interaction occurring *in vitro* between RNAP and PcrA$_{Bsub}$ is specific and strong.

This method was repeated using PcrA$_{Gst}$ and *B. subtilis* RNAP to determine if the interaction is conserved between different bacterial species. PcrA$_{Gst}$ was applied to the RNAP column (Figure 3.17A) and to the blank column (Figure 3.17B). The flow through, wash and elutions were transferred onto nitrocellulose membrane and detected using the anti-PcrA antibody (Section 2.5.3.1).
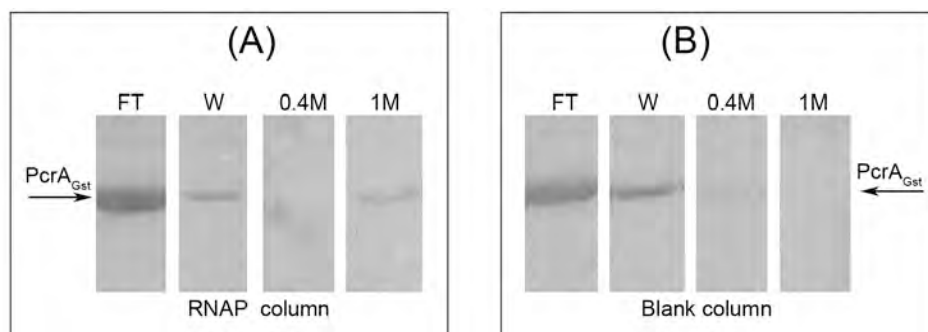
Figure 3.17: Interaction studies between PcrA$_{Gst}$ and *B. subtilis* RNAP using affinity chromatography. (A) RNAP column (B) blank column, FT – flow through, W – wash fraction, 0.4 M – 0.4 M NaCl elution and 1 M – 1 M NaCl elution. Western blots, detected using the anti-PcrA antibody.

The results of the affinity chromatography support the results of the PcrA$_{Bsub}$. The PcrA$_{Gst}$ was detected in the flow through, wash and 1 M elution from the RNAP column (Figure 3.17A), while it was only detected in the flow through and wash of the blank column (Figure 3.17B). This was a significant result because it showed the interaction may occur in other Gram-positive bacteria and that the PcrA$_{Bsub}$ could be replaced with the PcrA$_{Gst}$.

3.2.2.2 Far-Western blot analysis

Far-Western blotting was then used to examine the interaction between RNAP and PcrA in more detail and determine the specific PcrA binding site on RNAP. Fragments of the β and β' subunits of RNAP were used as the prey proteins on the membrane, which were resolubilised in refolding buffer (1 x TFWB + 5% skim milk) and probed with both PcrA$_{Bsub}$ and PcrA$_{Gst}$ (Section 2.5.3.2). The far-Western blot showed that both

PcrA$_{Bsub}$ and PcrA$_{Gst}$ have multiple sites of interaction with RNAP. The overlapping fragments of RNAP transferred to the nitrocellulose membrane are shown in the Coomassie-blue stained gel (Figure 3.18A). This illustrated that approximately equal amounts of each protein fragment were loaded on to the gel.
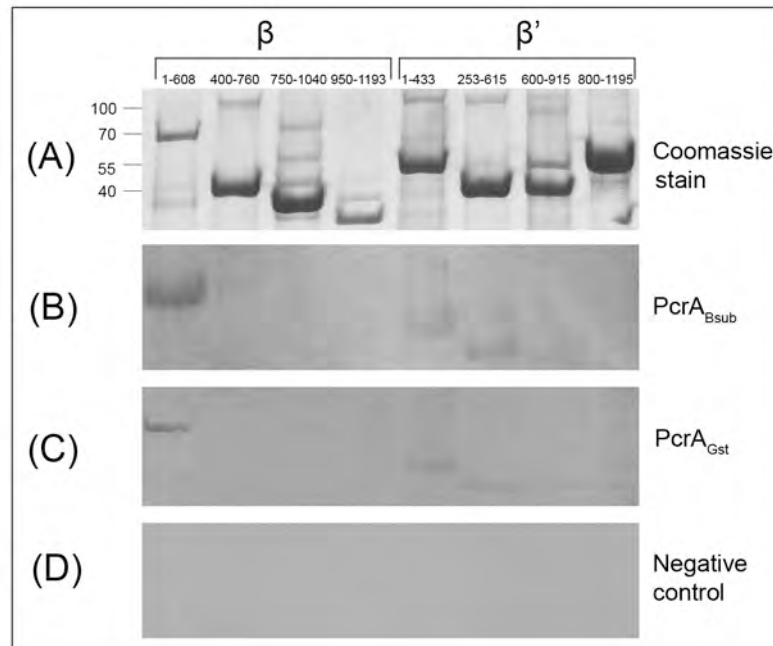
Figure 3.18: Far-Western blot of PcrA and RNAP fragments (A) Coomassie-blue staining of the RNAP fragments, transferred onto the nitrocellulose membrane. Numbers correspond to amino acid coordinates of the subunit fragment. (B) Membrane probed with PcrA$_{Bsub}$. (C) Membrane probed with PcrA$_{Gst.}$ (D) Negative control membrane without a protein probe. Numbers on the left indicate marker sizes (kDa).

Figure 3.18B shows the results of the PcrA$_{Bsub}$ probe, which was shown to bind to 3 different fragments of RNAP. Far-Western blotting was also performed on the α and ω subunits of RNAP, which did not interact with PcrA (not shown). The N-terminal region of β (1-608 aa), was expected to bind PcrA as this was the region used in the original yeast two-hybrid screen. The other two fragments of RNAP detected were in the N-terminal region of the β' subunit (1-433 aa and 253-615 aa). This result was surprising as it was expected that there would only be a single site of interaction between RNAP and PcrA. The interaction between PcrA and the β' subunit was not detected in the original yeast two-hybrid, as the β' fragments were found to be non-

functional. Fortunately the far-Western blot allowed for the immediate refolding and probing of the unstable β' fragments, which enabled the interaction to be detected. The results of the far-Western blot using PcrA$_{Gst}$ detected the same three RNAP fragments as the PcrA$_{Bsub}$ (Figure 3.18C). The identification of multiple binding sites between PcrA and RNAP support the affinity chromatography results, and additionally showed that the interaction between PcrA and RNAP was more complex that originally expected.

3.2.3 Investigation of the interaction in Gram-negative bacteria

3.2.3.1 E. coli RNAP affinity chromatography

After establishing a strong interaction between PcrA and RNAP the next step was to determine if the interaction was conserved with RNAP from other bacterial species. The interaction between PcrA and *E. coli* RNAP was investigated. This was for a number of reasons: firstly, *E. coli* is a model Gram-negative bacterium and PcrA is present in Gram-positive bacteria only. If the interaction was present it would suggest an SF1 helicase-RNAP interaction may be conserved in both Gram-positive and Gram-negative bacteria. Secondly the identification of an interaction with *E. coli* RNAP would be important for later functional studies.

Affinity chromatography was used to determine if there was an interaction occurring between PcrA and *E. coli* RNAP. This method was chosen because it is relatively quick and easy, and had been shown to give specific reliable results using PcrA (Section 3.2.2.1). *E. coli* RNAP was coupled to the affinity resin (Section 2.6.1) and PcrA$_{Gst}$ was applied to the *E. coli* RNAP and blank columns.
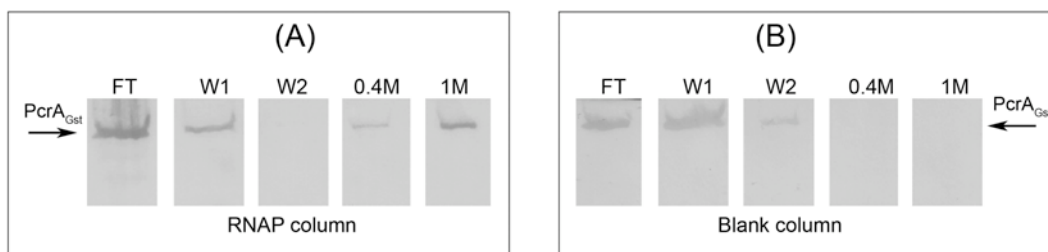
Figure 3.19: Interaction studies between PcrA$_{Gst}$ and *E. coli* RNAP using affinity chromatography. (A) *E. coli* RNAP column; (B) blank column. FT – flow through W1 & W2 – wash fractions, 0.4 M – 0.4 M NaCl elution and 1 M – 1 M NaCl elution.

The results of the RNAP column (Figure 3.19A) show that PcrA$_{Gst}$ was detected in the flow through, first wash and both the 0.4 M and 1 M NaCl elutions. The majority of the PcrA$_{Gst}$ was eluted after the addition of the 1 M NaCl. The absence of a signal in the second wash showed that all nonspecifically bound PcrA$_{Gst}$ was removed in the flow through and the first wash. The detection of the PcrA$_{Gst}$ in both NaCl elutions showed that there was a specific interaction occurring between RNAP and PcrA$_{Gst}$ and the addition of the NaCl elution buffer broke this interaction. The majority of the PcrA$_{Gst}$ bound to the RNAP column was eluted after the addition of the 1 M NaCl, which indicate sthere is a strong interaction with PcrA$_{Gst}$. The result of the blank column (Figure 3.19B) confirms that PcrA$_{Gst}$ did not bind nonspecifically to the resin. This was shown by the presence of PcrA$_{Gst}$ in the flow through and washes only.

This result was an exciting finding because as mentioned, PcrA is not found in Gram-negative bacteria. If this interaction that was identified *in vitro* is occurring *in vivo* in *E. coli,* it must be between RNAP and a different helicase. As previously mentioned (Section 3.1.3) UvrD is an SF1 helicase found in only Gram-negative bacteria, which shares sequence identity and structural homology to PcrA. Therefore UvrD is a good candidate for the RNAP-interacting helicase in Gram-negative bacteria, and these

results suggest that the major SF1 helicase in both Gram-positive and Gram-negative organisms is able to bind tightly to its cognate RNAP.

3.2.3.2 UvrD far-Western blot analysis

To investigate if UvrD could interact with RNAP the far-Western blot was used. This was for two reasons, firstly UvrD could not be purified (Section 3.2.1.3) and secondly the far-Western blot had been shown to identify the interaction correctly. In this case the reciprocal far-Western blot was employed, where UvrD and PcrA$_{Gst}$ were transferred onto the nitrocellulose membrane and probed with *B. subtilis* RNAP and purified fragments of RNAP (Section 2.5.3.2) previously found to bind to PcrA. The Coomassie-blue stained gel demonstrated that the protein loading was not equal and there was more PcrA present than UvrD (Figure 3.20A). This was due to the loading of the insoluble cell lysis fragment verses a purified protein sample. The experiment used the same negative control as previous far-Western blots, which involved no addition of probe protein to ensure that the antibodies did not detect the proteins transferred onto the membrane (Figure 3.20B). This control consistently showed that the antibodies used did not detect either the UvrD or PcrA proteins. An additional negative control experiment was also performed, using GreA, to ensure the UvrD did not bind non-specifically to any protein probe added (Figure 3.20C).
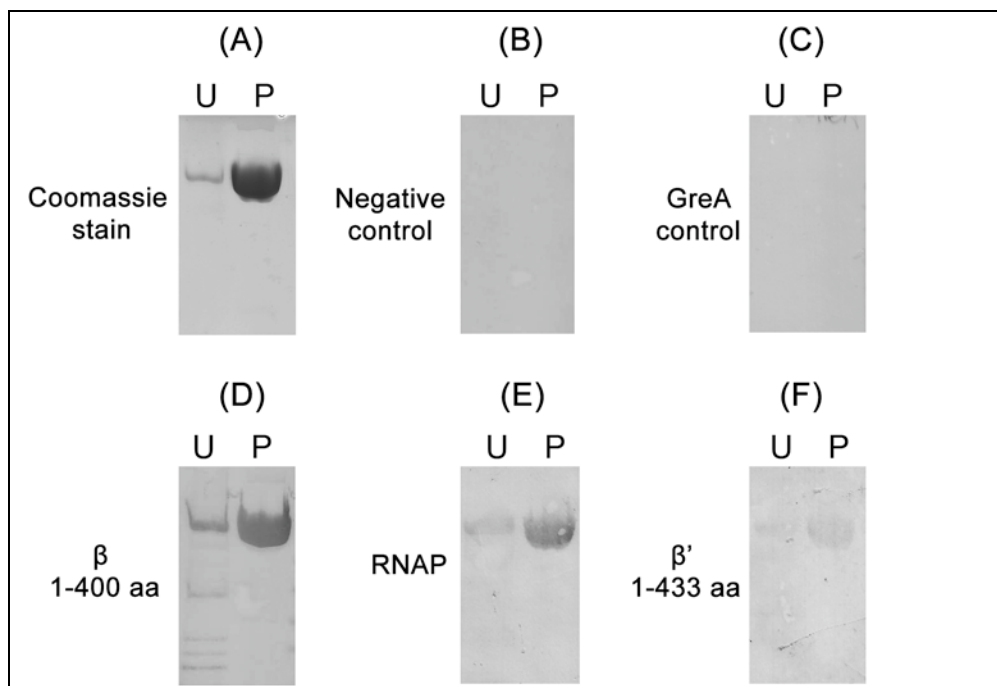
Figure 3.20: Far-Western blot of *E. coli* UvrD and *B. subtilis* RNAP interaction. (A) Coomassie-blue stained SDS-PAGE gel of insoluble UvrD and purified PcrA. (B) Negative control membrane, without protein probe added. (C) Membrane probed with *B. subtilis* GreA protein. (D) Membrane probed with *B. subtilis* β subunit (1-400 aa). (E) Membrane probed with *B. subtilis* RNAP (F) Membrane probed with *B. subtilis* β' subunit (1-433 aa). U – UvrD insoluble overproduction fragment, P – purified PcrA, used as a positive control.

Following the negative control results the membranes were probed with purified *B. subtilis* $\beta_{\text{1-400aa}}$ subunit, $\beta'_{\text{1-433aa}}$ subunit and whole RNAP. A strong interaction occurred between UvrD and the $\beta_{\text{1-400aa}}$ subunit fragment (Figure 3.20D). This is an interesting result as this is the region of RNAP initially detected as interacting with PcrA in the yeast two-hybrid assay. The results of the RNAP (Figure 3.20E) and β' subunit (Figure 3.20F) also show there is an interaction occurring between the two proteins. Overall this was an extremely interesting result that showed that the interaction between PcrA and RNAP might be conserved with SF1 helicases and RNAP in other bacterial species.

3.2.4 RNAP structure analysis

Initial results of the far-Western blots (Figure 3.18) suggested there were at least two separate binding sites for PcrA on RNAP, with PcrA interacting with both the β and β' fragments. To investigate the possible sites of PcrA binding on RNAP, the *B. subtilis* RNAP homology model (Johnston *et al*, 2009) was examined. The β fragment detected in the far-Western blot is highlighted in green and the two β' subunit fragments are highlighted in light blue, on the RNAP structure, with the overlapping region of the two β' fragments in dark blue (Figure 3.21).
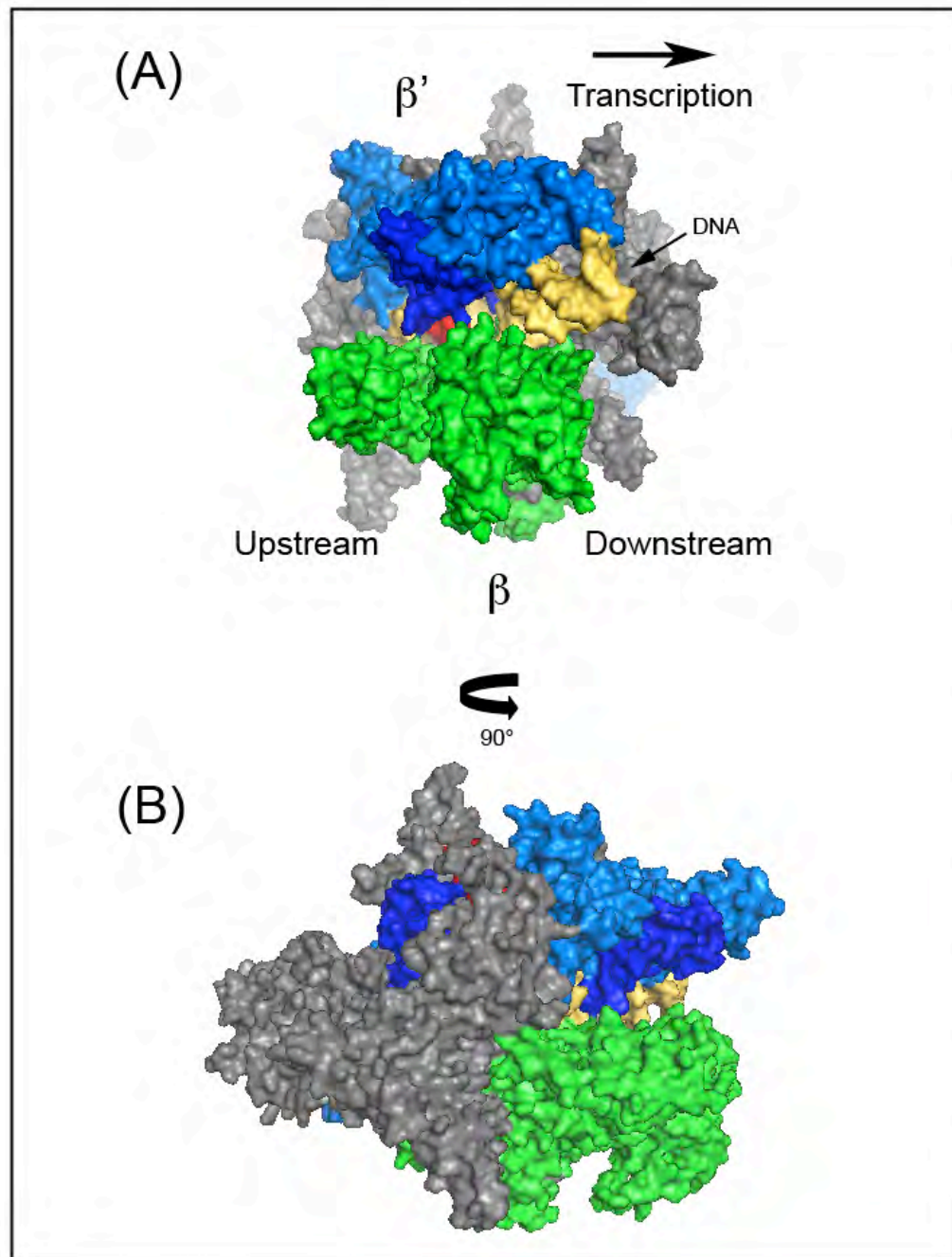
Figure 3.21: *B. subtilis* RNAP homology model. The three fragments detected in the PcrA far-Western blots are highlighted in β (1-608 aa) in green, β' (1-433 aa & 253-615 aa) in blue with the overlapping region in darker blue. (A) RNAP in the face-on orientation, the upstream face on the left and downstream face on the right. The direction of transcription is left to right. (B) The structure has been turned 90° to show the upstream side in profile; the direction of transcription is now going into the page.

When mapping the regions of interaction onto the homology structure of *B. subtilis* RNAP the binding position of PcrA seemed ambiguous. It was possible that multiple PcrA molecules could be bound to the different sites of RNAP *via* the CTD of PcrA (Figure 3.22A). Another possibility is that PcrA was binding to either the downstream or upstream face of RNAP, or possibly but less likely across the ββ' claw, where a single PcrA molecule would encompass both the β and β' subunit binding sites (Figure 3.22B)
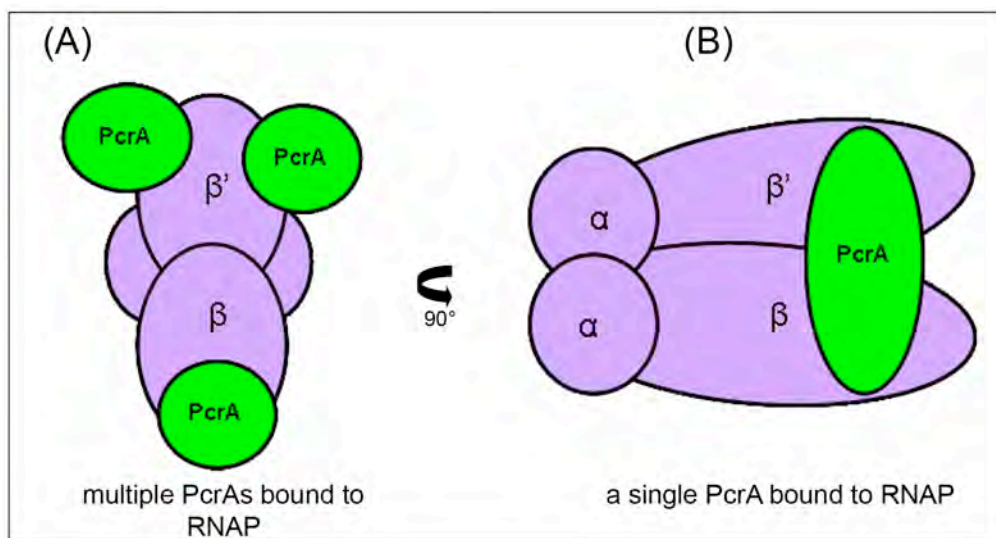
Figure 3.22: Schematic of the possible binding sites for PcrA on RNAP, based on the results of the far-Western blot. (A) The interaction may occur between multiple PcrA CTDs and the sites on the RNAP ββ' subunits. RNAP is shown in the face-on orientation, looking into the ββ'claw, which the DNA enters in the left to right direction. (B) The interaction may also occur between one PcrA molecule and RNAP, indicating that both the CTD and NTD of PcrA would be involved in the interaction. RNAP is shown in the side-on orientation and PcrA could be bound to either the downstream or upstream side.

At this stage, both models of binding between RNAP and PcrA were still a possibility and in order to investigate this further additional interaction studies were undertaken. The interaction studies were designed to utilize smaller fragments of both proteins in order to pinpoint the interaction sites and possible method of interaction between PcrA and RNAP.

3.2.5 Investigating the PcrA C-terminal domain (CTD)-β subunit interaction

3.2.5.1 Vector construction and protein purification

To begin the further investigation of interaction between PcrA and RNAP the C-terminal domain (CTD) of PcrA$_{Bsub}$ was used. This region was identified in the original yeast two-hybrid screen and previous work had produced a plasmid containing the PcrA$_{Bsub}$ CTD fragment with a C-terminal His$_6$-tag for overproduction and purification (Figure 2.23A, Withers 2007). The CTD overproduction was induced in *E. coli* BL21 (DE3) pLysS and the protein was purified (Section 2.5.1, 2.5.4.2; Figure 3.23B).
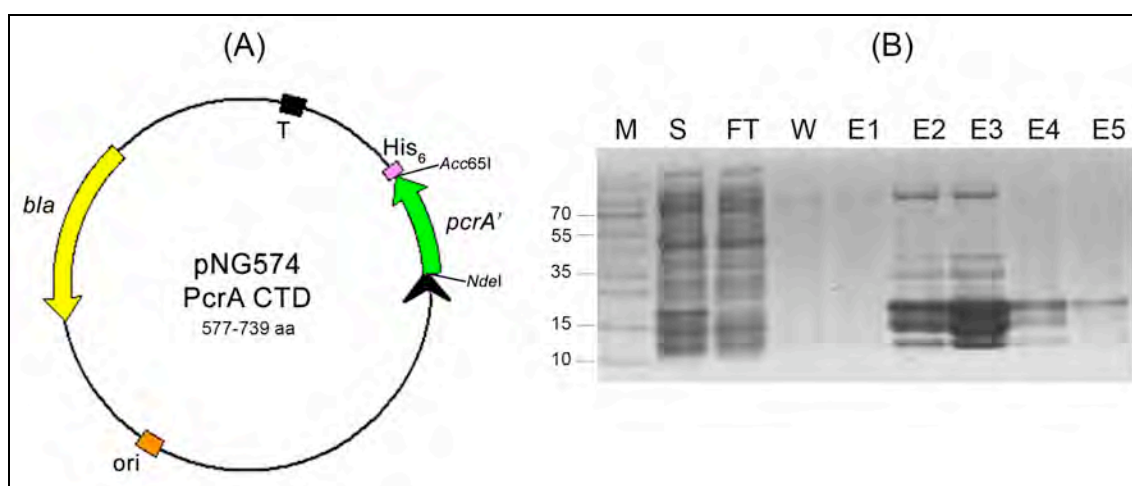
Figure 3.23: Purification of PcrA$_{Bsub}$ CTD. (A) pNG574 plasmid map. Ori – origin of replication, T - terminator, *bla* – ampicillin resistance gene   (B) SDS-PAGE gel of the PcrA CTD purification. M prestained protein ladder (Fermentas), sup – supernatant, FT – flow through, W – 45 mM imidazole wash, E1-5 – 500 mM imidazole elutions. Numbers on the left indicate marker sizes (kDa).

The elutions of the CTD (Figure 3.23B, E1-5) showed multiple protein bands present on the SDS-PAGE gel. This indicated the sample was still contaminated with low molecular weight proteins or the CTD was being degraded. The purification was repeated with the addition of protease inhibitors but this did not affect the results with the multiple bands still present in the final elutions (not shown).

To determine if the sample was contaminated or degraded during the purification of the CTD the dialysate was checked by Western blot (Figure 3.24).
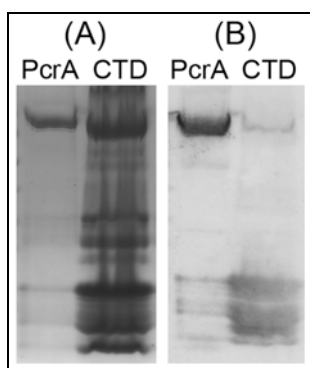
Figure 3.24: Purified CTD. (A) Coomassie-blue stained gel. (B) Western blot. PcrA – purified PcrA$_{Bsub}$, as a positive control. CTD – purified CTD protein.

When the Coomassie-blue stained gel showing the CTD dialysate (Figure 3.24A) was compared to the Western blot (Figure 3.24B) performed using the anti-PcrA antibody it showed that the CTD was being degraded. The lower molecular weight bands seen in the stained gel were also detected using the anti-PcrA antibody, indicating the bands represented smaller fragments of the PcrA$_{Bsub}$ CTD protein. There was some additional contamination of higher molecular weight proteins present in the concentrated dialysate sample seen in the stained gel when compared to the Western blot. It was hoped that the CTD protein in the sample would be adequate for far-Western blotting, as attempts to purify the CTD using other methods had been unsuccessful.

3.2.5.2 Far-Western blot analysis

The purified CTD was then used in the far-Western blot with the RNAP fragments as with the full length PcrA samples (Section 2.5.3.2) to determine where the PcrA CTD interacts with RNAP. Previous yeast two-hybrid results have shown there is an interaction between the CTD and the N-terminal region of the β subunit of RNAP. Due to the lack of yeast two-hybrid results using the β' subunit fragments, it was possible

that the PcrA CTD was also interacting with this region (Figure 3.22). Results of the

CTD far-Western blot showed that the interaction between PcrA CTD and RNAP

occurs between the CTD and the β subunit only. This was shown by the single band

detected in the CTD membrane, corresponding to the β subunit fragment 1-608 aa

(Figure 3.25B).
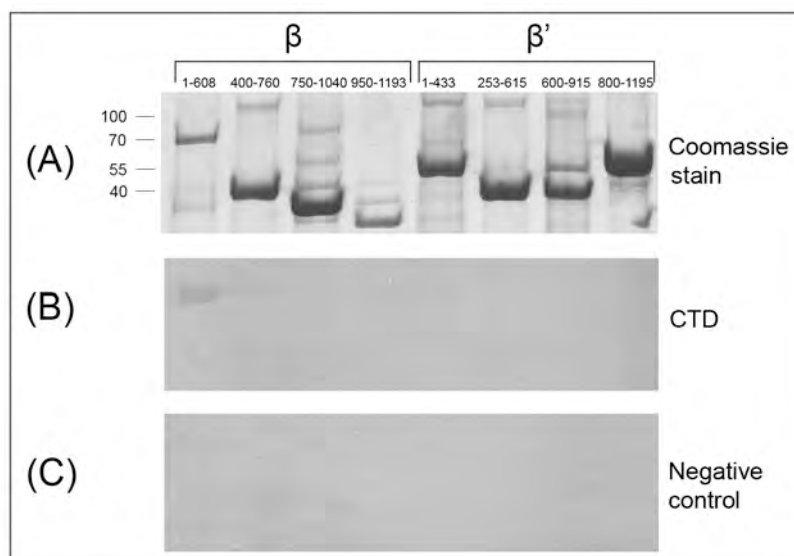
Figure 3.25: Far-Western blot analysis of PcrA$_{Bsub}$ CTD and RNAP fragments (A) Coomassie-blue staining of the RNAP fragments. (B) Membrane probed with CTD. (C) Negative control membrane without a protein probe. RNAP fragments: β – 1-608 aa, 400-760 aa, 750-1040 aa, 950-1193 aa. and β' – 1-433 aa, 253-610 aa, 600 – 915 aa, 800-1199 aa. Numbers on the left indicate marker sizes (kDa).

The interaction detected between the full length PcrA and the β' subunit fragments is

absent when the membrane was probed with the CTD. The Coomassie-blue stained gel

(Figure 3.25A) represented the loading of the RNAP fragments, while the blank control

membrane (Figure 3.25C) again showed the anti-PcrA antibody did not interact with the

RNAP fragments on the membrane.

The results of the CTD far-Western blot indicate a single PcrA molecule is interacting

with RNAP. The PcrA CTD was interacting with only the β subunit and therefore it

was assumed that the PcrA NTD binds to the β' subunit (Figure 3.22B). If the CTD

interacted with both the β and β' subunits this would have supported the idea that multiple PcrAs are binding to RNAP *via* the CTD (Figure 3.22A).

The CTD far-Western blot results also showed it was important to investigate the PcrA RNAP interaction further, including how PcrA is interacting with the β' subunit and the position of PcrA on RNAP in relation to transcription. If a single PcrA was binding to both the β and β' subunits, with the CTD binding to the $\beta_{1-608aa}$ subunit, it is possible that the PcrA could be binding to RNAP in three different orientations. This was due to the region of the β subunit fragment used in the far-Western blot encompassing both upstream and downstream sides and the tip of the β claw (Figure 3.21). Firstly the PcrA could be bound to the side of RNAP on either the upstream or downstream face (Figure 3.26A). Alternatively, PcrA could be binding across the face of the claw (Figure 3.26B). This type of interaction would interfere with the entry of the DNA into the claw and active site of RNAP during transcription initiation.
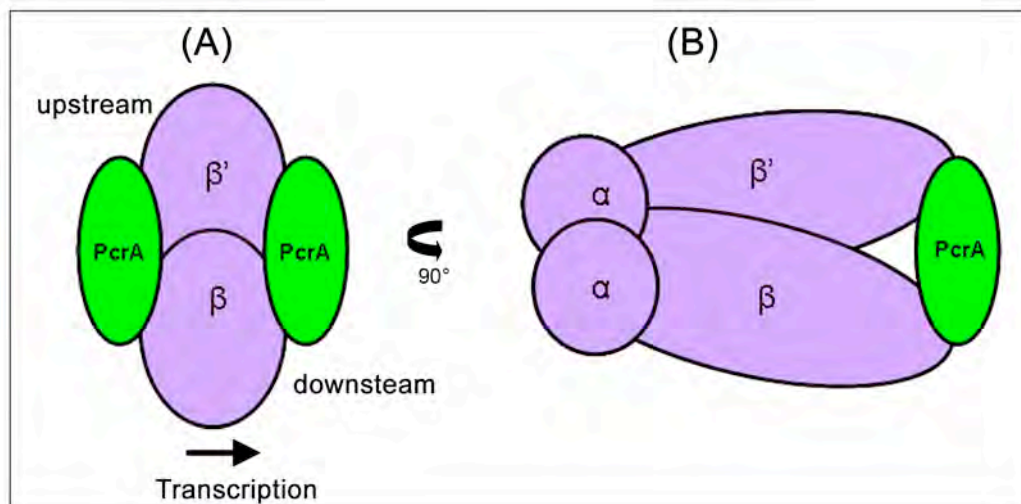
Figure 3.26: Schematic of the possible binding sites for PcrA on RNAP, assuming a single PcrA in involved in the interaction. (A) The interaction may occur between PcrA and RNAP on either the upstream or downstream face of RNAP. RNAP is shown in the face-on orientation, looking onto the ββ'claw, which the DNA enters in the left to right direction. (B) The interaction may also occur across the ββ' claw; RNAP structure is shown turned 90° to give the side-on view.

3.2.5.3 Further defining the PcrA CTD – β subunit interaction

To further specify the region of interaction smaller β subunit fragments were used in the far-Western blotting. The initial $\beta_{1-608aa}$ fragment was broken into 3 smaller fragments $\beta_{1-109aa}$, $\beta_{109-250aa}$ and $\beta_{250-427aa}$ and cloned into two vectors, with and without a His$_6$-tag. The overproduction of untagged protein was to enable the use of His-tagged probe proteins in the far-Western blots, when the β subunit fragments were transferred onto the membrane. Overproduction of His$_6$-tagged proteins would be used for purification if any of the fragments produced soluble proteins (Section 2.5.1).

3.2.5.3.1 β subunit sub-fragments – vector construction

The three β subunit sub-fragments were highlighted in the structure of RNAP in the face-on orientation (Figure 3.27A). The $\beta_{1-109aa}$ region was highlighted in red (Figure 3.27B) and is on the upstream face of RNAP with the next fragment, the $\beta_{109-250aa}$ highlighted in yellow (Figure 3.27C) and stretching across to the downstream face of RNAP. The third fragment is the $\beta_{250-427aa}$ region, highlighted in pink (Figure 3.27C) and is on the bottom of the β subunit, close to the N-terminal region highlighted in red. There is an additional protein fragment $\beta_{427-608aa}$ highlighted in blue (Figure 3.27D), which makes up the end of the original $\beta_{1-608aa}$ fragment detected in the far-Western blot. This region was not overproduced because this region is also found in the second β subunit fragment (400-760 aa) used in the original far-Western blots (Figure 3.18). PcrA does not interact with the $\beta_{400-760aa}$ fragment in the far-Western blot and therefore it was thought that PcrA would not bind to the $\beta_{427-608aa}$ fragment. This region also looks like it is slightly buried within the β subunit structure and therefore would not easily contact other proteins following RNAP assembly.
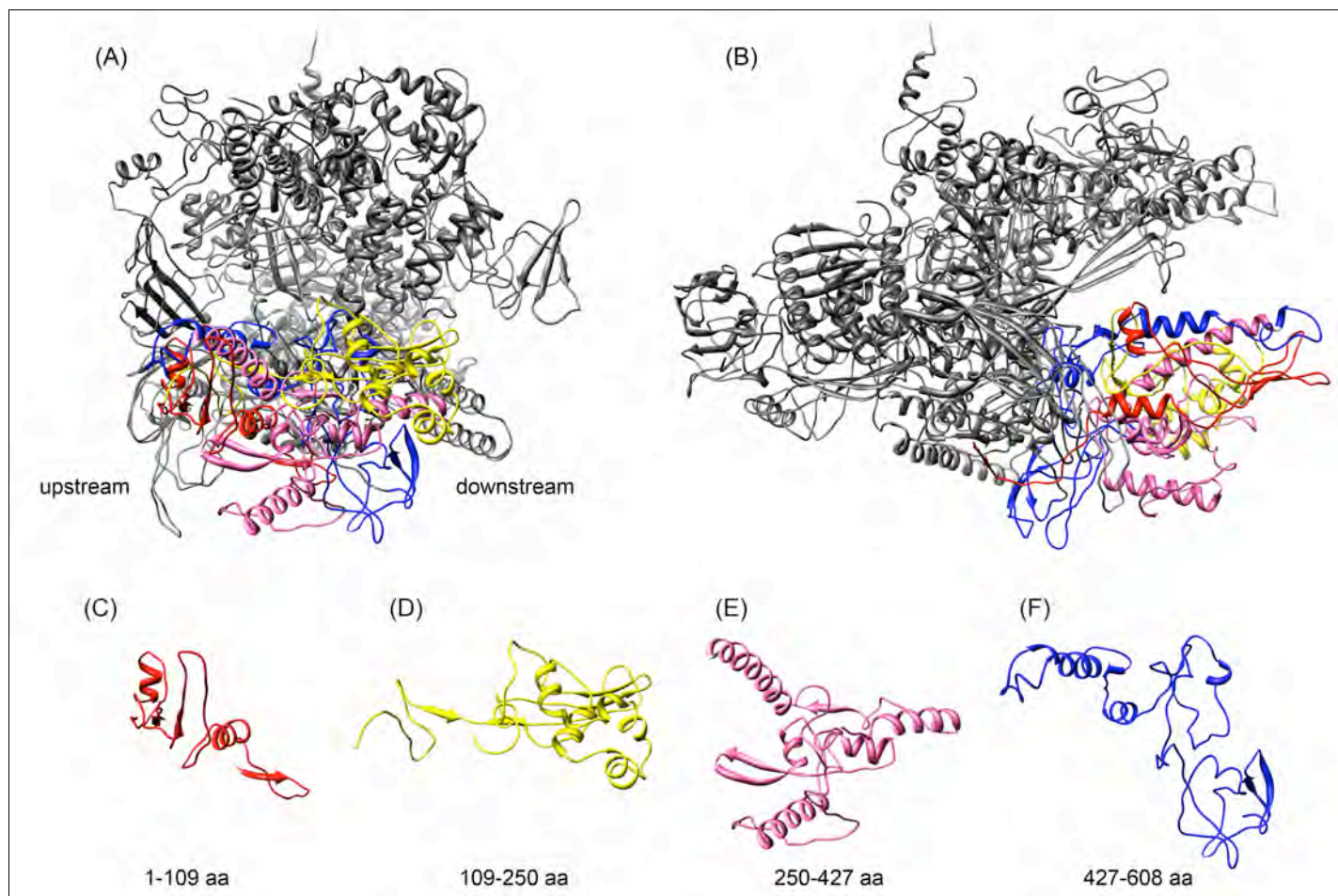
Figure 3.27: RNAP with the β sub-fragments highlighted. (A) RNAP structure in the face on orientation. (B) RNAP structure turned 90° for side on view, with β subunit on the bottom. The four fragments of the β subunit are highlighted. 1-109 aa in red. 109-250 aa in yellow, 250-427 aa in pink and 427-608 aa in blue. (C) 1-109 aa region only. (D) 109-250 aa region only (E) 250-427 aa region only. (F) 427-608 aa region only.

The construction of the plasmids was checked using restriction digest checks (Figure 3.28D) and DNA sequencing to ensure the β subunit fragments were correctly cloned into both the pETMCSIII and pNG209 vectors. The plasmid maps of the β subunit fragments in the parent vector pNG209 are shown below: pNG841 ($\beta_{1-109aa}$, Figure 3.28A), pNG842 ($\beta_{109-250aa}$, Figure 3.28B) and pNG857 ($\beta_{250-427aa}$, Figure 3.28.C)
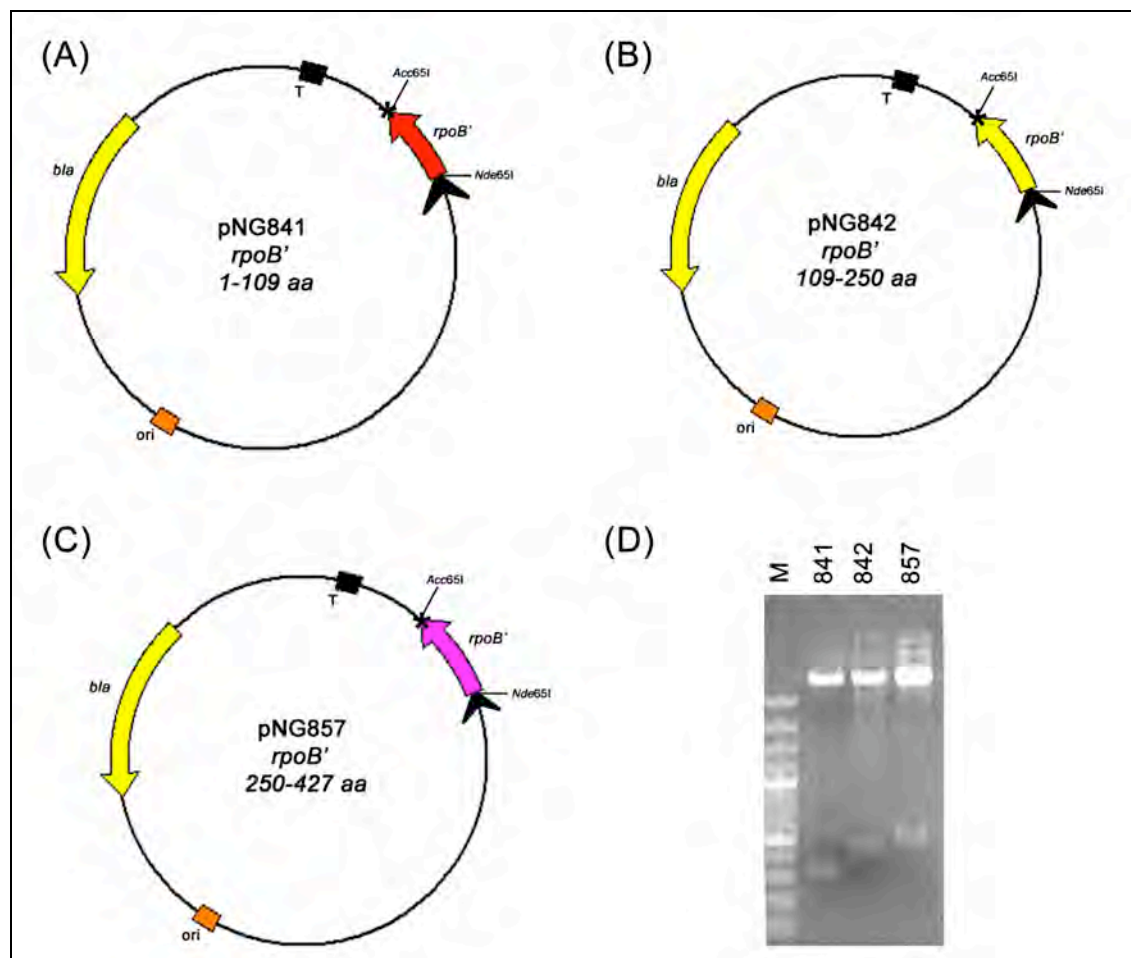
Figure 3.28: Cloned fragments of the β subunit (*rpoB*). (A) pNG841 plasmid map. (B) pNG842 plasmid map. (C) pNG857 plasmid map. (D) Positive digest checks of each plasmid. Features shown on plasmid maps are the T7 RNAP promoter (black arrow) and terminator (T), ampicillin resistance (*bla*) and the plasmid origin of replication in *E. coli* (ori), and stop codon (*)

The β sub-fragments were then overproduced, along with the positive control $\beta_{1-608aa}$ fragment. The overproduced fragments are indicted by arrows on the Coomassie-blue stained gel (Figure 3.29A), and the same fragments were confirmed using Western blotting with the anti-β antibody (Figure 3.29B).
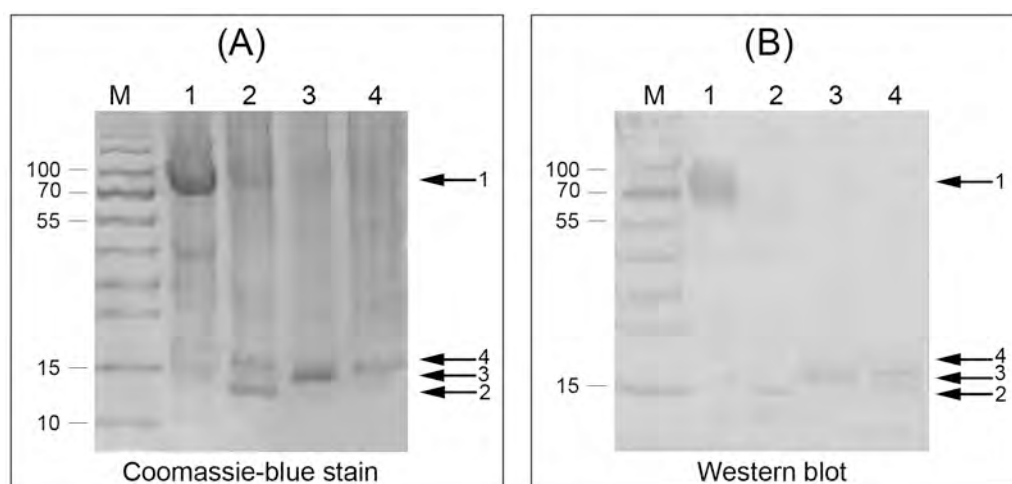
Figure 3.29: Overproduction and Western blot checks of the β sub-fragments. (A) Coomassie-blue stain SDS-PAGE gel of the whole cell inductions samples. (B) Western blot of the samples using the anti-β antibody. M – prestained protein ladder (Fermentas), 1 – $\beta_{1-608aa}$, 2 – $\beta_{1-109aa}$, 3 – $\beta_{109-250aa}$ and 4 – $\beta_{250-427aa}$. Arrows indicate the overproduced fragment in the corresponding lane. Numbers on the left indicate marker sizes (kDa).

Following the successful overproduction, these β sub-fragments could be used in the far-Western blot and probed with PcrA. The identification of the smaller region of the β subunit responsible for interacting with PcrA would help provide information on the position of the PcrA-RNAP interaction.

3.2.5.3.2 Far-Western blot analysis

The β sub-fragments overproduced above, $\beta_{1-608aa}$, $\beta_{1-109aa}$, $\beta_{109-250aa}$ and $\beta_{250-427aa}$ were transferred to nitrocellulose membrane and probed (Section 2.5.3.2) with both full length $PcrA_{Bsub}$ and $PcrA_{Bsub}$ CTD. Firstly the Coomassie-blue stained gel (Figure 3.30A), showed that the sub-fragments were present on the membrane in similar quantities. The membrane probed with $PcrA_{Bsub}$ (Figure 3.30B) showed that the $\beta_{1-608aa}$

positive control was detected, but none of the three sub-fragments were found to interact with PcrA$_{Bsub}$. This trend was repeated when the CTD was used as the probe (Figure 3.30C), again only the $\beta_{1\text{-}608aa}$ fragment was detected. The negative control blot, which did not contain a protein probe, was blank, as expected (Figure 3.30D)
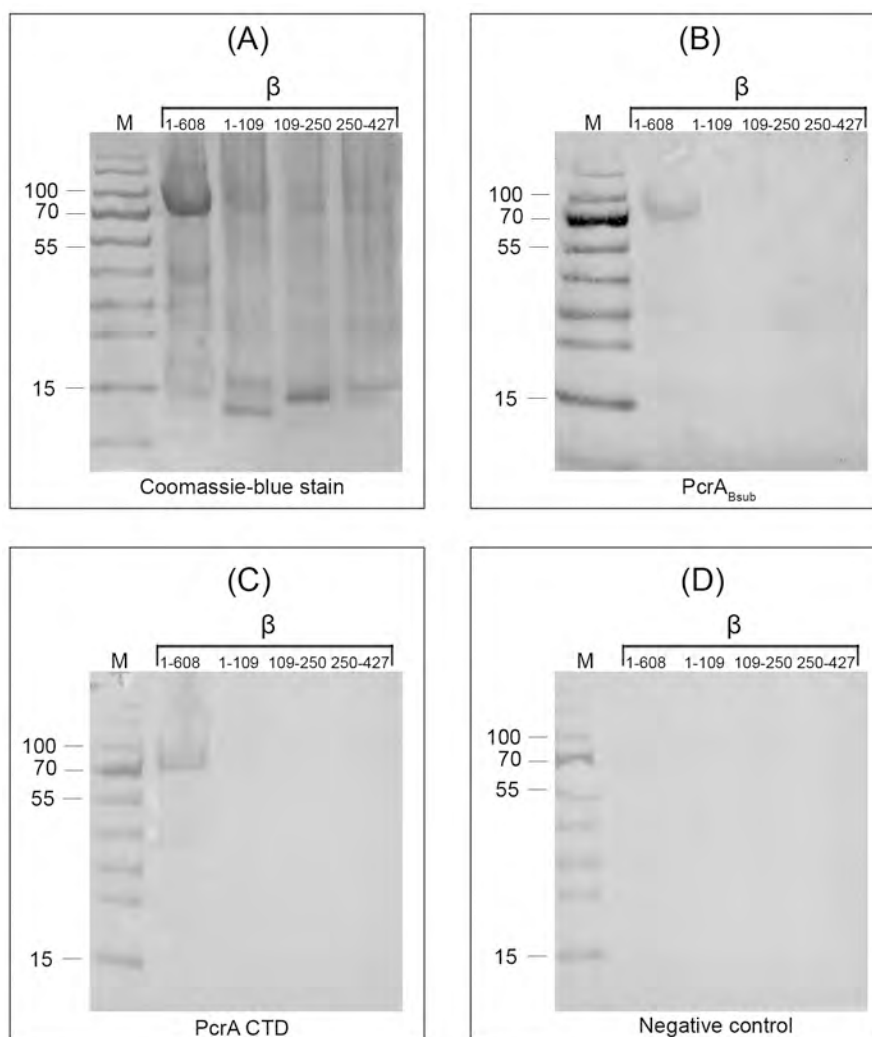
Figure 3.30: Far-Western blot of β sub-fragments. (A) Coomassie-blue stain of β sub-fragments. (B) The membrane was probed with PcrA$_{Bsub}$. (C) The membrane was probed with PcrA$_{Bsub}$ CTD. (D) Negative control – no probe added. M – prestained protein ladder (Fermentas), $\beta_{1\text{-}608aa}$, $\beta_{1\text{-}109aa}$, $\beta_{109\text{-}250aa}$ and $\beta_{250\text{-}427aa}$. Numbers on the left indicate marker sizes (kDa).

There could be a couple of different reasons for not detecting an interaction between PcrA$_{Bsub}$ and the β sub-fragments. Firstly, the smaller fragments of the β subunit do not overproduce as well as the larger fragment and therefore there is less of the protein available for PcrA$_{Bsub}$ to bind. Secondly, the sub-fragments were created in an attempt

to overproduce soluble protein, but instead it may have separated the regions which interact with PcrA. For example if PcrA interacts with the 70-150 aa region of the β subunit, this region is now broken into two different fragments and PcrA may not bind to either. Unfortunately this result means the site of interaction between PcrA and the β subunit cannot be further defined and cannot provide more information on the position of PcrA when bound to RNAP. Other methods will instead be used to determine this, including the investigation of the PcrA NTD-β' subunit interaction.

3.2.6 PcrA N-terminal domain (NTD) – β' subunit interaction

The next step was to investigate the binding site between the β' and the N-terminal domain (NTD) of PcrA$_{Bsub}$. Again smaller regions of both the β' subunit and PcrA were overproduced for use in far-Western blotting.

3.2.6.1 Vector construction and protein purification

To investigate the regions of interaction between PcrA and RNAP the proteins were broken down into smaller fragments for protein overproduction and purification. This first required the construction of overproduction plasmids containing the PcrA NTD fragments. The regions identified for cloning were chosen to isolate different features in the protein (Figure 3.31).
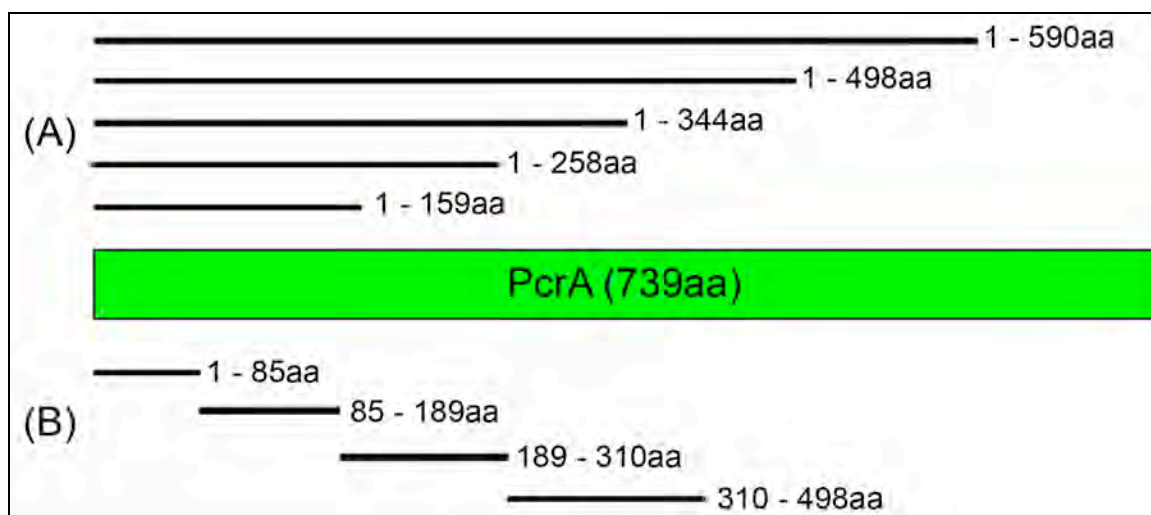
Figure 3.31: A schematic showing the N-terminal fragments of PcrA that were cloned and overproduced. The fragments are shown in amino acids (aa). (A) N-terminal fragments that increase in size all starting from the 1 aa. (B) N-terminal step wise smaller fragments.

Initially, NTD fragments of PcrA of increasing size were cloned into expression vectors for overproduction and affinity purification. These fragments all started from amino acid 1 as shown in the top of the schematic (Figure 3.31A) and were cloned into parent vector pNG209 containing a C-terminal His$_6$-tag. Small scale overproduction trials in *E. coli* BL21 (DE3) pLysS (Section 2.5.1.2) were used to determine if the PcrA fragments would be soluble. Unfortunately while all the fragments were overproduced, they were all insoluble.

Insoluble overproduction of the NTD proteins was seen in the whole cell sample (wcs) and insoluble fraction (in) of four of the NTD fragments PcrA$_{1-159aa}$ (Figure 3.32A), PcrA$_{1-258aa}$ (Figure 3.32B), PcrA$_{1-498aa}$ (Figure 3.32D) and PcrA$_{1-590aa}$ (Figure 3.32E). In another attempt to overproduce soluble NTD proteins the fragments were also cloned into another overproduction vector, pNG698. This vector is similar to pNG209 except that the C-terminal His$_6$-tag was replaced with a C-terminal CHIP-tag (His$_{10}$ and

StrepII), which can help increase the solubility. However, the resulting clones did not overproduce soluble protein either (not shown).
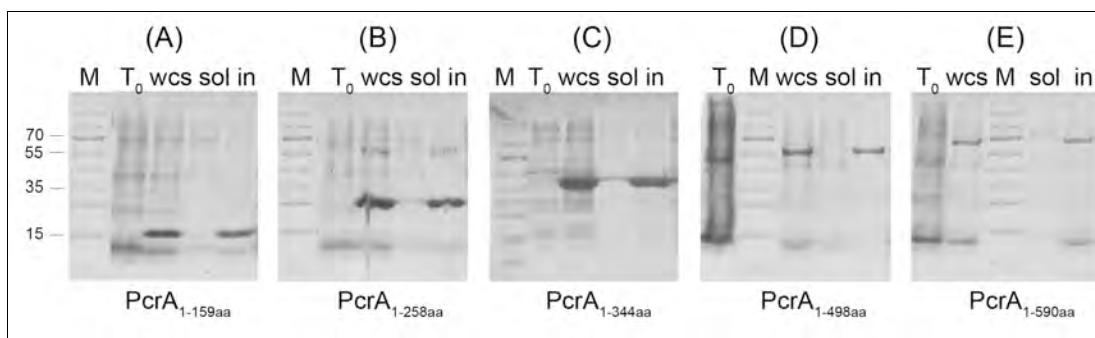
Figure 3.32: Solubility checks of PcrA NTD overproduction. (A) pNG710 - PcrA$_{1-159aa}$. (B) pNG712 - PcrA$_{1-258aa}$. (C) pNG714 – PcrA$_{1-344aa}$. (D) pNG716 - PcrA$_{1-498 aa}$. (E) pNG718 - PcrA$_{1-590 aa}$. M – protein marker – PAGE ruler (Fermentas), T$_0$ – un-induced sample, wcs – T$_5$ whole cell sample, T$_5$ sol – soluble fraction, T$_5$ in – insoluble fraction. Numbers on the left indicate marker sizes (kDa).

One of the NTD fragments was shown to be slightly soluble during the small scale overproduction: PcrA$_{1-344aa}$ (Figure 3.32C). The protein overproduction was up-scaled and the protein purified (Section 2.5.4.2). The results of this purification can be seen in Figure 3.33. The 500 mM imidazole elutions (Figure 3.33A E1-E3) show that the purified protein samples contained some contaminants but the PcrA$_{1-344aa}$ fragment was the major protein present. Elutions were kept for dialysis and storage and separated into two samples, E2-E4 were the more contaminated and became dialysate 1 (D1), while E5-E6 were less concentrated but had a higher purity and become dialysate 2 (D2) (Figure 3.33B & C). The dialysate samples were checked using SDS PAGE and Western blot. The Coomassie-blue stained gel (Figure 3.33B) showed the D1 sample was still highly contaminated, and possibly was being degraded but the D2 sample had a small amount of purified protein present. These samples were also checked *via* Western blotting utilizing the anti-His antibody (Figure 3.33C), which showed that the His$_6$-tagged protein was present in both samples. The anti-PcrA antibody could not be used

to detect the protein as the anti-PcrA antibody does not react with the N-terminal region of PcrA, which is discussed below.
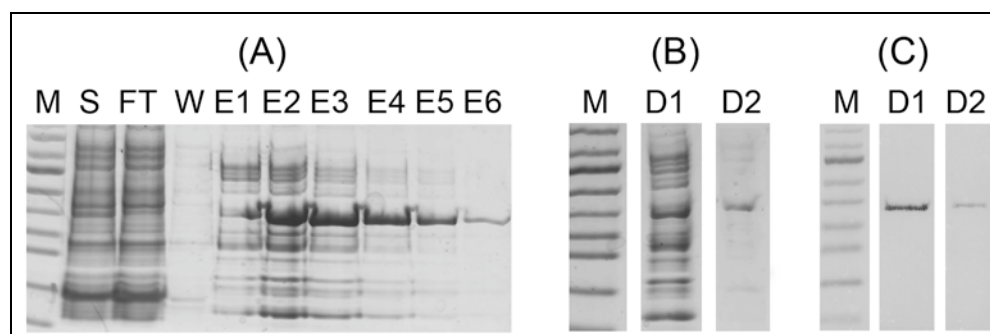
Figure 3.33: $PcrA_{1-344aa}$ purification from pNG714. (A) Purification of $PcrA_{1-344aa}$. (B) Dialysate after the purification in two different aliquots. (C) Western blot of the dialysates using the anti-His antibody. M – protein marker – PAGE ruler (Fermentas), S – supernatant, FT – flow through, W – 45 mM imidazole wash, E1-6 – 500 mM imidazole elutions, D1 – dialysis sample 1, D2 – dialysis sample 2.

The anti-PcrA antibody was tested against the NTD fragments and the anti-His antibody was used as a positive control because all the fragments were His_6-tagged. To test the antibody, the overproduced proteins were transferred to duplicate nitrocellulose membranes and detected with either the anti-PcrA or anti-His antibodies. The Coomassie-blue stained SDS-PAGE gel in Figure 3.34A showed that the overproduced PcrA fragments were equally loaded onto the gel. The Western blot using the anti-His antibody (Figure 3.34B) detected all of the PcrA fragments seen with the Coomassie-blue stain. The membrane using the anti-PcrA antibody shows only the last two fragments could be detected (Figure 3.34C). These two fragments correspond to amino acids 1-498 and 1-590. The next largest PcrA fragment (1-344 aa) could not be detected with the anti-PcrA antibody. Therefore, the anti-PcrA antibody is only able to detect epitopes after amino acid 344. This was highly unexpected as the polyclonal antibody was raised against the full length PcrA protein. This lack of cross reactivity between protein and antibody has been previously reported in the literature where, for example, anti-RfaH antibodies were shown to recognise epitopes in the CTD only (Belogurov *et al*, 2010).
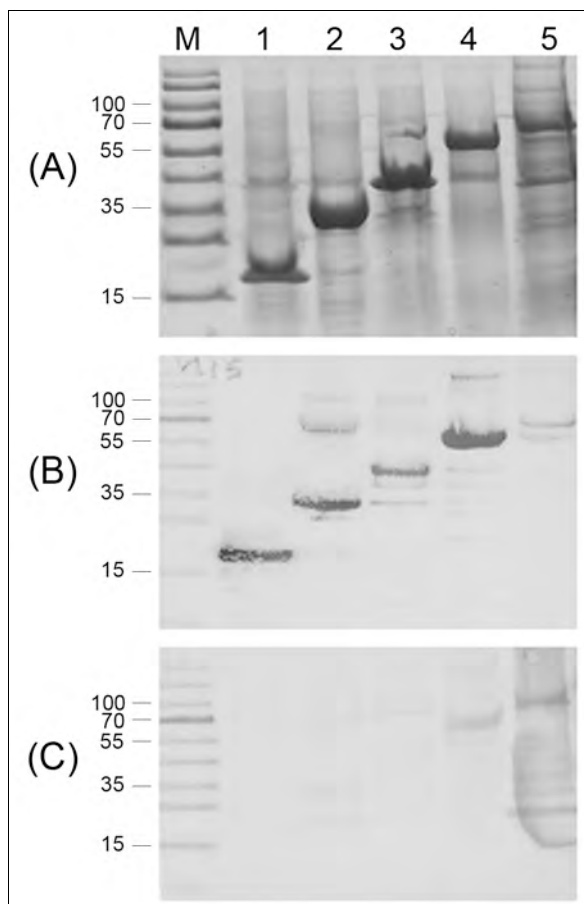
Figure 3.34: PcrA NTD fragments. (A) Coomassie-blue stained SDS-PAGE gel of the overproduced proteins. (B) Western blot – anti-His antibody. (C) Western blot – anti-PcrA antibody. M - prestained protein ladder (Fermentas). 1 – PcrA$_{1-159aa}$, 2 – PcrA$_{1-258aa}$, 3 – PcrA$_{1-344aa}$ 4 – PcrA$_{1-498aa}$, 5 – PcrA$_{1-590aa}$. Numbers on the left indicate marker sizes (kDa).

3.2.6.2 Initial far-Western blot analysis

Due to the fact that the overproduced PcrA$_{Bsub}$ NTD fragments were mostly insoluble and could not be easily purified, the far-Western blot was used to investigate the interaction between the PcrA NTD and RNAP. The PcrA fragments were transferred onto the nitrocellulose membrane. NTD proteins on the membrane were refolded and the membrane was probed with RNAP and β and β' fragments (Section 2.5.3.2).
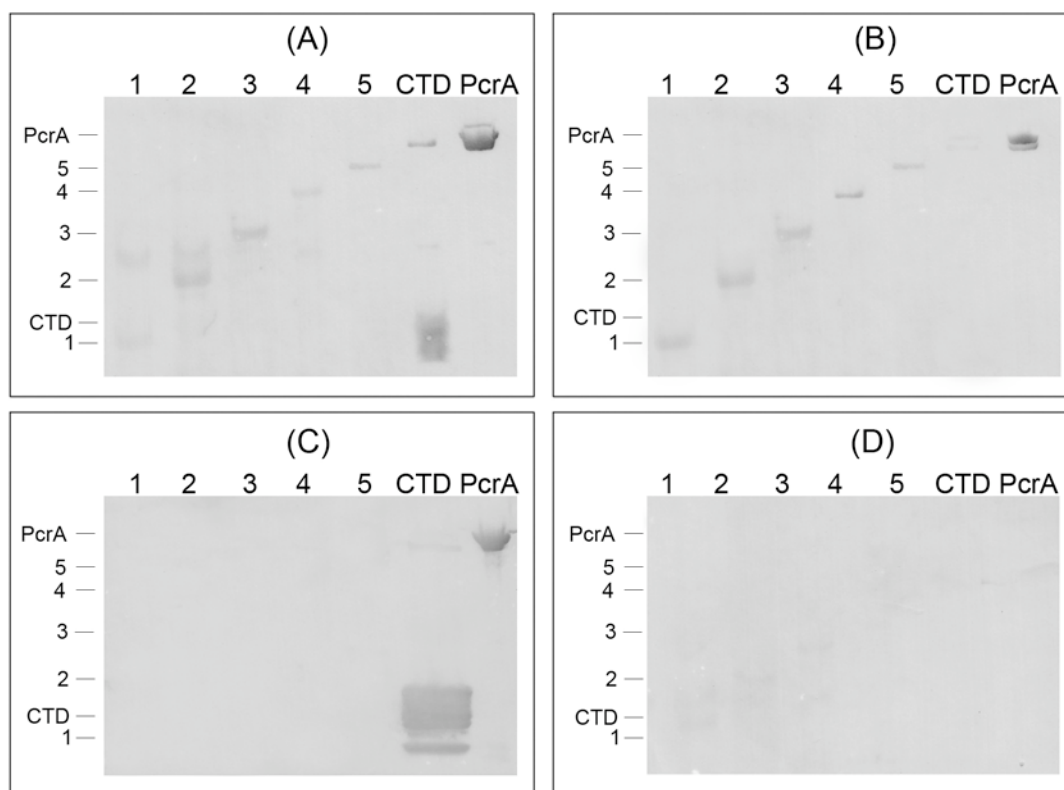
Figure 3.35: Far-Western blot of PcrA$_{Bsub}$ NTD. (A) The membrane was probed with *B. subtilis* RNAP. (B) The membrane was probed with purified β' subunit fragment (1-433 aa). (C) The membrane was probed with purified β subunit fragment (1-400 aa). (D) Negative control – no probe added. 1 – PcrA$_{1-159aa}$, 2 – PcrA$_{1-258aa}$, 3 – PcrA$_{1-344aa}$, 4 – PcrA$_{1-498aa}$, 5 – PcrA$_{1-590aa}$, CTD – purified PcrA$_{Bsub}$ CTD, PcrA – purified PcrA$_{Bsub}$. The position of expected bands was marked on the left, with the numbers corresponding to the lane.

Figure 3.35A shows that RNAP interacts with all of the PcrA$_{Bsub}$ NTD fragments as well as the purified PcrA$_{Bsub}$ CTD and full length PcrA$_{Bsub}$. This result was repeated when the purified fragment of the β'$_{1-433aa}$ subunit was used as the protein probe (Figure 3.35B). This result suggested the interaction between RNAP and PcrA NTD occurs in the first 159 aa, as this region is present in all NTD fragments. No interaction was detected between the PcrA$_{Bsub}$ CTD and the β'$_{1-433aa}$ subunit fragment, as expected (CTD, Figure 3.35B). When the membrane was probed with the purified β$_{1-400aa}$ subunit fragment (Figure 3.35C) the results showed that the β$_{1-400aa}$ subunit interacted with only the CTD and the full length PcrA$_{Bsub}$. Figure 3.35D showed that non-specific interactions between the anti-β' antibody and the PcrA on the membrane did not occur.

This result supported the earlier far-Western blot data using the CTD as the probe and also supports the theory that a single PcrA molecule interacts with RNAP *via* both the N-terminal and C-terminal regions.

3.2.6.3 Modeling the interaction sites on PcrA structure

When examining the crystal structure of PcrA$_{Gst}$ the first 159 aa stretches across two sides of the structure (yellow, Figure 3.36A). It was therefore thought that PcrA could not be interacting with RNAP with this entire region and it was hoped that by creating smaller fragments of PcrA for use in the far-Western blots, the correct region of interaction could be identified. Additional work was performed to separate this region into two smaller fragments, the 1-85 aa fragment (green, Figure 3.36D) and the 85-189 aa fragment (blue, Figure 3.36D) that were on the opposite sides of the PcrA molecule (Figure 3.36C).
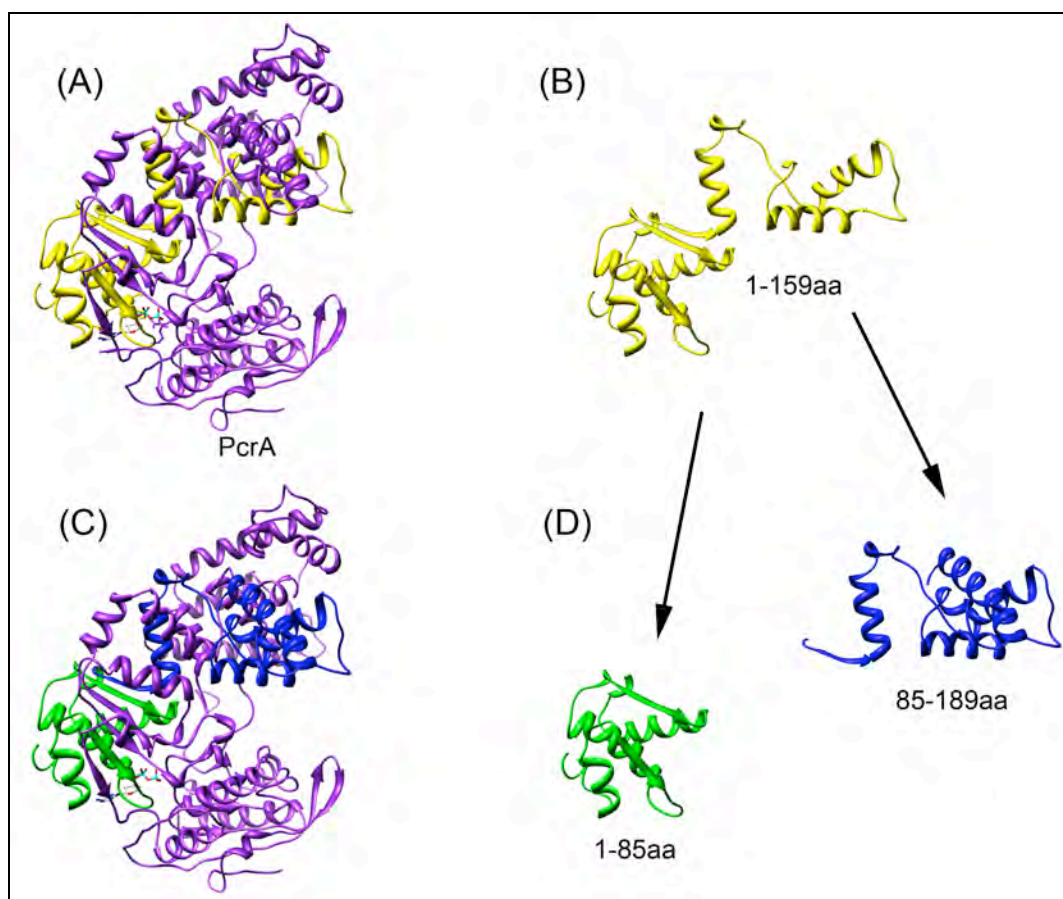
Figure 3.36: PcrA$_{Gst}$ crystal structure. (A) PcrA with the first 159 aa highlighted in yellow. (B) Region 1-159 aa only. (C) PcrA with 1 – 85 aa region highlighted in green and the 85-189 region highlighted in blue (D) The two smaller fragments only.

To see the position of these two N-terminal fragments more clearly, the 1-85 aa (green) and 85-189 aa (blue) regions were again highlighted on the surface rendered PcrA structure (Figure 3.37). The CTD (violet) is also highlighted, along with the DNA (red) in complex. This shows that the CTD and PcrA$_{85-189aa}$ region are both part of the 'crab claw' structure, where DNA enters (Figure 3.37A) and could both be interacting with RNAP at the same time. If you rotated the structure 90° into the page the PcrA$_{1-85aa}$ region is shown to be on the opposite side of PcrA, compared to the PcrA$_{85-189aa}$ region (Figure 3.37B). In this orientation is it clear that both of these regions cannot be binding to RNAP, along with the CTD, which is on the inside of the page. If the molecule is rotated another 90° in the same direction, now both the PcrA$_{1-85aa}$ and the

CTD are visible (Figure 3.37C) and, although they are separated, both regions could be interacting with the β and β' subunits of RNAP. It is important to remember that the C-terminal 72 amino acids were not resolved in the crystal structure so that does provide some flexibility when modeling the interaction of PcrA and RNAP.
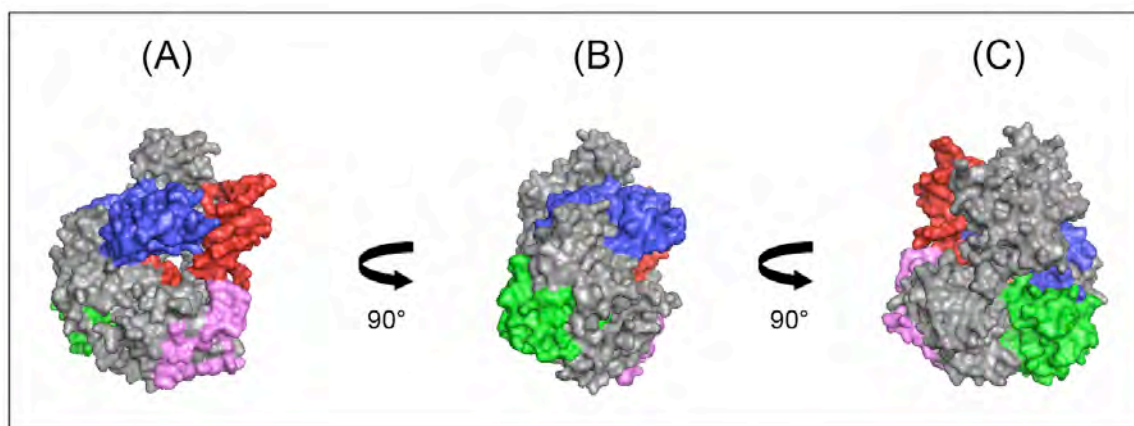
Figure 3.35: Space filled models of the PcrA$_{Gst}$ crystal structure. (A) PcrA in same orientation as in Figure 3.34 (above). (B) PcrA$_{Gst}$ turned 90° into the page. (C) PcrA$_{Gst}$ turned 90° again. 1-85 aa – green, 85-189 aa – blue, CTD - violet and DNA – red.

3.2.6.4 Further analysis of the PcrA NTD-β' subunit interaction

3.2.6.4.1 NTD vector construction and overproduction

Smaller fragments of PcrA (Figure 3.31B) were then cloned into expression vectors. When designing these fragments the crystal structure of the PcrA$_{Gst}$ and protein alignments of the PcrA$_{Gst}$ and PcrA$_{Bsub}$ protein sequences were used to create protein fragments that would have a higher chance of being correctly folded. This would hopefully produce soluble protein fragments. Additionally, the fragments were designed to be cloned into expression vectors containing different affinity tags for purification. This was done as different tags may produce more soluble proteins for purification and the tags could be exploited during the far-Western blot interaction studies. The need for the different affinity tags was not only for ease of purification but also due to the lack of cross reactivity of the anti-PcrA antibody to the N-terminal regions of PcrA.

The expression vectors chosen to clone the smaller sequential fragments of PcrA were pNG698 and pNG651 (Figure 3.36). These fragments were not cloned into pNG209 as the larger fragments were already shown to be insoluble when overproduced using this vector. Both pNG651 and pNG698 are based on pNG209. The C-terminal His_6 tag was removed and replaced with a CHIP-tag (His_{10} + StrepII – pNG698) and a GST-tag (pNG651) (G. Doherty, and X. Yang, University of Newcastle, unpublished). The StrepII tag is an eight-residue tag that binds to an engineered streptavidin, called Strep-Tactin and can be eluted with desthiobiotin (Lichty *et al*, 2005). GST stands for glutathione *S*-transferase and creates protein fusions that can be purified using glutathione agarose (Lichty *et al*, 2005). Both vectors have the same T7 promoter for IPTG inducible protein expression in the appropriate *E. coli* host cells (Figure 3.36).
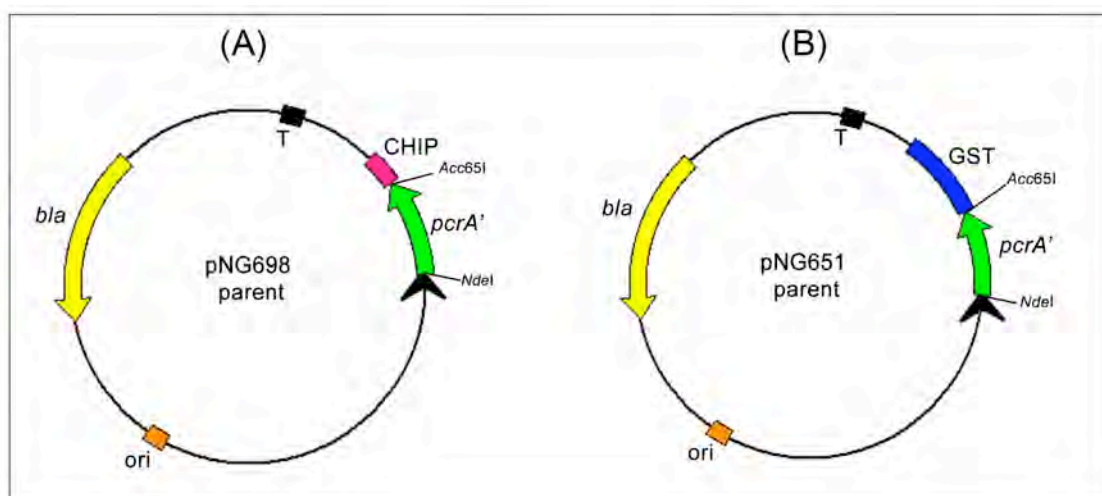
Figure 3.36: PcrA NTD plasmids. All of the smaller NTD fragments were cloned into two parent vectors with differing affinity tags. (A) The parent vector pNG698 with a C-terminal CHIP tag, which is a His_{10} + StrepII tag. (B) The parent vector pNG651 with a 3C-GST C-terminal tag. Features shown are the T7 RNAP promoter (black arrow head) and terminator (T), ampicillin resistance gene (*bla*) and the plasmid origin of replication in *E. coli* (ori).

Four smaller fragments of $PcrA_{Bsub}$ NTD were then cloned into the two parent vectors (Table2.2) and tested for protein overproduction and solubility. Three of the four smaller NTD fragments were successfully cloned into the pNG651 parent, but the 85-189 aa fragment could not be cloned. This parent plasmid, as mentioned earlier, has a

C-terminal GST-tag. GST is a large tag of ~26 kDa, which can sometimes aid the production of soluble protein (Harper & Speicher, 2011). Small scale overproduction and solubility checks (Section 2.5.1.5) were performed on: PcrA$_{GST1-85aa}$ (Figure 3.37A), PcrA$_{GST189-310aa}$ (Figure3.37B) and PcrA$_{GST310-498aa}$ (Figure 3.37C).

Figure 3.37: Protein overproduction and solubility checks of NTD-GST fusion proteins. (A) PcrA$_{GST1-85aa}$. (B) PcrA$_{GST189-310aa}$. (C) PcrA$_{GST310-498aa}$. M – prestained protein ladder (Fermentas). T$_0$ – un-induced sample, WCS – T$_3$ whole cell sample, Sol – T$_3$ soluble fraction, In – T$_3$ insoluble fraction. Numbers on the left indicate marker sizes (kDa).

Induction of the three NTD fragments showed overproduced protein in the T$_3$ whole cell sample and the insoluble fraction, which was not seen in the soluble fraction (Figure 3.37). The PcrA$_{GST189-310aa}$ sample was slightly different because the induced whole cell sample did not show any overproduced protein. Other induction conditions were trialed, including dropping the induction temperature and the IPTG concentration and changing the *E. coli* host strain, but this did not increase the solubility of the protein fragments (not shown). Therefore the NTD fragments could not be purified and would be used in the far-Western blot as the membrane bound proteins. There were fragments of the β$_{1-400aa}$ and β'$_{1-433aa}$ subunits of RNAP that had been purified, which were used as the probe proteins to identify the NTD region of PcrA that was involved in the interaction.

Finally, the PcrA$_{GST}$ fusion proteins were tested using an anti-GST Western blot (Figure 3.38B). A duplicate gel was also Coomassie-blue stained to show the loading of the fragments (Figure 3.38A). It showed the three PcrA$_{GST}$ fragments were successfully overproduced. The Western blot showed the overproduced fragments seen on the Coomassie-blue stained gel were also detected using the anti-GST antibody and therefore were the target PcrA$_{GST}$ fusion protein fragments.
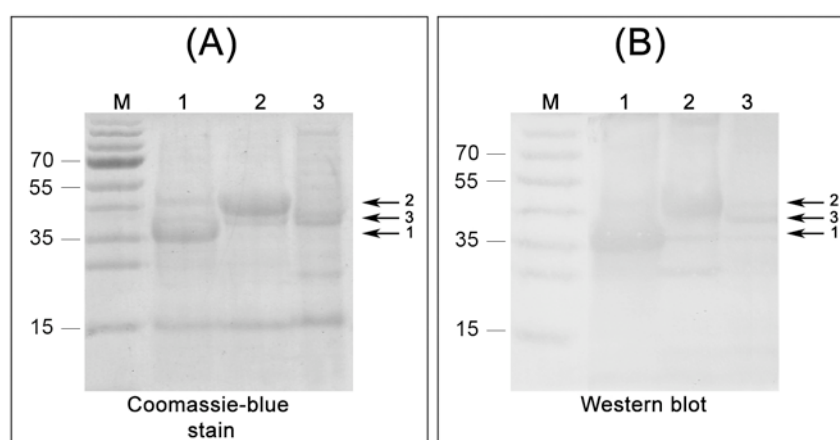
Figure 3.38: (A) Coomassie-blue stain SDS-PAGE gel of the NTD–GST fusion proteins. (B) Western blot of the samples using the anti-GST antibody. M – prestained protein ladder (Fermentas), 1 – PcrA$_{GST1-85aa}$, 2 – PcrA$_{GST189-310aa}$, 3 – PcrA$_{GST310-498aa}$. Numbers on the left indicate marker sizes (kDa).

The NTD fragments were also cloned into pNG698, which meant that either the commercially available anti-Strep or the anti-His antibodies could be used for detection of the proteins. All four PcrA fragments were successfully cloned into the parent vector pNG698 (Table 2.2): pNG834 – PcrA$_{CHIP1-85aa}$, pNG835 – PcrA$_{CHIP85-189aa}$, pNG836 – PcrA$_{CHIP189-310aa}$, and pNG837 – PcrA$_{CHIP310-498aa}$.

Overproduction and solubility checks were then carried out on all four NTD fragments. The CHIP-tag was a much smaller tag than the GST tag and the smaller C-terminal tag had previously produced soluble PcrA CTD protein. The first fragment, PcrA$_{CHIP1-85aa}$ showed the presence of an overproduced protein band in the whole cell sample (wcs)

compared to the un-induced sample (T_0), as indicated by the arrow (Figure 3.39A). Unfortunately the overproduced protein was seen in the insoluble fraction (in) only. This result was repeated for the fourth fragment, PcrA$_{CHIP310-498aa}$ (arrow, Figure 3.39D), where again the overproduced protein was seen in the whole cell sample and the insoluble fraction only. The remaining two fragments PcrA$_{CHIP85-189aa}$ (Figure 3.39B) and PcrA$_{CHIP189-310aa}$ (Figure 3.39C) showed no protein overproduction in the whole cell sample when compared to the T_0 sample.
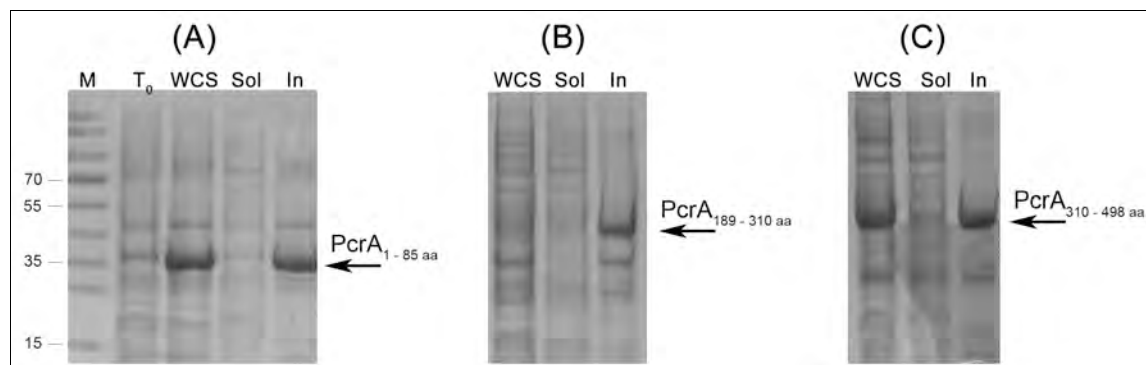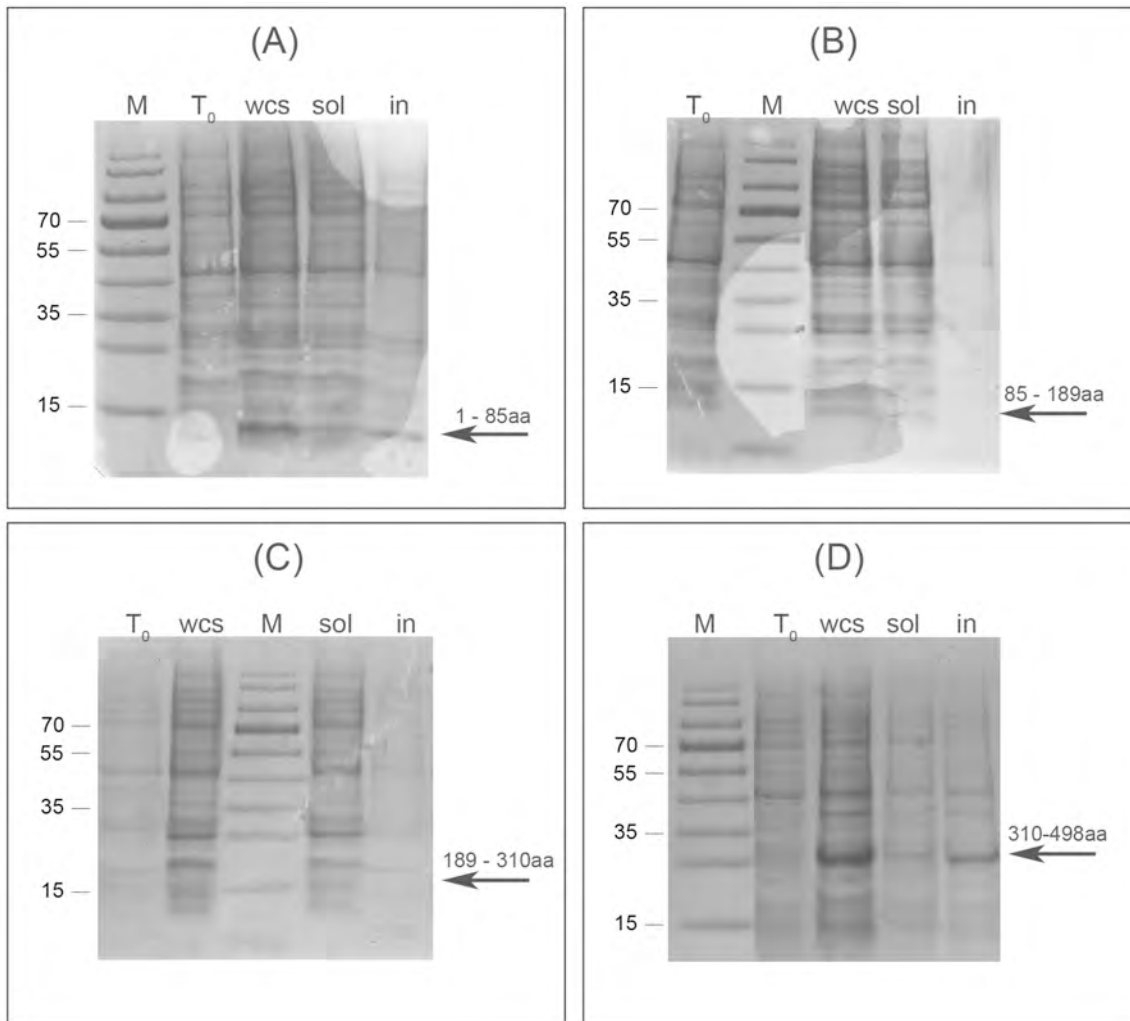

Figure 3.39: Protein overproduction and solubility checks of NTD-CHIP fusion proteins. (A) PcrA$_{1-85aa}$. (B) PcrA$_{CHIP85-189aa}$. (C) PcrA$_{CHIP189-310aa}$. (D) PcrA$_{CHIP310-498aa}$. M - prestained protein ladder (Fermentas), T_0 – un-induced sample, wcs – T_3 whole cell sample, sol – T_3 soluble fraction, in – T_3 insoluble fraction. Numbers on the left indicate marker sizes (kDa).

Due to the unsuccessful overproduction of soluble PcrA fragments from the two parent plasmids, alternative overproduction conditions were trialed. These were also performed on the two fragments, which were not overproduced. Changes in overproduction conditions included using different *E. coli* host strains and IPTG concentrations, ranging from 0.05 – 0.5 mM IPTG. Additionally, the temperature of the induction was lowered from 37 °C to room temperature, which can also assist in the overproduction of soluble proteins. An alternative medium was also used – auto induction medium, which does not require the addition of IPTG. Unfortunately none of the changes made to the induction conditions helped overproduce PcrA fragments (not shown).

All four plasmids were sequenced to confirm the correct DNA fragment was present and contained no mutations. Western blotting was then used to assess the presence of the fragments containing the CHIP-tag. Blotting was carried out using both the anti-His and anti-Strep II antibodies, to confirm the two fragments PcrA$_{CHIP1-85aa}$ and PcrA$_{CHIP310-498aa}$ that were overproduced contained the CHIP-tag and to confirm the remaining fragments (PcrA$_{CHIP85-189aa}$ and PcrA$_{CHIP189-310aa}$) were not being overproduced. The Coomassie-blue stained gel showed the overproduced PcrA fragments were difficult to distinguish from cellular proteins in the four samples, as indicated by the arrows (Figure 3.40A). This was unexpected as previously two fragments: PcrA$_{CHIP1-85aa}$ and PcrA$_{CHIP310-498aa}$ were shown to overproduce the target fragment.
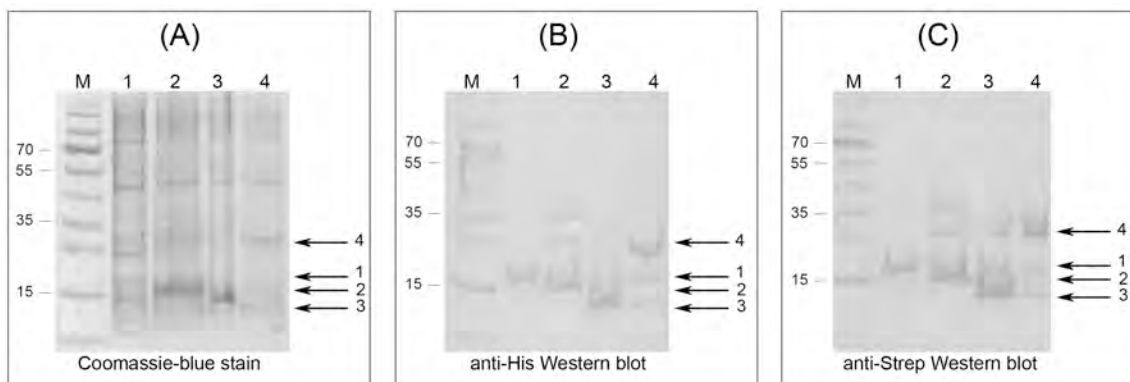
Figure 3.40: (A) Coomassie-blue stained SDS-PAGE gel of NTD-CHIP fusion proteins. (B) Western blot using the anti-His antibody. (C) Western blot the anti-Strep antibody. M - prestained protein ladder (Fermentas), 1 – PcrA$_{1-85aa}$, 2 – PcrA$_{85-189\ aa}$, 3 – PcrA$_{189-310aa}$, 4 – PcrA$_{310-498aa}$. Numbers on the left indicate marker sizes (kDa).

Both the anti-His antibody blot (Figure 3.40B) and the anti-Strep II antibody blot (Figure 3.40C) showed that the fragments are not greatly overproduced. Only PcrA$_{CHIP85-189aa}$ and PcrA$_{CHIP189-310aa}$ (Figure 3.40A, 2 & 3) are visible as an over-represented band on the stained gel, but proteins containing the affinity tag were being expressed in all four samples. It was hoped that these CHIP-fusion proteins could be purified but unfortunately this was not the case, due to both the low level overproduction and the insolubility of the protein fragments (not shown). Instead the proteins were transferred onto nitrocellulose membrane and used in the far-Western blot interaction studies.

3.2.6.4.2 Far-Western blot analysis

The overproduction of the PcrA NTD sub-fragments was greater when the fragments were fused to a large GST tag and therefore these fragments were initially used in the far-Western blot. The 3 GST-fusion proteins: PcrA$_{GST1-85aa}$, PcrA$_{GST189-310aa}$, PcrA$_{GST310-498aa}$, along with full length PcrA$_{Bsub}$ were transferred to the membrane and probed with RNAP and β'$_{1-433aa}$. The Coomassie-blue stained gel (Figure 3.41A) shows all proteins

were present in similar quantities and the negative control blot (Figure 3.41B) was appropriately blank.

The results from probing with *B. subtilis* RNAP (Figure 3.41C) showed an interaction was occurring between RNAP and full length PcrA, as expected. There was an additional interaction identified, between RNAP and $PcrA_{189-310aa}$, as indicated by the arrow. This was interesting because earlier far-Western blotting using the NTD fragments suggested the interaction occurred, at least in part in the first 159 aa of PcrA (Figure 3.35). The results of probing with $β'_{1-433aa}$ also supported the results of the RNAP blot, where an interaction with the $PcrA_{189-310aa}$ region was again detected. The use of the two separate probes, RNAP and the $β'_{1-433aa}$ fragment confirms that the interaction between PcrA does extend past the first 159 amino acids to include some of 189-310 aa region. This result also suggests that the initial result indicating the $PcrA_{1-159aa}$ region is responsible for the interaction now does not include the $PcrA_{1-85aa}$ region and that the further investigation of the $PcrA_{85-159aa}$ region was required.
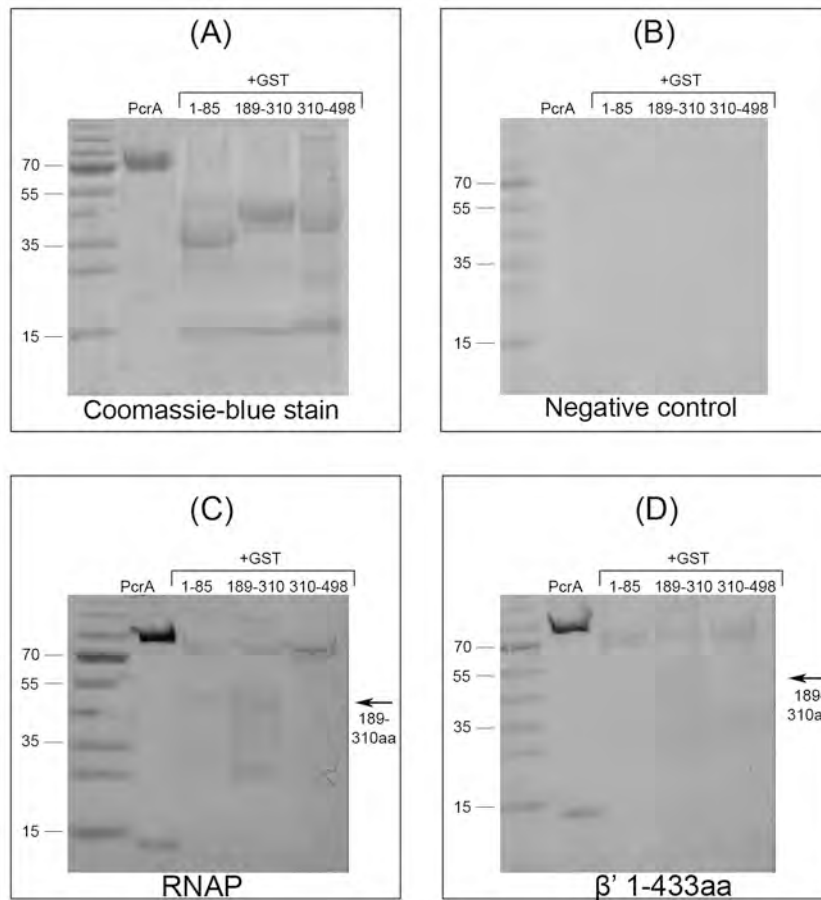
Figure 3.41: Far Western blot of PcrA NTD-GST fusion fragments. (A) Coomassie-blue stained SDS-PAGE gel of PcrA and the NTD–GST fusion proteins. (B) Negative control – no probe added. (C) The membrane was probed with *B. subtilis* RNAP. (D) The membrane was probed with purified $β'_{1-433aa}$. M – prestained protein marker (Fermentas), PcrA – full length $PcrA_{Bsub}$, 1-85 – $PcrA_{GST1-85aa}$, 189-310 – $PcrA_{GST\ 189-310aa}$, 310-498 – $PcrA_{GST310-498aa}$. Numbers on the left indicate marker sizes (kDa).

The $PcrA_{85-189aa}$ region could not be successfully cloned with a GST tag, therefore, it was still possible that RNAP and specifically the β' subunit could be interacting with both $PcrA_{85-189aa}$ and $PcrA_{189-310aa}$ fragments. In order to determine if RNAP interacts with the $PcrA_{85-189aa}$ region the NTD-CHIP fusion proteins were used in the far-Western blot. An interaction detected with $PcrA_{85-189aa}$ would explain the original NTD far-Western blot results, which identified an interaction occurring in the first 159 aa of PcrA.

Again the PcrA-CHIP fusion proteins were transferred onto the membrane and probed with both RNAP and the $\beta'_{1-433aa}$ subunit. The Coomassie-blue stain gel showed that the CHIP-fusion proteins were not greatly overproduced (Figure 3.42A), but previous Western checks confirmed that four fragments are present (Figure 3.40). The negative control membrane was again blank, as expected (Figure 3.42B). The results of the RNAP probe (Figure 3.42C) supported the results of the GST-fusion far-Western blot, RNAP was interacting with $PcrA_{CHIP189-310aa}$. There may also be a very faint interaction with $PcrA_{CHIP85-189aa}$, and possibly with $PcrA_{CHIP1-85aa}$, although there are nonspecific bands present across all membranes that are more evident. For that reason the interaction between RNAP and these fragments cannot be confirmed and additionally the lack of interaction between RNAP and $PcrA_{GST1-85aa}$ would suggest that the results required further investigation before the exact site of interaction is identified.

When the membrane was probed with the $\beta'_{1-433aa}$ fragment (Figure 3.42D), no interactions were detected, not even with full length PcrA. The full length PcrA had been shown to bind to the $\beta'_{1-433aa}$ fragment previously (Figure 3.18) and was therefore used as a positive control in this far-Western blot. The lack of interaction between the $\beta'_{1-433aa}$ fragment and PcrA indicated that the experiment was not successful and the absence of an interaction between the $\beta'_{1-433aa}$ fragment and the PcrA NTD fragments was not an accurate result.
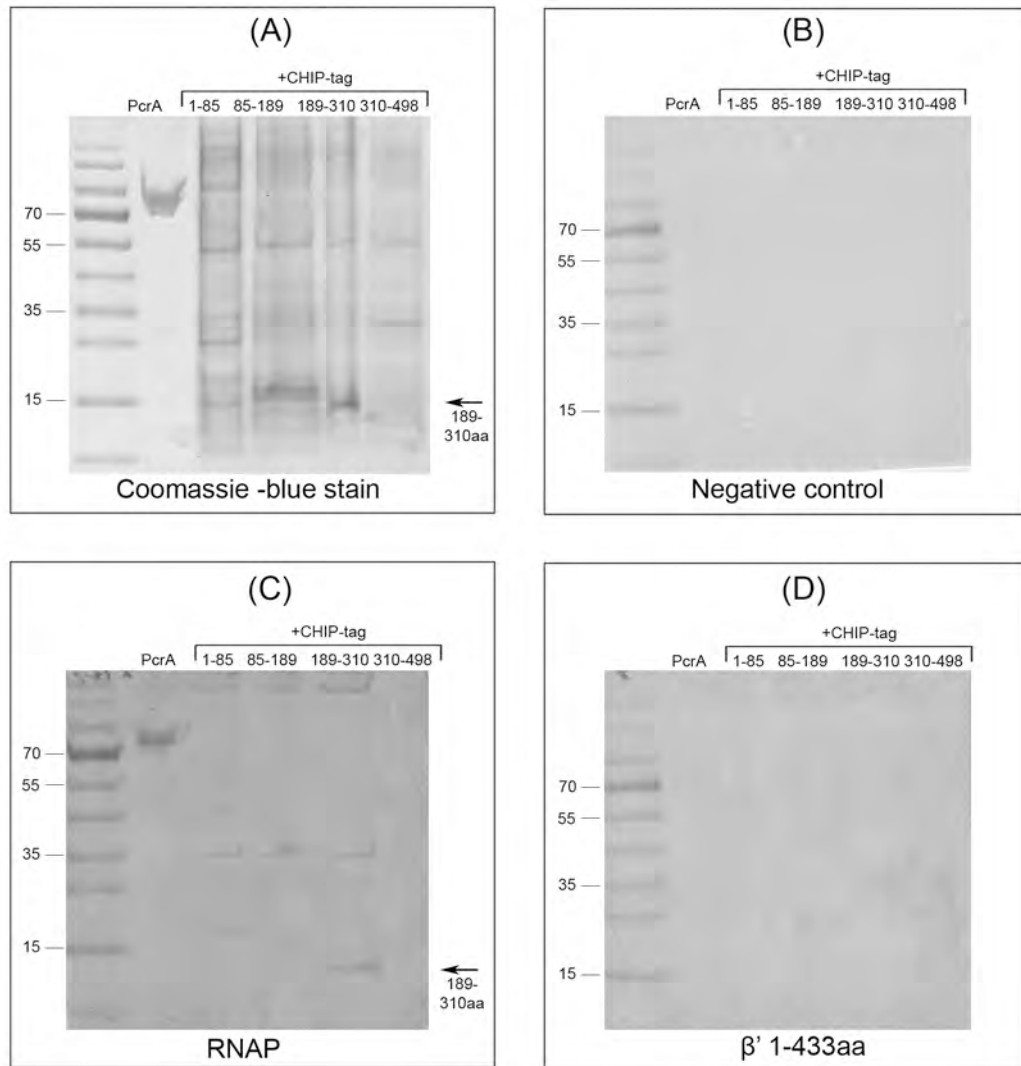
Figure 3.42: Far Western blot of PcrA NTD-CHIP fusion fragments. (A) Coomassie-blue stained SDS-PAGE gel of PcrA and the NTD–CHIP fusion proteins. (B) Negative control – no probe added. (C) The membrane was probed with *B. subtilis* RNAP. (D) The membrane was probed with purified $\beta'_{1\text{-}433aa}$. M – prestained protein marker (Fermentas), PcrA – full length PcrA$_{Bsub}$, 1-85 – PcrA$_{CHIP1\text{-}85aa}$, 85-189 – PcrA$_{CHIP85\text{-}189aa}$, 189-310 – PcrA$_{CHIP189\text{-}310aa}$, 310-498 – PcrA$_{CHIP310\text{-}498aa}$. Numbers on the left indicate marker sizes (kDa).

3.2.7 Further investigation of the PcrA NTD-β' subunit interaction

3.2.7.1 β' subunit sub-fragment – protein overproduction

Another way to identify the specific sites of interaction between the β' subunit and PcrA NTD was using sub-fragments of the β' subunit, which were also overproduced. It was hoped that these smaller RNAP fragments could be used in conjunction with the PcrA NTD fragments to identify specific regions of interaction. This would help identify the position of PcrA bound to RNAP in relation to the direction of transcription and hopefully provide insights in to any possible roles PcrA may have during transcription.

This did not require cloning of the sub-fragments of the β' subunit as previous studies had created plasmids for protein overproduction. These fragments were $\beta'_{1-102aa}$ (pNG750), $\beta'_{103-227aa}$ (pNG751) and $\beta'_{228-310aa}$ (pNG752) (Table 2.2, Kunalan, 2009), which encompassed some of the β' region known to interact with PcrA. Results from the original far-Western blot (Figure 3.16) showed PcrA interacted with $\beta'_{1-433aa}$ (pNG482) and $\beta'_{253-610aa}$ (pNG483). The smaller fragments are highlighted on the structure of RNAP in the face-on orientation (Figure 3.43A) and the side-on orientation (Figure 3.43B). The $\beta'_{1-102aa}$ fragment is in green (Figure 3.43C) and is on the upstream face of RNAP (Figure 3.43A). The $\beta'_{103-227aa}$ fragment is in blue (Figure 3.43D) and on the downstream face of RNAP and finally the $\beta'_{228-310aa}$ fragment is in pink (Figure 3.43E) and found between the green and blue regions on the front and bottom of the β' subunit, inside the claw formed between the β and β subunits (Figure 3.43A).
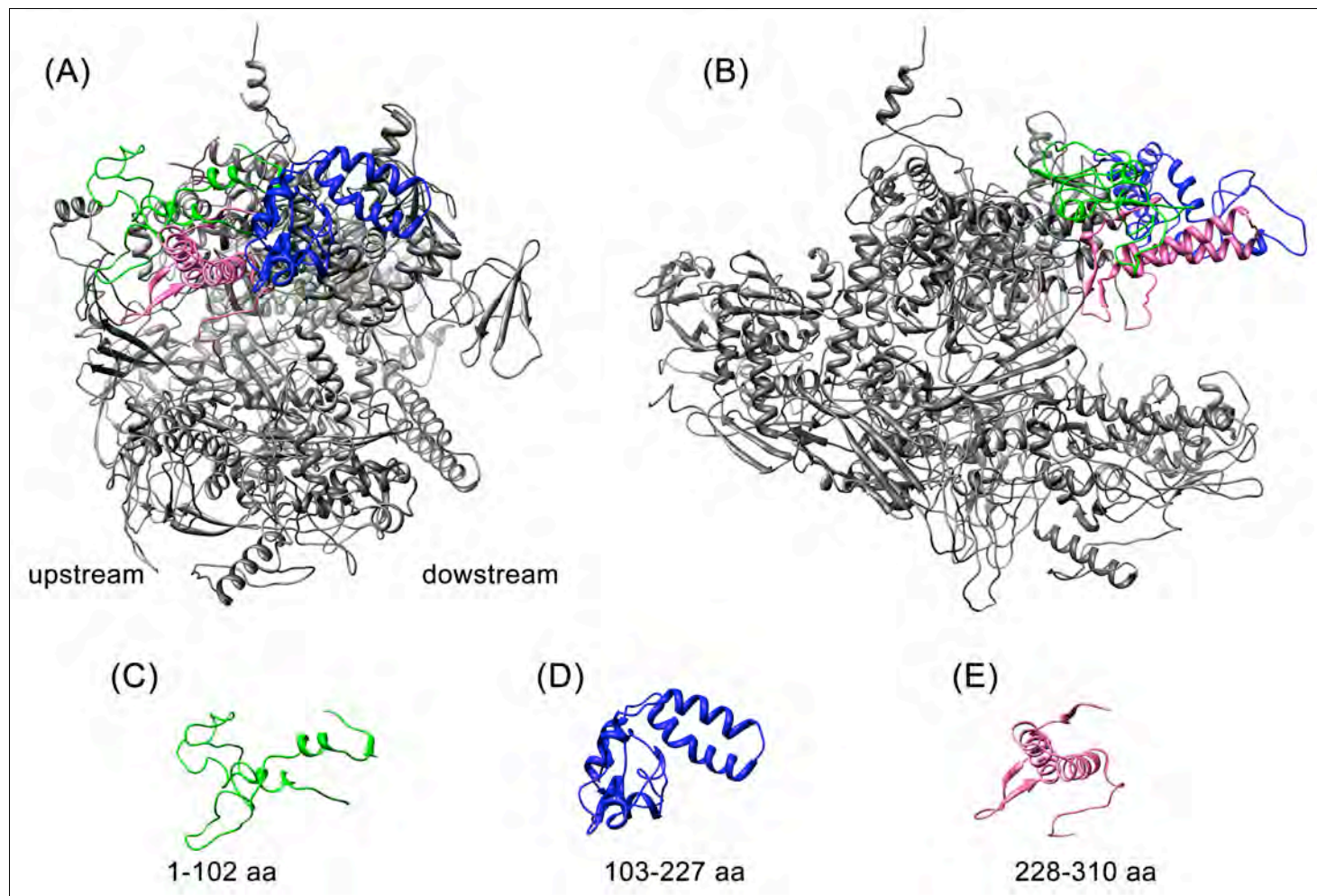
Figure 3.43: RNAP with the β' fragments highlighted. (A) RNAP structure in the face on orientation, β' subunit on the top. (B) RNAP structure turned 90° in side on orientation. The three fragments of the β' subunit are highlighted. 1-102 aa in green. 103-227 aa in blue and 228-310 aa in pink. (C) 1-102 aa region only. (D) 103-227 aa region only (E) 228-310 aa region only.

Small scale protein overproduction checks were performed using $\beta'_{1-102aa}$, $\beta'_{103-227aa}$ and $\beta'_{228-310aa}$ in *E. coli* BL21 (DE3) pLysS (Section 2.5.2.2).
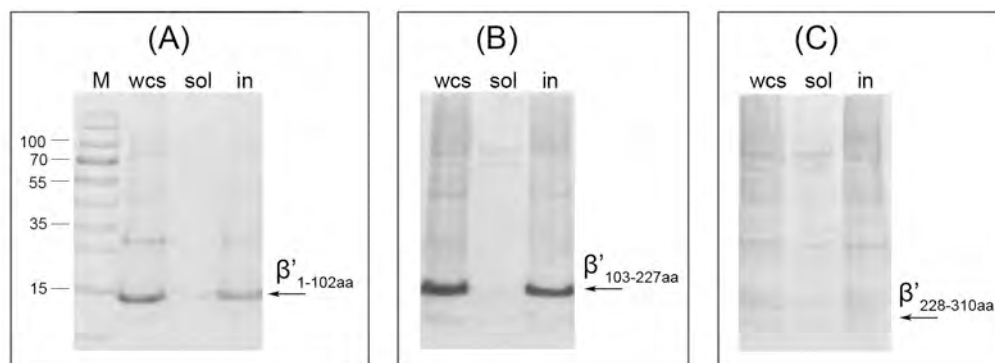
Figure 3.44: Protein overproduction and solubility checks of the β' sub-fragments. (A) $\beta'_{1-102aa}$. (B) $\beta'_{103-227aa}$. (C) $\beta'_{228-310aa}$. M- prestained protein ladder (Fermentas), wcs – whole cell sample, sol – soluble fraction and in – insoluble fraction. Numbers on the left indicate marker sizes (kDa).

The results of the solubility checks showed that both the $\beta'_{1-102aa}$ (Figure 3.44A) and $\beta'_{103-227aa}$ (Figure 3.44B) proteins were overproduced and seen in the whole cell sample (wcs) and in the insoluble fraction (in). The insoluble fraction was used in far-Western blots and probed with PcrA. Unfortunately the $\beta'_{228-310aa}$ fragment was not overproduced at the same level (Figure 3.44C) as the other fragments. To confirm the $\beta'_{228-310aa}$ fragment was overproduced a Western blot was performed using the anti-β' antibody (Section 2.5.3.1). This showed that all four β' fragments were detected, including the $\beta'_{228-310aa}$ fragment (Figure 3.45) and therefore this fragment was included in the far-Western blots.
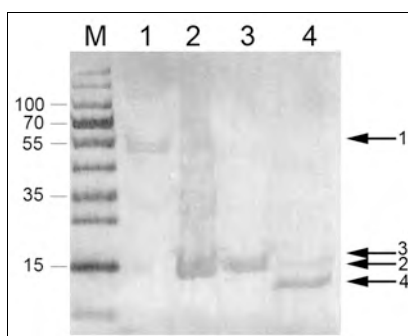
Figure 3.45: Western blot of the overproduced β' fragments. $1 - \beta'_{1-433aa}$, $2 - \beta'_{1-102aa}$, $3 - \beta'_{103-227aa}$, $4 - \beta'_{228-310aa}$. M – prestained protein ladder (Fermentas). Arrows indicate the β' fragments, in corresponding lanes. Numbers on the left indicate marker sizes (kDa).

3.2.7.2 Far-Western blot analysis

The $\beta'_{1-102aa}$, $\beta'_{103-227aa}$ and $\beta'_{228-310aa}$ fragments were transferred onto nitrocellulose membranes and probed with PcrA$_{Bsub}$ and PcrA$_{Gst}$ (Section 2.5.3.2). Unfortunately the membranes could not be probed with PcrA NTD fragments as they could not be purified.
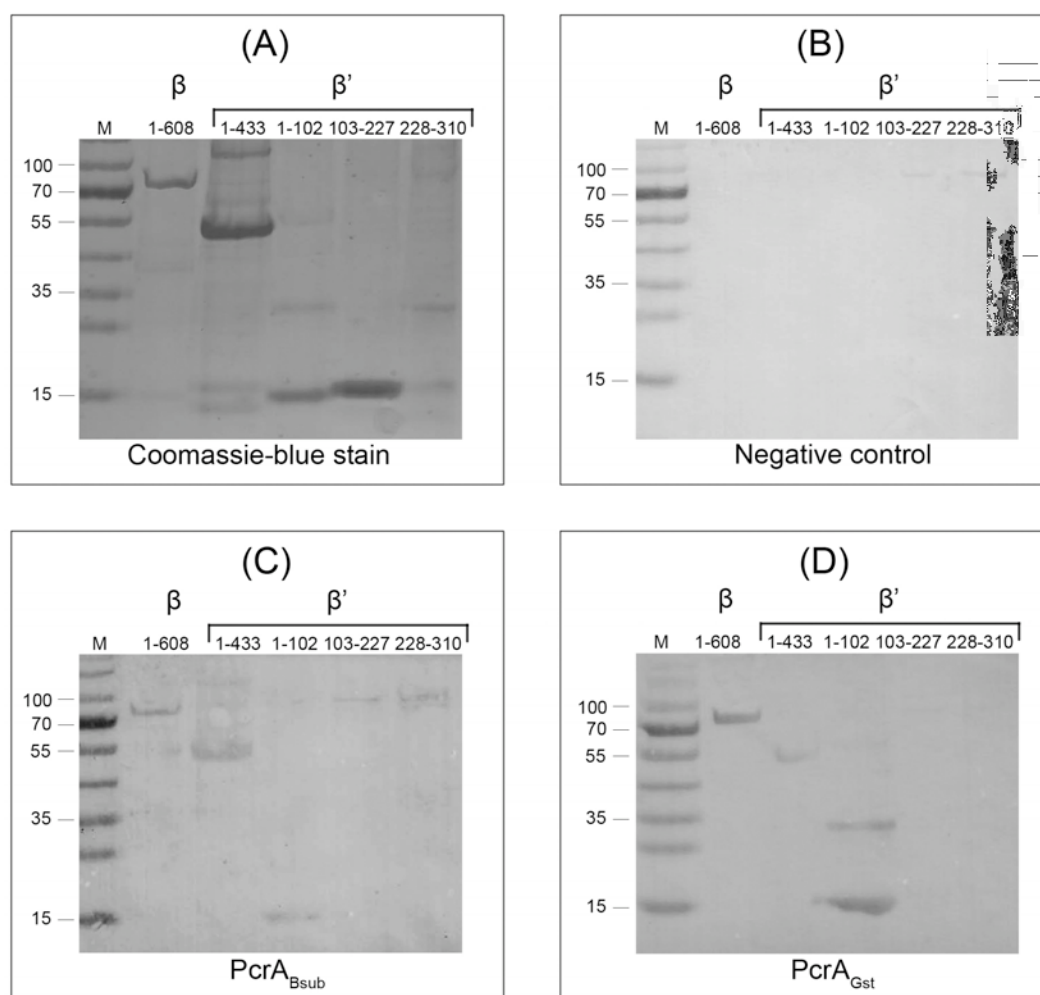
Figure 3.46: Far-Western blot of β' sub-fragments. (A) Coomassie-blue stained SDS PAGE gel. (B) Blank membrane – no protein probe was added. (C) Membrane probed with PcrA$_{Bsub}$. (D) Membrane probed with PcrA$_{Gst}$. M- prestained protein ladder (Fermentas), $\beta_{1-608aa}$, $\beta'_{1-433aa}$, $\beta'_{1-102aa}$, $\beta'_{103-227aa}$, $\beta'_{228-310aa}$.

The results of the far-Western blot indicated PcrA was binding to the first 102 amino acids of the β' subunit (Figure 3.46). Positive control RNAP fragments were used, which were the fragments previously detected in the full length PcrA far-Western blots (Figure 3.16). The β' subunit fragments were probed with both the PcrA$_{Bsub}$ and

PcrA$_{Gst}$ proteins. The PcrA$_{Bsub}$ probe detected an interaction occurring between PcrA and β$_{1-608aa}$, β'$_{1-433aa}$ and β'$_{1-102aa}$ samples (Figure 3.46C). PcrA$_{Gst}$ also showed an interaction occurred with the same β$_{1-608aa}$, β'$_{1-433aa}$ and β'$_{1-102aa}$ samples (Figure 3.46D). When mapping the smaller site of β' PcrA interaction on to the structure of RNAP it indicated that the interaction would most likely be occurring on the upstream face of RNAP.

3.3 Discussion

3.3.1 Modeling the interaction

The interaction between PcrA and RNAP was confirmed in this body of work. The use of *in vitro* interaction studies has shown that the interaction between these two proteins is a complex one, with multiple binding sites. The affinity chromatography showed the interaction was a strong interaction, which was only disrupted with the addition of 1 M NaCl (Figure 3.16). Many protein interactions are broken at lower NaCl concentrations, in fact the interaction between RNAP and known transcription factor GreA was disrupted after the addition of 0.4 M NaCl.

RNAP is a large multi-subunit enzyme and therefore it was important to determine the position of the PcrA interaction, which would help to determine the reason for the strong interaction between the two proteins. The far-Western blot was used to determine the specific regions of the RNAP β and β' subunits that interact with PcrA. The results showed that there were multiple binding sites between RNAP and PcrA. In fact the N-terminal regions of both the β and β' subunits interact with PcrA. The

multiple sites of interaction could be a reason the strong interaction could not be broken with 0.4 M NaCl.

The region of interaction was further specified using additional far-Western blots and the final model (Figure 3.47) involves the β subunit (light blue, 1-400 aa) binding to PcrA CTD (violet, 577-739 aa) and the β' subunit (1-102 aa, green and possibly 228-310 aa, pink) binding to PcrA (189-310 aa, yellow and possibly 85-189 aa, blue). The regions of PcrA involved in the interaction are shown in Figure 3.47A and B. Figure 3.47B shows that these three fragments of PcrA are all one face of PcrA and could therefore be interacting with RNAP together. The regions of RNAP involved in the interaction are shown in Figure 3.47C and D. More work is needed to determine a smaller region of the β subunit that is responsible for the interaction with PcrA.
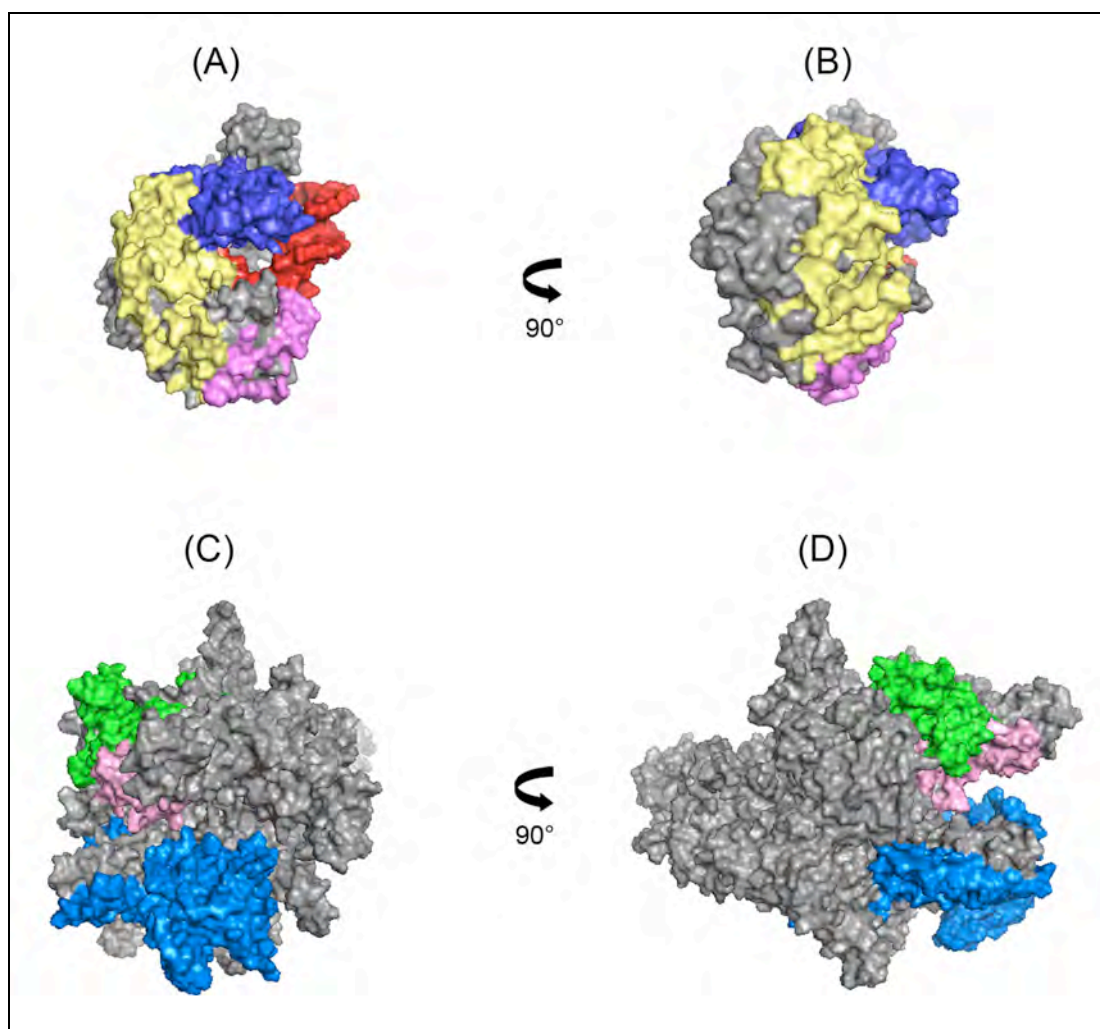
Figure 3.47: Interacting regions of PcrA and RNAP. (A) PcrA, 85-189 aa – dark blue, 189-310 aa yellow, unstructured CTD – violet and DNA – red. (B) PcrA turned 90° into the page. Shows all three fragments are on same face of PcrA. (C) RNAP, β subunit 1-400 aa - light blue, β' subunit 1-102 aa – green and β' subunit 228-310 aa – pink. (D) RNAP turned 90° into the page, so the upstream face of RNAP is in profile. Note: molecules are not to scale.

There is some speculation on the sites of interaction, particularly with the $\beta'_{228\text{-}310aa}$ and PcrA$_{85\text{-}189aa}$. Far-Western blotting did not show an interaction between the $\beta'_{228\text{-}310aa}$ subunit fragment and PcrA, but it was possible that the low overproduction of the $\beta'_{228\text{-}310aa}$ fragment was responsible for not detecting the interaction, rather than the $\beta'_{228\text{-}310aa}$ region not being involved in the interaction. The reason for implicating the $\beta'_{228\text{-}310aa}$ fragment in the interaction with PcrA was due to the $\beta'_{253\text{-}615aa}$ fragment interacting with PcrA, along with the $\beta'_{1\text{-}433aa}$ fragment in the initial far-Western blot (Figure 3.18). It was thought that the overlapping region of the two fragments (253-433 aa) may be the

region of interaction, which encompasses the $β'_{228-310aa}$ fragment. To conclusively prove this region of the $β'$ subunit interacts with PcrA, further work on this fragment would be needed. $PcrA_{85-189aa}$ has been similarly implicated in the interaction with the $β'$ subunit of RNAP. When increasing-sized fragments of the PcrA NTD were investigated, far-Western blotting indicated that all the fragments were interacting with the $β'$ subunit (Figure 3.35). The region of PcrA that was common to all fragments was 1-159 aa, which was then broken down into two smaller fragments (1-85 aa and 85-189 aa) for further studies. Results using the fragments fused to a GST affinity tag, showed $PcrA_{1-85aa}$ did not interact with the $β'$ subunit (Figure 3.41) and the fragments fused to the CHIP affinity tag showed a slight interaction with both $PcrA_{85-189aa}$ and $PcrA_{1-85aa}$ (Figure 3.42).

There were some issues within this experimental work, which prevented finding a conclusive answer to the regions of both PcrA and RNAP involved in the interaction. Firstly there were difficulties in overproducing and purifying many of the proteins required for the *in vitro* interaction studies. To overcome this, alternative proteins were utilised, such as the substitution of *G. stearothermophilus* PcrA for *B. subtilis* PcrA. These proteins have very high levels of identity so didn't adversely affect the interaction studies. In fact the use of the $PcrA_{Gst}$ with the *B. subtilis* RNAP in the affinity chromatography was the first indication that the interaction between PcrA and RNAP may be conserved in other bacterial species. There were additional problems with using fragments of the *B. subtilis* PcrA protein, specifically the NTD fragments. The fragments were not soluble and did not react with the anti-PcrA antibody. The lack of reactivity with the antibody was particularly disappointing as the polyclonal anti-PcrA antibody was raised against the purified full length protein. This made the *in vitro*

interaction studies more difficult to manipulate, as both required the use of antibodies and Western blotting for detection of the interaction. Without using the anti-PcrA antibody, antibodies for fusion affinity tags must be used, which obviously means the affinity tag must not be present on the second protein.

3.3.2 Extending the interaction into Gram-negative bacteria

After showing PcrA$_{Gst}$ interacts with the *B. subtilis* RNAP, the investigation of the interaction was extended further to include the Gram-negative bacterium, *E. coli*. Initial affinity chromatography confirmed that the interaction between PcrA$_{Gst}$ and *E. coli* RNAP was conserved. As previously mentioned this was surprising as PcrA is found in Gram-positive bacteria only. This result led to the investigation of the PcrA homologue – UvrD. PcrA and UvrD are both non-replicative helicases, the roles of which have not yet been fully defined *in vivo.* Far-Western blotting showed that UvrD interacts with the same region of *B. subtilis* RNAP as PcrA. Unfortunately the interaction between UvrD and *E. coli* RNAP could not be confirmed using affinity chromatography, due to solubility issues. This was not a concern as the conserved nature of the interaction between the Gram-positive and Gram-negative bacteria suggested strongly that an interaction occurs between UvrD and RNAP in *E. coli*.

The presence of a conserved interaction between RNAP and similar SFI helicases in Gram-positive and Gram-negative bacteria would suggest that this interaction may provide a role of a DNA helicase during transcription *in vivo*, but at this stage the role of PcrA (and UvrD) is unknown. It was hoped that further investigation of this interaction

would provide additional information that could elucidate the possible roles of PcrA during transcription

3.4 Conclusions

This work has shown there is a strong interaction occurring between PcrA and RNAP, which has multiple binding sites between the two proteins. The interaction identified was stronger than the interaction between RNAP and a known transcription factor. Although the majority of the work was carried out using the proteins of *B. subtilis* the use of PcrA$_{Gst}$ showed that the interaction is conserved in other Gram-positive bacteria. Further investigation also identified an interaction between UvrD and RNAP, which suggested that the interaction between like SF1 helicase and RNAP was also conserved in Gram-negative bacteria. The extent of the interaction in the different bacterial species and the strong, complex interaction suggests that the interaction could have a role during transcription. In order to identify the functional importance of this interaction further investigation was required and mutagenesis was then used to provide additional information on the specific site of interaction (Chapter 4).

Chapter 4:

Mutagenic analysis of the

PcrA CTD–β subunit

interaction

Chapter 4: Mutagenic analysis of the PcrA CTD – β subunit interaction

4.1 Introduction

4.1.1 Mutagenic yeast two-hybrid assay

Following the initial yeast two-hybrid investigation that identified the interaction between PcrA and the β subunit of RNAP, a mutagenic yeast two-hybrid screen was performed, as a continuation of this study. The mutagenic screen used a fragment of the *B. subtilis* RNAP β subunit (amino acids 250-407) as the bait and the prey proteins were randomly mutated PcrA CTD (612-739 aa) (M-F. Noirot-Gros, INRA, France). The results indicated the interaction could be disrupted by single amino acid mutations and the screen identified 6 amino acid changes that abolished the interaction between PcrA and the β subunit, all of which occurred at the C-terminus of PcrA (692-739 aa).

There is currently no structural information for this region of PcrA, as the crystal structure of *G. stearothermophilus* PcrA was resolved for the first 652 amino acids only (Subramanya *et al*, 1996). This makes it difficult to tell if these amino acids are surface residues and the changes represent a genuine disruption of an amino acid interaction between PcrA and the β subunit, or if the mutation causes a change in the protein conformation, which in turn disrupts the interaction.

The 6 amino acid changes and their positions are highlighted in the PcrA alignment (Figure 4.1), which shows they are conserved throughout other Gram-positive bacteria. The isoleucine at amino acid position 718 is conserved through a range of Gram-positive bacteria, while the other 5 mutated residues are not completely conserved. Interestingly, the alanine at position 734 which was mutated to valine is already present as valine in the PcrA sequence of *Corynebacterium diptheriae* (Figure 4.1). This mutation was still investigated as *C. diptheriae* is only distantly related to *B. subtilis* and the alanine is highly conserved at this position in the other bacteria. It is also possible that the valine in *C. diptheriae* PcrA is the result of a sequencing error.
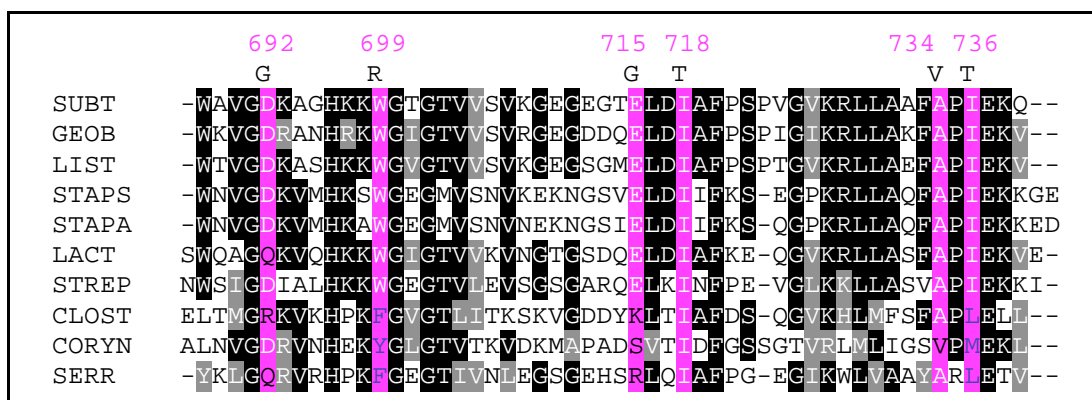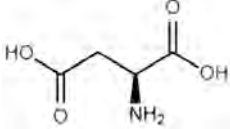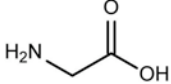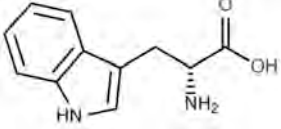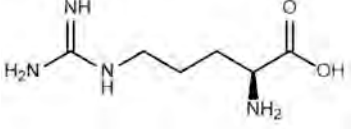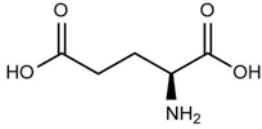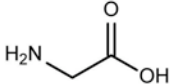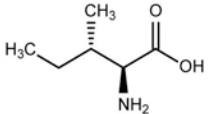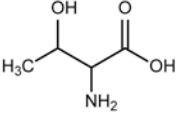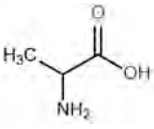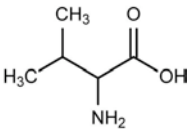
Figure 4.1: Amino acid alignment of the C-terminus of PcrA. *B. subtilis* is the first sequence; the amino acid which was mutated is highlighted in pink. The amino acid change and the position of the mutations are seen above each mutation. (SUBT – *Bacillus subtilis,* GEOB – *Geobacillus stearothermophilus* LIST – *Listeria innocus,* STAPS – *Staphylcoccus saprophyticus,* STAPA – *Staphylcoccus aureus,* LACT - *Lactobacillus sakei,* STREP – *Streptococcus pneumoniae,* CLOST – *Clostridium perfringes,* CORYN – *Corynebacterium diphtheriae,* SERR - *Serratia marcescens.*)

The amino acid changes identified in the mutagenic screen are shown in detail in Table 4.1. The mutations have introduced a size or charge change in the amino acid, which may influence the interaction with the β subunit. For example the mutation at position 692 was an aspartic acid (D) replaced with glycine (G), which replaces the acidic residue for a non-polar residue. A tryptophan (W) was replaced with arginine (R) at position 699, which replaces a large hydrophobic residue with a basic residue. The

glutamic acid (E) to glycine (G) at position 715 also removes the acidic residue, which is replaced by a smaller non-polar amino acid.   The isoleucine (I) to threonine (T) mutation was identified twice, at positions 718 and 736.   This change removes a hydrophobic residue and replaces it with the polar threonine residue (Table 4.1).   It is important to consider the changes to the individual amino acids, as the position of the mutations in relation to the PcrA surface structure is unknown.

Table 4.1: A representation of the amino acid changes in the PcrA mutants. The amino acid changes were identified using yeast two-hybrid mutagenic screening.

AA	Wild type	Mutation
692	Aspartic acid (D) Acidic - pKa = 3.9	Glycine (G) Small, non polar
699	Tryptophan (W) Aromatic, non polar	Arginine (R) Basic –pKa = 12.48
715	Glutamic acid (E) Acidic – pKa = 4.07	Glycine (G) Small, non polar
718 & 736	Isoluecine (I) Hydrophobic, non polar	Threonine (T) Nuceolophilic, polar
734	Alanine (A) Small, non polar	Valine (V) Hydrophobic

Initially the mutations of PcrA were introduced into the *B. subtilis* chromosome with the aim of assessing the effect of disrupting the CTD-β subunit interaction. There were a few issues to overcome when designing this set of experiments. Firstly, PcrA is an essential protein in *B. subtilis,* with a role in the RecFOR pathway. Similarily, if the interaction between PcrA and the β subunit is essential, introducing the mutations into the *B. subtilis* chromosome would produce a lethal phenotype. This would mean further *in vivo* studies on the role of the mutations would be extremely challenging. Secondly, PcrA is known to bind to other proteins within the cell. When performing the yeast

two-hybrid, it was found that the changes also disrupted the interaction between PcrA and UvrB, UvrC and YxaL, in addition to the β subunit. This suggests the mutations may cause a conformational change in the CTD, or that the same CTD region of PcrA is able to interact with several different binding partners. A common binding site for multiple proteins as been shown previously; an example is the single stranded DNA binding (SSB) protein, where the CTD interacts with several DNA replication and repair proteins (Shereda *et al*, 2008; Costes *et al*, 2010). The disruption of these additional interactions would have to be considered when interpreting the results of the *in vivo* studies.

4.1.1.1 Other protein partners

PcrA interacts with several other proteins, from a range of pathways within the cell. These include UvrB and UvrC, which are part of the nucleotide excision repair (NER) pathway. NER is a process of repairing UV-induced DNA lesions (reviewed by Van Houten *et al*, 2005). The three main components of the NER pathway are the proteins UvrA, UvrB and UvrC. Each of these proteins has its own role within the pathway. Initially, UvrA and UvrB form a complex which locates DNA damage. Following the location of the damage, UvrB binds to the DNA lesion and UvrA is released. UvrC is then recruited to the complex and introduces incisions into the DNA positioned either side of the lesions. UvrD then removes the damaged DNA and Pol I fills in the gap. Finally, ligase rejoins the nicked DNA (reviewed by Goosen & Moolenaar, 2001; reviewed by Van Houten *et al*, 2005; Malta *et al*, 2006). If the disruption of the PcrA interaction with UvrB and UvrC affects the NER pathway, the resulting increase of DNA lesions on the *B. subtilis* chromosome may produce a phenotypic change. Therefore any changes seen after introducing the PcrA mutations would have to be

investigated to determine if they are a result of disrupting the interaction with RNAP or the UvrBC proteins. In order to determine the effect of disrupting the PcrA-UvrBC interaction specifically the PcrA mutations were introduced into a UvrB mutant strain. Strain HCR-9 has a point mutation in UvrB, which prevents the protein from functioning (Okubo & Romig, 1966). The results of these transformations would indicate that any phenotypic changes common to both parent strains were not a result of disrupting the NER pathway. Changes seen only in the wild type strain would suggest the PcrA-UvrBC interaction was responsible for the phenotype.

The third protein-protein interaction that was disrupted when PcrA was mutated was the PcrA-YxaL interaction. YxaL is a protein with no known function, but its amino acid sequence shows similarity to a family of proteins that have a repeated domain in their sequence, called a β-propeller (Fulop & Jones, 1999). β-propeller proteins have a structure consisting of repeated β-sheets, arranged around a central axis, like the blades of a propeller. These proteins have a high diversity in both sequence and function and are found in a wide variety of species (Tachino *et al*, 2007). YxaL has been shown to enhance the processivity of PcrA in helicase assays, which allows PcrA to translocate greater lengths of DNA. Additionally, YxaL did not alter the ATPase or DNA-binding properties of PcrA (Noirot-Gros *et al*, 2002b). It has been shown that there was no readily observable phenotype in the *yxaL* knockout strain by replacing the *yxaL* gene with the spectinomycin resistance gene (Noirot-Gros *et al*, 2002b).

4.1.2 Integration of *pcrA* mutations into the *B. subtilis* chromosome

To introduce the *pcrA* mutations into the *B. subtilis* chromosome an integration vector pSG1164 was used (Lewis & Marston, 1999). This vector integrates the mutations into the *pcrA* gene in the chromosome *via* single crossover (Figure 4.2). The recombination event between the chromosomal DNA and the plasmid DNA occurs within the region of homology (500 bp of '*pcrA*), so the mutated *pcrA* was under the control of the wild type promoter. Any genes downstream of *pcrA* would then be under the control of the plasmid xylose promoter (Figure 4.2B). The gene downstream from *pcrA* in *B. subtilis* 168 is *ligA* (encodes a DNA ligase), which is essential and therefore the resultant strains require the presence of xylose for growth.
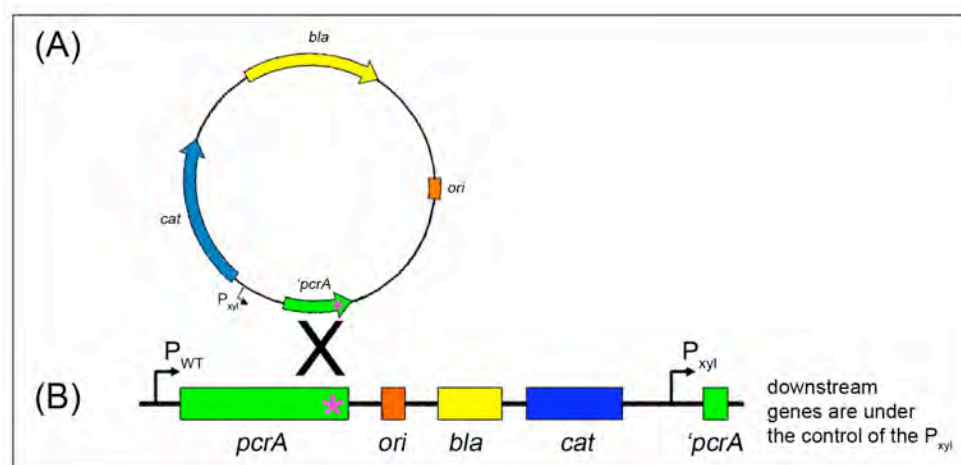
Figure 4.2: Integration of the *pcrA* mutations into the *B. subtilis* chromosome. (A) Plasmid containing the *pcrA* mutations for integration into the chromosome. (B) Final chromosomal construct – after recombination, mutant *pcrA* is under control of wild type promoter (P_{WT}) and downstream genes are under the control of the xylose promoter (P_{xyl}). * - mutation in *pcrA*.

4.1.3 *In vitro* interaction studies

In conjunction with the *in vivo* analysis of the effects of the PcrA mutations, *in vitro* mutagenesis interaction studies were also carried out to investigate the interaction

directly. The *in vitro* assays required cloning of the mutated gene fragments into expression vectors for the overproduction and purification of the proteins. The purified proteins were then used in binding assays, as in Chapter 3.

4.1.4 Aims

This work aimed to investigate the specific amino acids involved in the PcrA CTD-β subunit interaction using mutagenesis. The effects of the mutations were examined using both *in vitro* and *in vivo* interaction assays.

4.2 Results

4.2.1 *In vitro* mutagenic interaction studies

4.2.1.1 PcrA C-terminal domain (CTD) mutants

Four mutations were introduced to into the CTD overproduction vector (pNG574; Table 2.2) using site directed PCR mutagenesis (Section 2.3.1.3). As previously mentioned (Chapter 3, Section 3.2.6.1) the wild type CTD region when purified is very unstable and is rapidly degraded. This was shown to be due to protein instability and not contaminants *via* Western blotting with PcrA antibody (Figure 3.22). This protein degradation also occurred after the purification of the mutant CTD proteins (Figure 4.3).
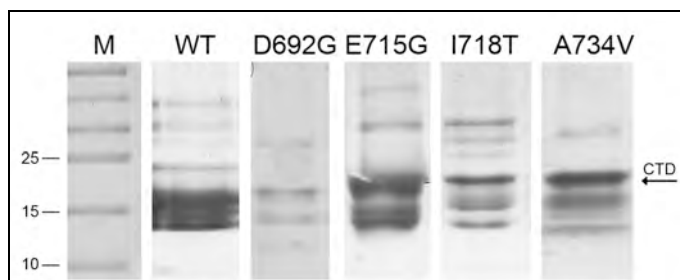
Figure 4.3: Purified *B subtilis* PcrA mutant CTDs. M – prestained protein ladder (Fermentas), WT – wild type CTD , D692W mutation, E715G mutation, I718T mutation, A734V mutation, I736T mutation.

The purified mutant proteins, although degraded, were still used in the far-Western blot. This was for two reasons: firstly, degraded wild type CTD had been successfully used in the far-Western blot (Section 3.2.6.2) and secondly, PcrA and fragments of PcrA had been difficult to overproduce and purify. Other attempts were made to purify the CTD, including: using an N-terminal His_6 tag, which did not generate any protein overproduction (Harriott, 2006) and an N-terminal GST tag, which couldn't be cleaved after protein overproduction and purification (Withers 2007). Therefore, the current overproduction levels and protein purification were optimal and it was hoped that the level of degradation would not impact greatly in the results of the far-Western blots.

For the far-Western blots, the $\beta_{1-608aa}$ subunit fragment was transferred onto nitrocellulose membrane and re-natured (Section 2.5.3.2). The membranes were then probed with equal amounts of PcrA CTD proteins and the interaction was detected using the anti-PcrA antibody (Section 2.5.3.3).
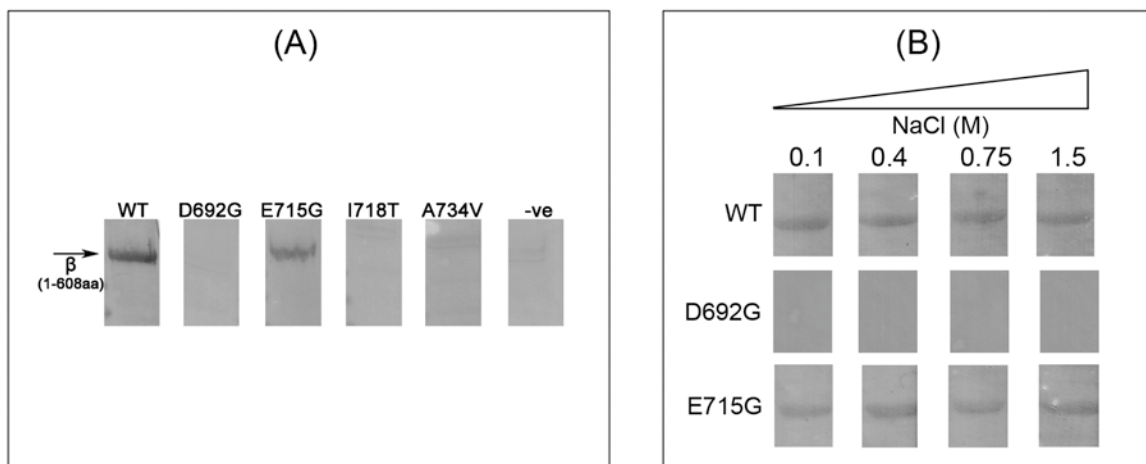
Figure 4.4: Far-Western blot between the β subunit (1-608aa) and CTD PcrA mutants. (A) Four CTD mutant protein were used as probe, as well as wild type CTD, as the positive control. The wild type and the 715 mutation bound to the β subunit (1-608aa). (B) Two mutations, E715G (binds β 1-608aa) and D692G (disrupted binding) were tested, along with wild type, with increasing NaCl washes. WT – the wild type CTD. D692G – CTD with D692G mutation. E715G – CTD with E715G mutation, 718T – CTD with I718T mutation and A734V – CTD with A734V mutation.

The results of the far-Western blot showed the wild type protein (WT, Figure 4.4A) binds to the $\beta_{1-608aa}$ fragment as expected. An interesting result was that mutant protein CTD$_{E715G}$ also bound to the $\beta_{1-608aa}$ fragment. The remaining mutant CTD proteins did not bind the fragment, as expected from the yeast two-hybrid results. Although the far-Western blot is not quantitative, it was thought that the strength of the interaction between the two protein fragments could be qualitatively tested by increasing the NaCl concentration of the membrane washes. Therefore, the far-Western blots were repeated, this time using only the CTD$_{WT}$, CTD$_{E715G}$ and CTD$_{D692G}$ proteins. After probing with the proteins, but before addition of the anti-PcrA antibody, the membranes were subjected to a series of NaCl washes (0.1-1.5 M), shown in Figure 4.4B. The wild type PcrA protein has previously been shown to bind tightly to RNAP using affinity chromatography (Chapter 3, Figure 3.14) so it was expected that the interaction would survive the first three NaCl (0.1 M, 0.4 M and 0.75 M) washes in the far-Western blot. The 1.5 M NaCl wash was expected to break the interaction because 1 M NaCl was used to disrupt the interaction in the affinity chromatography (Section 3.2.2.1). The

results showed the CTD_{WT} remained bound to the β subunit, following the 1.5M washes. This indicates that the interaction formed in the far-Western blot, required a higher NaCl concentration or a longer incubation time to disrupt the interaction. The CTD_{D692G} protein was used as a negative control, as it had already been shown to have no binding to the β fragment and again an interaction was not detected (Figure 4.4B). The mutant CTD_{E715G} protein gave the same binding pattern as the CTD_{WT} protein (Figure 4.4B), and washing the membrane with 1.5M NaCl failed to break the protein interaction. This result strongly suggested that the CTD_{E715G} mutant protein does not have a decreased binding affinity for the $β_{1-608aa}$ fragment and acts similar to the CTD_{WT} protein.

4.2.1.2 Full length PcrA mutants

Two mutations were then introduced into the full length *pcrA* plasmid (pSMG69, Table 2.2) for purification; these were D692G and E715G (pNG832 and pNG824; Table 2.2). From Figure 4.4A it was shown that the D692G mutation in the purified CTD disrupted the β subunit interaction, while the E715G mutation did not. The next step was to use the far-Western blot to investigate the effect on the RNAP binding of full length PcrA mutant proteins. The mutations were introduced using PCR site-directed mutagenesis (Section 2.3.1.3) and purified using the four-step purification protocol (Section 2.5.4.1)
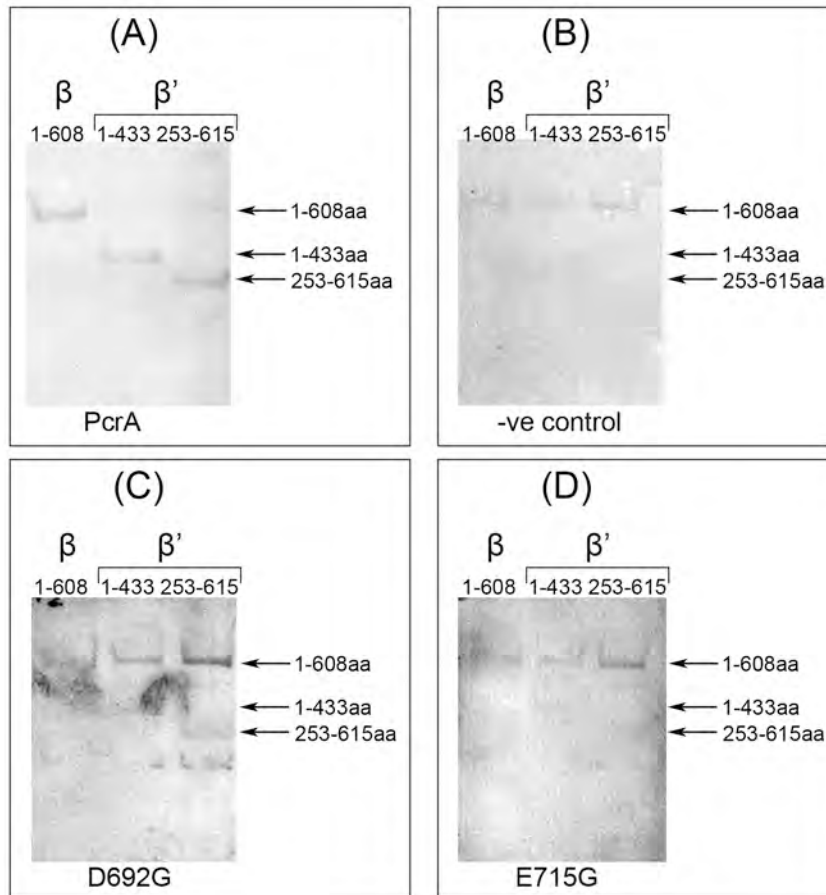
Figure 4.5: Far-Western blot of full length PcrA mutants. (A) PcrA – wild type +ve control. (B) –ve control – no probe. (C) D692G mutation. (D) E715G mutation.

The three fragments of RNAP that PcrA are known to bind to ($\beta_{1\text{-}608aa}$, $\beta'_{1\text{-}433aa}$ and $\beta'_{253\text{-}615aa}$) were transferred to the membrane. The membranes were probed with wild type PcrA$_{Bsub}$ (Figure 4.5A) and the mutant proteins, PcrA$_{D692G}$ (Figure 4.5C) and PcrA$_{E715G}$ (Figure 4.5D). The positive control shows that PcrA is binding to the three RNAP fragments as expected (Figure 4.5A). When assessing the results of the mutant proteins probes, the interactions were harder to identify. The signal intensity of the blots was very low. The interaction is present between the mutants and the β' subunit fragments (1-433aa and 253-615aa), as indicated by arrows (Figure 4.5 C&D). Unfortunately the interaction with the β subunit (1-608aa) cannot be determined, due to the presence of a non-specific band, which is present in all three lanes on both mutant probe blots and the negative control blot (Figure 4.5B). These blots required a longer detection time, which

may have allowed the non-specific bands to develop on the membrane. It was expected that the interaction between PcrA$_{D692G}$ and the β subunit would be disrupted, as shown in the mutant CTD far-Western blot (Figure 4.4A), while PcrA$_{E715G}$ would still interact with the β subunit. If the PcrA NTD aids in establishing the interaction between the CTD and the β subunit then the interaction may still occur with the full length PcrA$_{D692G}$, even though it was disrupted using the truncated PcrA CTD$_{D692G}$. Unfortunately the interaction or lack there of, between PcrA$_{D692G}$ and the β subunit could not be determined using this far-Western blot. Attempts were made to increase the mutant protein concentrations but the results could not be improved.

4.2.2 *In vivo* mutagenesis

4.2.2.1 Vector construction and integration

The integration of the mutations into the *B. subtilis* chromosome first required the construction of a suitable plasmid vector. Previously, plasmids containing the wild type 3' fragment of *pcrA* (pNG515; Table 2.2) and the mutations D692G (pNG554) and I718T (pNG576) were constructed (Harriott, 2006; Withers, 2007). The same method was used to create the remaining mutagenic vectors, E715G (pNG763) and A734V (pNG724) (Table 2.2). The remaining two mutations W699R and A736V were not cloned into the integration vector.

The mutagenic *pcrA* inserts were created using site-directed PCR mutagenesis (Section 2.3.1.3). Integration vector pSG1164 (Lewis & Marston, 1999) was used, where the *gfpmut1* gene was excised using the *Acc*65I and *Spe*I restriction enzymes and the *pcrA* fragment was ligated in its place (Figure 4.6B & C). The plasmids were checked using

restriction digests (Section 2.3.4) to show the *pcrA* clones had an insert at ~500 bp instead of the *gfpmut1* (Figure 4.6D).
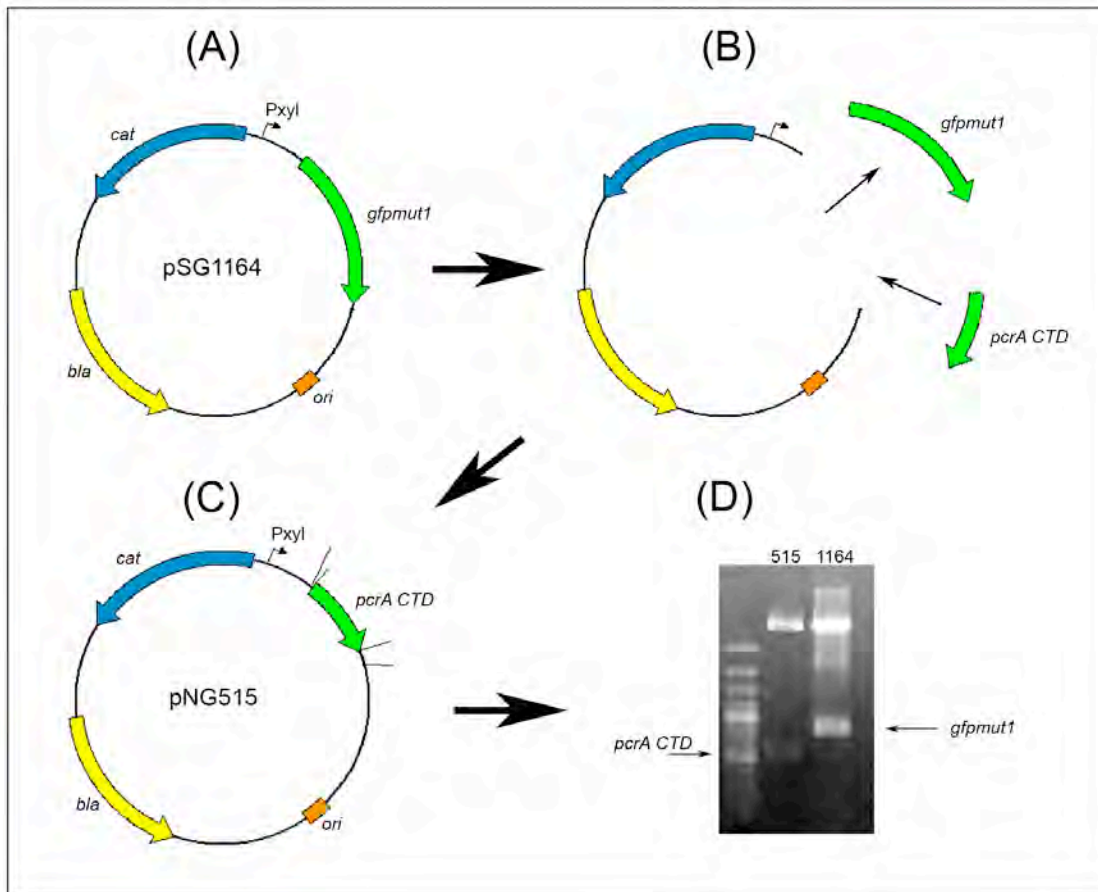
Figure 4.6: (A) Construction of pNG515. *gfpmut1* was removed and replaced with the *pcrA* fragment using the restriction enzymes *Spe*I and *Acc*65I. (B) pNG515 plasmid map, *gfpmut1* was successfully replaced with wild type *pcrA* fragment. (C) Digest check of pNG515, positive clone in lane 2 is indicated by the green arrow and lane 3 shows pSG1164 digest check. All mutagenic *pcrA* vectors were created using the same method (Harriott 2006).

The *pcrA* mutations are all found in the last 150 bp of the 500 bp C-terminal fragment that was integrated into the chromosome (Figure 4.7A). The recombination event can occur anywhere within this 500 bp fragment, indicated by the black crosses (Figure 4.7A). Due to the fact that the mutations occur at the end of the gene fragment it is more likely the recombination event will occur before the mutation, resulting in the incorporation of the mutation into the full length copy of *pcrA* (Figure 4.7B). Chromosomal sequencing reactions were used to determine the position of the mutation

after integration (Section 2.3.7.1). Sequencing reactions were carried out on the PCR

products from the *B. subtilis* chromosomal DNA using specially designed primers. The

forward primer binds upstream of the 3' fragment within the full length *pcrA* gene and

the reverse primer binds downstream of the insert in the plasmid (blue arrows, Figure

4.7B). If the sequence contained the mutation then the recombination event had

occurred before the mutation, integrating it into the full length copy of *pcrA*.

Conversely, if the sequencing results were wild type the recombination event must have

occurred after the mutation, which resulted in the mutation being present in the

downstream partially duplicated gene fragment (Figure 4.7C).
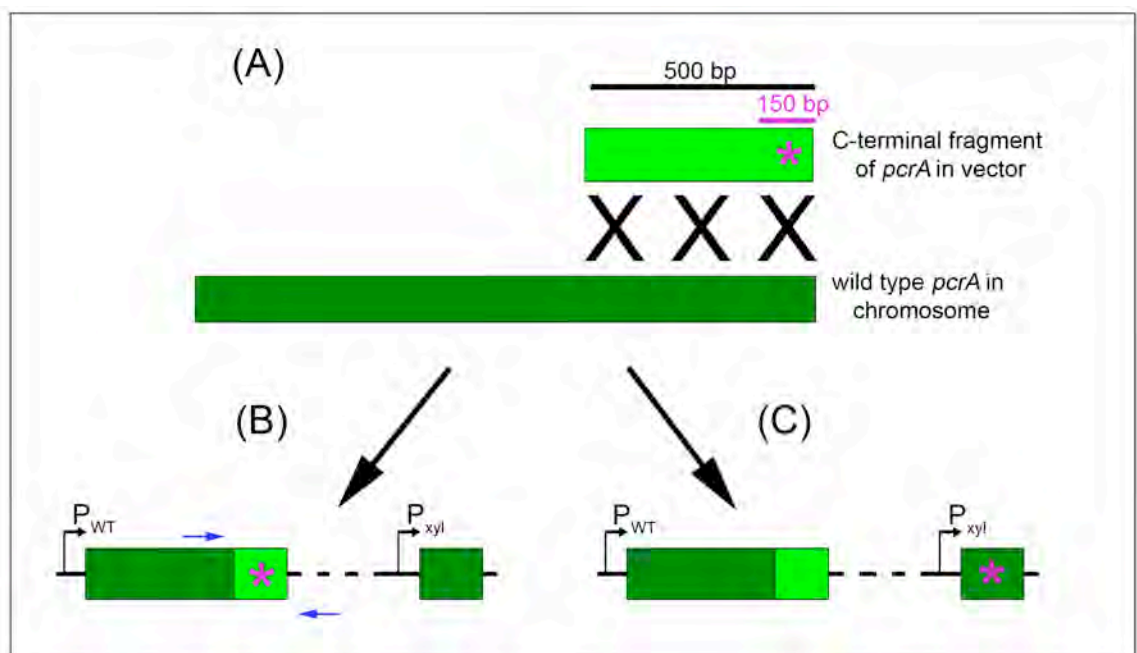
Figure 4.7: Integration of the *pcrA* mutations into the *B. subtilis* chromosomal DNA. (A) The 500 bp fragment of *pcrA* and the positions of the mutation, which are all located in the final 150 bp. Integration into the chromosome can occur at any position within this fragment, indicated by the large black crosses. (B) Recombination occurs before the site of mutation, where the mutation is integrated into the full length *pcrA* gene under the control of the wild type promoter. (C) Recombination occurs after the mutation site, where mutation is found in the *pcrA* fragment under the xylose promoter. Blue arrows represent the primer binding sites used in the PCR and sequencing reactions that were performed to determine the site of integration (Harriott, 2006).

4.2.2.2 Transformation efficiency

The mutant vectors and wild type control were then introduced into the *B. subtilis* chromosome. *B. subtilis* was transformed (Section 2.4.2) with equal amounts of the respective plasmids. Following the transformation experiments (10 replicates), colonies were counted and transformation efficiencies (number of colonies/µg of DNA) for each were calculated for both the wild type and mutant plasmids (Figure 4.8).
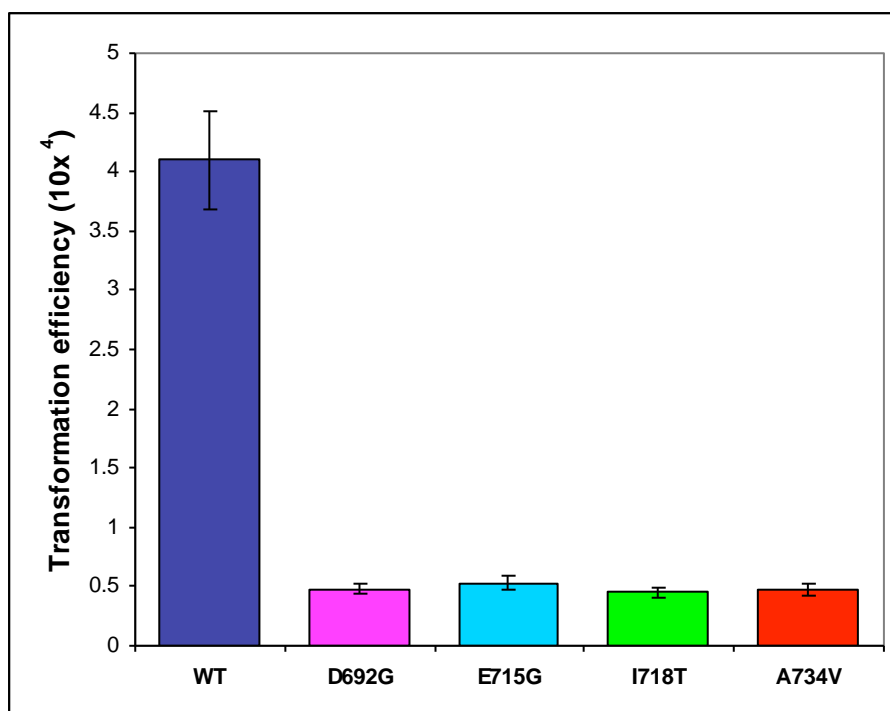
Figure 4.8: Comparison of the transformation efficiency (TE) of PcrA mutations transformed into *B. subtilis* 168. Transformation efficiency (x10^4) is calculated as the number of colonies/µg of DNA and standard deviation bars included. WT is the wild-type control (dark blue), D692G mutation (pink), E715G mutation (light blue), I718T mutation (green) and A734V mutation (red).

All the mutants had a significant drop in transformation efficiency when compared to the wild type plasmid, which was used as a positive control (Figure 4.8). The drop in transformation efficiency was approximately 90% in all four mutations, as shown in Table 4.2.

Table 4.2: The percentage transformation efficiency relative to wild-type when the *pcrA* mutations were introduced into the *B. subtilis* 168 chromosome.

Strain	Percentage TE (%)
WT	100
D692G	11.7
E715G	12.9
I718T	10.9
A734V	11.4

This result must be carefully interpreted and further investigation was required. It was possible that the introduction of the mutations produced a lethal phenotype and the drop in efficiency was the result of low frequency crossover after the point mutation. Sequencing was used to determine if the mutation was present in the full length copy of the *pcrA* gene under the wild type promoter. If the mutation was lethal, indicating that the PcrA-β subunit interaction was essential then it would be expected that all the resultant sequences returned would be wild type.

This was not the case; in fact the majority of the chromosomal constructs sequenced contained the mutations (Table 4.3). This meant that the recombination event occurred before the position of the mutations within the plasmid and the mutations were present in the full length copy of *pcrA* under the control of the wild type promoter.

Table 4.3: *In vivo* mutagenesis sequencing results. The *B. subtilis* chromosomal DNA was amplified by PCR and sequenced for the four different mutations.

Mutation	Mutation present	No mutation	Total number of colonies sequenced
D692G	16	4	20
E715G	21	4	25
I718T	18	2	20
A734V	17	3	20

This result was slightly unexpected as a lethal phenotype for the mutations was the initial reasoning for the lowered transformation efficiency, before receiving the results of the sequencing. Instead, over 80% of colonies sequenced contained the mutation present within full length *pcrA*. Another possible explanation for the lowered transformation efficiency (with respect to the interaction of PcrA with RNAP) was the presence of suppressor mutations occurring within the *rpoB* gene, which compensated for the *pcrA* mutation. To check for suppressor mutations the same procedure was followed. The region of *rpoB* corresponding to amino acids 250-427 from all 4 of the *pcrA* mutant strains was PCR amplified and sent for sequencing. All returned a wild type *rpoB* sequence. At this stage the cause of the drop in the transformation efficiency was unexplained and further work and phenotypic analysis was needed. The resultant *B. subtilis* strains used for further characterization were: BS494 – wild type control (transformed with pNG515), BS495 – D692G (transformed with pNG554), BS496 – E715G (transformed with pNG763), BS497 – I718T (transformed with pNG576) and BS498 – A734V (transformed with pNG724) (Table 2.1).

Another explanation for the drop in the transformation efficiency could be due to the disruption of PcrA interactions with additional proteins. To investigate this possibility the mutant plasmids were transformed into the *uvrB* mutant strain - HCR-9 (Okubo & Romig, 1966). If the same drop in transformation efficiency occurred, it would indicate that the disruption of the UvrB/C interaction with PcrA was not responsible. The resultant transformation efficiencies are shown in Figure 4.9.
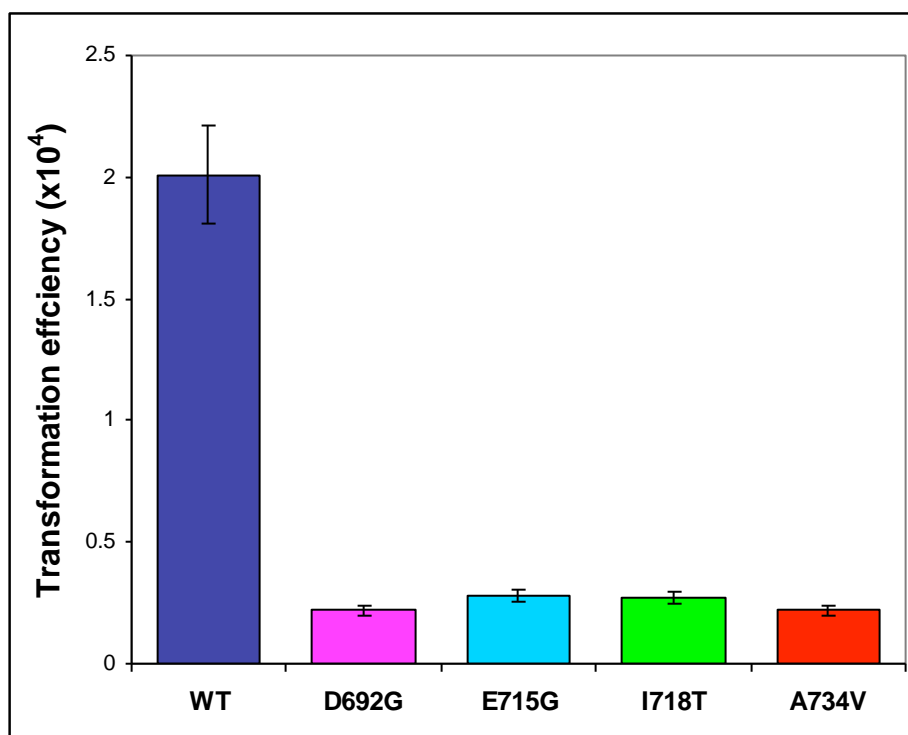
Figure 4.9: Comparison of the transformation efficiency (TE) of *pcrA* mutations transformed into *B. subtilis* HCR-9. Transformation efficiency (x10^4) is calculated as the number of colonies/μg of DNA. Standard deviations are shown. WT is the wildtype control (dark blue), D692G mutation (pink), E715G mutation (light blue), I718T mutation (green), A734V mutation (red).

The results show that the drop in transformation efficiency also occurred when the *B. subtilis* HCR-9 parent strain was used (Figure 4.9). The transformation efficiency of the wild type plasmid was approximately half compared to the *B. subtilis* 168 strain, consistent with previous reports (Okubo and Romig, 1966). The transformation efficiency of the mutant plasmids followed the same trend and decreased further. Again the percentage drop in the transformation efficiency of the mutants, compared to the WT was calculated (Table 4.4). This showed that although the wild type transformation efficiency was lowered in the HCR-9 strain the mutant strains again showed an approximate 90 % drop in transformation efficiency.

Table 4.4: The percentage drop in transformation efficiency when the *pcrA* mutations were introduced into the *B. subtilis* HCR-9 chromosome, compared to the wild type *pcrA*.

Strain	Percentage TE (%)
WT	100
D692G	10.9
E715G	13.9
I718T	13.4
A734V	10.9

This suggested that the drop in transformation efficiency seen when introducing the mutants into the two parent strains was not caused by disrupting the PcrA-UvrBC interaction. Therefore this could be due to disrupting the PcrA-β subunit interaction. The YxaL mutant strain of *B. subtilis* had already been shown to display no phenotypic changes (Noirot-Gros *et al*, 2002b) so transformation of the *pcrA* mutations was not performed with that strain as it was not expected to provide any additional data. The resultant mutant PcrA strains were then examined to determine if any phenotypic changes were observable.

4.2.2.3 Growth kinetics of pcrA mutant strains

Growth curves were performed to determine if the PcrA mutations were having an effect on the growth rate of the *B. subtilis* strains. Absorbance at OD_{600} was measured at 10 min intervals using a shaking, incubating plate reader (Section 2.7.2). This format allowed for 6 replicates of each strain during every growth curve experiment. The growth kinetics of all four PcrA mutant strains: D692G (Figure 4.10A), E715G (Figure 4.10B), I718T (Figure 4.10C) and A734V (Figure 4.10D) were compared to the *B. subtilis* wild type strain BS494. BS494 was created by transforming the wild type *pcrA* fragment into *B. subtilis* 168 to use as an appropriate control.
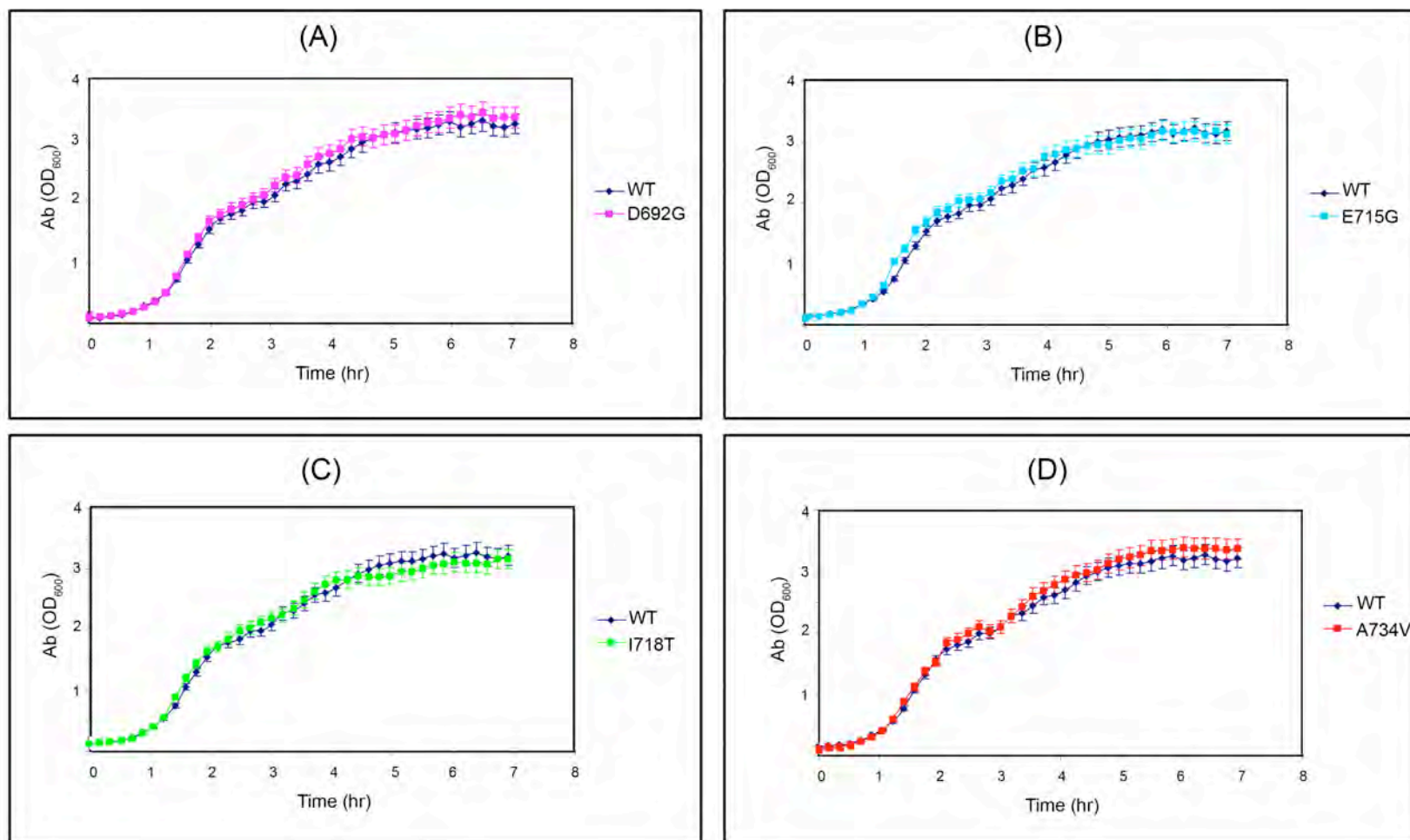
Figure 4.10: Growth curves of the mutant PcrA *B. subtilis* strains compared to the wild type strain BS494 (WT ♦) in LB media. (A) BS495 - D692G mutation (■). (B) BS496 – E715G mutation (■). (C) BS497 – I718T (■). (D) BS498 – A734V (■).   Standard deviations are shown.

The growth curves showed that all four PcrA mutant strains had the same growth kinetics as the wild type strain. The doubling times of each of the strains were also calculated using the growth curve data ($\log_{OD600}$ vs. time) and this also showed that there are minimal differences between the growth rates of the mutant and wild type strains (Table 4.5).

Table 4.5: Doubling times of the mutant *pcrA B. subtilis* strains, compared to the wild type control strain. All strains were grown in selective LB media.

Strain	Doubling Time (min)	Std deviation
WT	30.81	±0.45
D692G	29.90	±0.40
E715G	30.90	±0.39
I718T	30.48	±0.31
A734V	28.73	±0.34

4.2.2.4 Further cytological analysis of pcrA mutant strains

Following the results of the growth kinetics, further analysis was required to determine if the strains had a wild type appearance to match the growth kinetics. Microscopy was used to measure and compare the average cell length of the strains. Again the strains were grown at 37 °C in LB medium to mid exponential phase, when samples were taken and microscope slides prepared (Section 2.8). Average cell sizes were calculated using MetaMorph software (v 7.1.3.0, Universal Imaging Corps, USA) and a total of 400 cells/strain were counted (Table 4.6).

Table 4.6: Cell length of the mutant *pcrA B. subtilis* strains, compared to the wild type control strain. All strains were grown in selective LB media.

Strain	Average length (µm)	Std deviation
WT	5.48	±1.21
D692G	5.41	±1.21
E715G	5.42	±1.33
I718T	5.26	±1.32
A734V	5.55	±1.22

The average cell size between all five strains was shown to be similar (Table 4.6). No change in average cell length was observed, when comparing the mutant strains to the wild type strain. This indicated the *pcrA* mutations were not affecting the *B. subtilis* cell size.

The growth kinetics and cytological analysis of cell lengths showed that the introduction of the *pcrA* mutations had had no effect on the cells after the initial drop in transformation efficiency. This indicated that the disruption of the PcrA-RNAP interaction did not adversely affect the *B. subtilis* cells. However, following the *in vitro* interaction studies (Section 3.2.2.2) it was now known that there were additional points of contact in the PcrA-RNAP interaction, other than between the CTD of PcrA and the β subunit. As shown in Chapter 3, the N-terminal region of PcrA was found to bind the N-terminal segment of the β' subunit. This interaction may be able to compensate for the C-terminal mutations in PcrA *in vivo*. Further studies were performed to determine whether mutant PcrA was still able to form a complex with RNAP.

4.2.2.5 Mutagenic tandem affinity purifications (TAP-tag)

To further investigate the PcrA-RNAP interaction in the mutant PcrA *B. subtilis* strains, tandem affinity purification was used to isolate RNAP complexes. This method utilized the dual SPA-tag (FLAG and Calmodulin-binding peptide (CBP)) on the β sub-unit of RNAP. The tandem purifications were performed under non-denaturing conditions (Section 2.9) to isolate the interacting protein complexes (Delumeau *et al*, 2011). This experiment was designed to confirm wild-type PcrA could be purified in complex with RNAP (Delumeau *et al*, 2011) and secondly, to determine whether disruption of the PcrA CTD-β subunit interaction prevented the RNAP-PcrA complex from forming. The significance of this result could support the previous *in vitro* interaction studies that had identified an additional site of interaction between RNAP and PcrA.

The chromosomal DNA of the *rpoB-spa* strain (EU299, Table 2.1) was transformed into the PcrA mutant strains (Section 2.4.2; Table 2.1). If the PcrA mutations completely disrupted the interaction between PcrA and RNAP, PcrA would not be detected in the elutions of the tandem purifications.
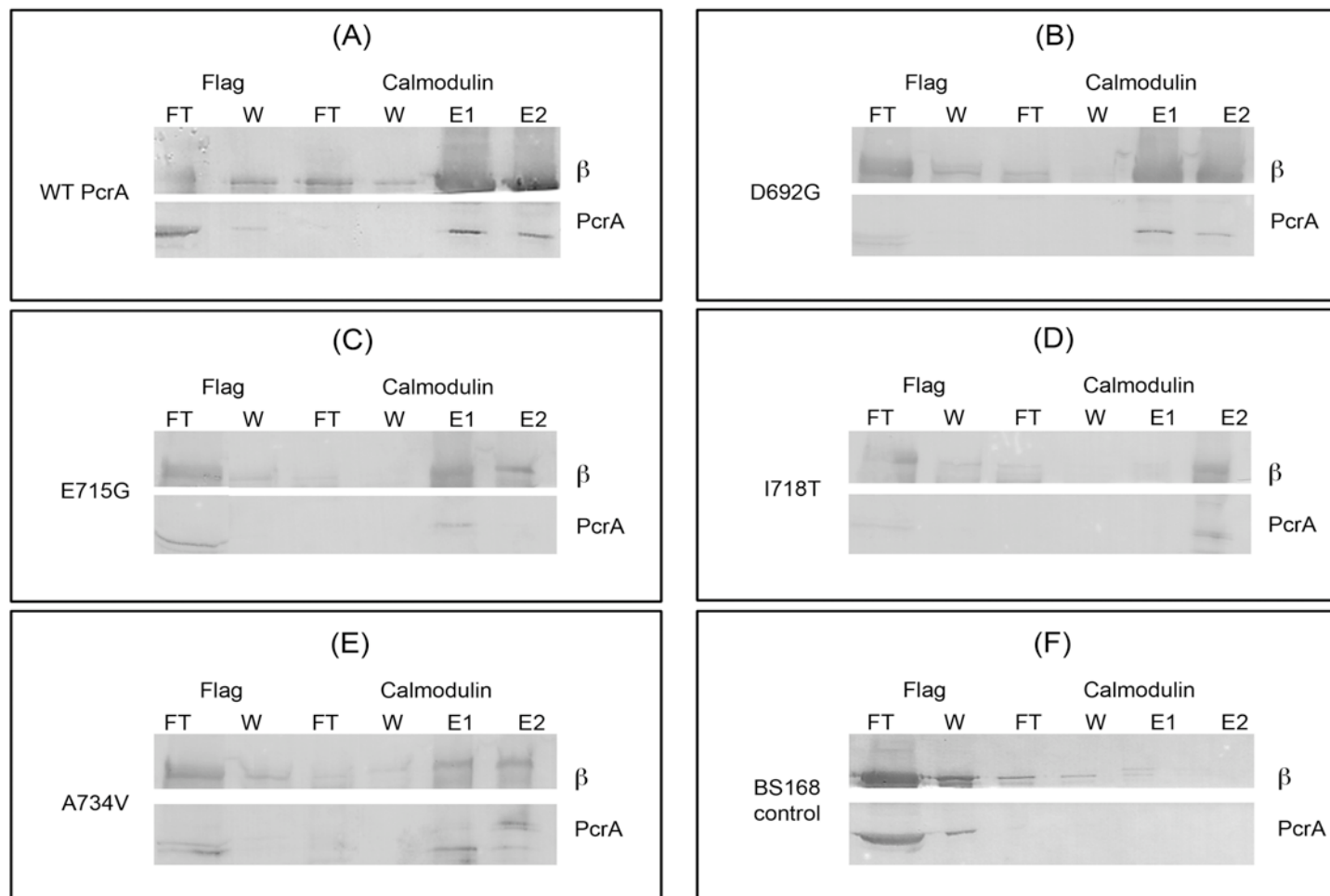
Figure 4.11: Tandem affinity purification (TAP-tag) of RNAP. (A) Positive control – BS500 - wild type PcrA. (B) BS501 – D692G. (C) BS502 - E715G. (D) BS503 – I718T. (E) BS504 – A734V and (F) Negative control - BS168. Two purifications were performed - FLAG and Calmodulin, FT – flow through, (W) – wash, (E1, E2) – elutions.

The results of the TAP-tag purifications showed that the mutant PcrA proteins still interacted with RNAP *in vivo*. The positive control of wild type PcrA (Figure 4.11A) showed that PcrA was detected in both elutions (E1 & E2), after the calmodulin-binding peptide affinity purification. PcrA was also detected in the elutions of the mutant strains (Figure 4.11B-E). The negative control used (Figure 4.11F) was wild type 168 *B. subtilis*. This strain did not contain tagged *rpoB* therefore the β subunit and PcrA were not expected in the final elutions (Figure 4.11F - E1 & E2). The mutant strain purifications successfully identified the specific PcrA-RNAP interaction and isolated RNAP and PcrA in complex.

This result also indicated the additional sites of interaction occurring between RNAP and PcrA are sufficient for the proteins to remain in complex. Although the introduced PcrA mutations disrupted the interaction occurring between the β subunit and PcrA there was still an interaction occurring between PcrA and RNAP *in vivo*.

4.3 Discussion

The analysis of the PcrA CTD–β subunit interaction was undertaken as a result of a mutagenic yeast two-hybrid assay, which was completed before the *in vitro* interaction studies (Chapter 3) were finalised. When the PcrA mutations were identified, it was believed that the interaction between PcrA and RNAP existed between PcrA CTD and the β subunit only.

The first step after receiving these interaction disrupting mutations in PcrA, was to clone the mutations and confirm the mutations disrupted the interaction. The result of

the far-Western blots using the mutant CTD proteins showed that one PcrA mutation (CTD$_{E715G}$) was still interacting with the β subunit fragment (Figure 4.4A). All other mutants did not bind to the β subunit, which confirmed the yeast two-hybrid results. Although the far-Western blot is not a quantitative technique, it was thought that the strength of the CTD$_{E715G}$-β subunit interaction, could be compared to the wild type CTD, using increasing NaCl washes to break the protein interactions. This additional far-Western blot showed that the wild type and CTD$_{E715G}$ proteins both remained bound to the β subunit on the membrane after washing the membrane with up to 1.5 M NaCl (Figure 4.4B) and furthermore, this suggests the CTD$_{E715G}$ mutant does not have a reduced affinity for the β subunit.

To follow the *in vitro* results using the mutant CTD proteins, two of the PcrA mutations, D692G and E715G, were then introduced into the full length PcrA protein and the proteins were purified. The introduction of the single amino acid change into the PcrA sequence may have had an effect on the solubility of the overproduced PcrA protein. It was thought that the mutations may be having an effect on the solubility as both mutant proteins (PcrA$_{D692G}$ and PcrA$_{E715G}$) did not overproduce and purify as well as the wild type protein did (Section 3.2.6.1). The lowered quantity of purified mutant proteins obtained made the *in vitro* interaction studies difficult to optimise. The aim of purifying the two mutations (D692G – interaction disrupting and E715G – still interacts with the β' subunit) was to determine if the presence of the PcrA NTD could restore the interaction between the CTD and the β subunit, in the presence of the mutation or if the interaction was still disrupted in the full length PcrA$_{D692G}$ protein. The PcrA$_{E715G}$ protein was included to see if the mutation had an adverse effect on the full length PcrA binding, which was absent in the CTD$_{E715G}$ protein. Unfortunately the results obtained

from the far-Western blots using the PcrA mutant proteins (Figure 4.5) were difficult to interpret. The interaction between PcrA proteins and the N-terminal region of the β' subunit was detected, but non-specific bands in the Western blot were present on the membrane, which prevented the detection of the interaction between the mutant PcrAs and the β subunit fragment. The non-specific bands most likely occurred due to the increased incubation time required to detect the lower protein concentrations.

After confirming the PcrA mutations disrupted the interaction, the effect was examined *in vivo,* where four mutations, including E715G, were introduced into the *B. subtilis* chromosome. PcrA is essential for cell viability and currently the essential nature of PcrA is attributed to a role in homologous recombination repair (Petit *et al*, 1998). If the interaction between the CTD and the β subunit was also essential, introducing the mutations would produce a lethal phenotype. When mutations were introduced, the transformation efficiency was lowered by up to ninety percent, when compared to the positive control. Initially, when these results were obtained it was thought that the interaction was indeed essential and a lethal phenotype was seen. The small number of colonies obtained could have been due to segregation against the mutation. To confirm this, sequencing of the chromosomal DNA was carried out and showed that there was no segregation against the mutation, in fact most of the sequences contained the PcrA mutations. Another reason for the lowered transformation efficiency was the rescue of a lethal phenotype by the presence of suppressor mutations in the interacting region of the β subunit. Therefore, the next step was to check the sequence of the N-terminal region of β to identify potential mutations. The results did not identify any suppressor mutations, which suggested the disruption of the CTD-β subunit interaction did not produce a lethal phenotype.

The resultant mutant PcrA *B. subtilis* strains were then examined to determine other phenotypic changes, in addition to the lowered transformation efficiency. The cytological studies, including the growth kinetics, DNA localisation (not shown) and cell length did not identify any phenotypic changes, when compared to the wild type strain. This work unfortunately did not provide any additional information regarding the possible cause for the lowered transformation efficiency.

To investigate the interaction between the mutant PcrA proteins and RNAP *in vivo*, tandem affinity purification (TAP-tag) was used. TAP-tagging can be effectively used to isolate protein complexes *in vivo* (Yang *et al*, 2008a). PcrA had been shown to co-purify with RNAP, which had been fused to a dual purification SPA-tag (Delumeau *et al*, 2011). The *in vivo* interaction with wild type PcrA was confirmed in this work and the TAP-tagging purification was performed with the four PcrA mutant strains. The results of the TAP-tagging showed the mutant PcrA proteins were co-purified with RNAP (Figure 4.11). This suggested that either the NTD-β' interaction was all that was required for PcrA and RNAP to interact *in vivo*, or alternatively that the full length protein can overcome the point mutations that disrupted the CTD-β interaction.

The mutagenic analysis of the CTD-β subunit interaction was investigated, as it may provide a role for PcrA during transcription. The region of the β-subunit which interacts with PcrA also interacts with Mfd. Mfd can remove stalled RNAP elongation complexes from DNA and recruit DNA excision repair machinery to sites of DNA lesions (Selby & Sancar, 1993a; Selby & Sancar, 1993b). Mfd is not a DNA helicase but it has conserved helicase motifs and can translocate on DNA (Section 1.3.3). In addition to releasing stalled complexes Mfd can also restart elongation complexes by

replacing the 3' end of the RNA back in the RNAP active site (Figure 4.12, Park *et al*, 2002; Roberts & Park, 2004).
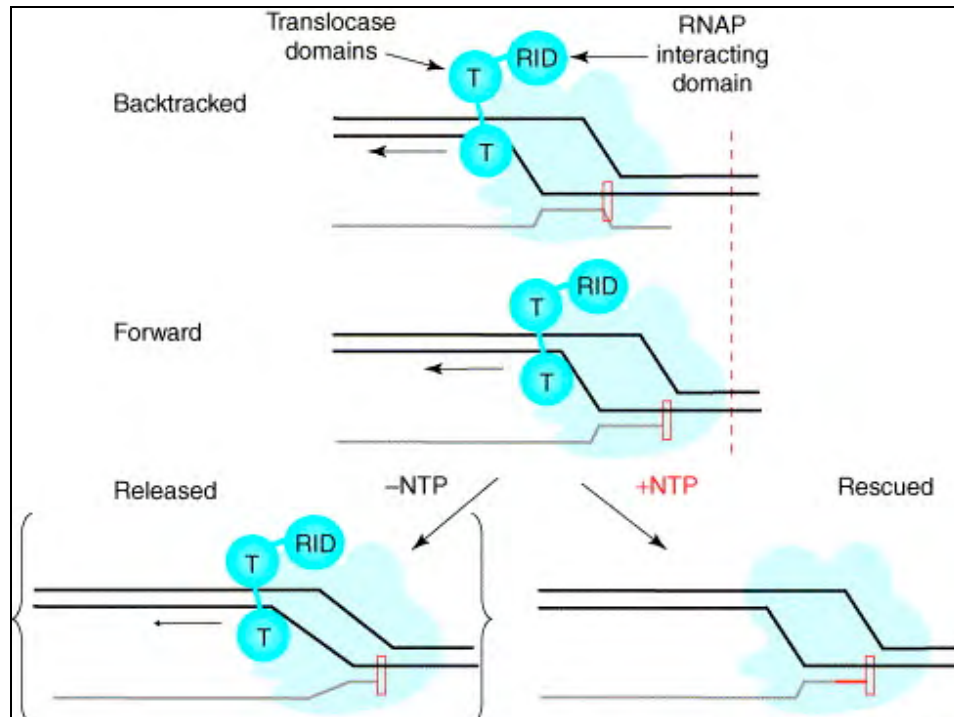
Figure 4.12: The effect of Mfd on RNAP. Top shows backtracked RNAP, bound to Mfd through the RNAP interacting domain (RID). The two Mfd translocase domains (T) translocate in the direction of the arrow, which pushes the RNA forward until the RNA 3' end is in the active site (red rectangle). If NTP is present (right), transcription elongation can continue. If no NTPs are available, Mfd causes the collapse of the transcription bubble and RNAP dissociation occurs (Roberts & Park, 2004)

PcrA has also been shown to remove proteins bound to DNA and the highly processive translocase nature of PcrA could mean PcrA could remove stalled RNAP and recruit DNA repair enzymes (such as UvrB/C). The repair of DNA lesions is important to ensure the essential processes of RNA transcription and DNA replication can occur. Stalled RNAP complexes can also block the progression of DNA replication forks, which is detrimental to cell viability. To ensure cell viability there are levels of redundancy within important cellular processes, meaning PcrA and Mfd could have a shared role in transcription. This possible role of PcrA made the mutagenic analysis of

the CTD-β subunit interaction an important part of this project and although this role could not be shown in this work it still remains a possibility.

4.4 Conclusions

This chapter investigated the specific amino acids that disrupted the interaction between the PcrA CTD and the β subunit of RNAP. This mutagenesis work was started before the complex nature of the PcrA-RNAP interaction was known and the additional binding sites on the β' subunit were discovered.

The important results from this work showed that although the mutations disrupted the CTD-β subunit interaction *in vitro* the proteins remained in complex *in vivo*, as shown in the tandem affinity purification. This work requires further investigation into the reasons for the lowered transformation efficiency when introducing the mutations into the *B. subtilis* chromosome, as cytological studies performed did not show any other phenotypic changes. It was hoped that this work could provide further information in the specific site of interaction between PcrA and the β subunit and also provide insights into the possible roles of PcrA during transcription.

Chapter 5:

PcrA and RNAP in complex –

Single particle analysis using

negative staining and transmission

electron microscopy

Chapter 5: PcrA and RNAP in complex –

Single particle analysis using negative staining and transmission

electron microscopy

5.1 Introduction

The characterisation of the interaction between PcrA and RNAP has shown that there are multiple binding sites between the two proteins. Far-Western blotting (Chapter 3) suggests that PcrA interacts with the upstream face of RNAP. To support these results the structure of the PcrA-RNAP complex was investigated.

5.1.1 Single particle reconstruction

The PcrA-RNAP complex is large and it would be difficult to solve the structure using X-ray crystallography. Instead, negative staining transmission electron microscopy (TEM), single particle analysis and 3D reconstruction were used. Single-particle reconstruction is typically used when solving the structure of larger viruses or macromolecular protein assemblies, as the lower size limit for successful results using this technique is approximately 200 kDa (Ludtke, 2010). Additionally, when negative staining TEM is used to visualise the particles the final resolution of the structures is limited to 15-30 Å. It was hoped that although the structure of the PcrA-RNAP complex would be low resolution it would provide information on the position of PcrA, when interacting with RNAP.

The EMAN (Electron Micrograph ANalysis) software package was used to create the 3D reconstructions from the single particle images. This software breaks the reconstruction process into three main stages: particle selection, creation of an initial model, and refinement (Figure 5.1, Ludtke *et al,* 1999). The single particles are selected from raw micrographs. This can be done either manually or semi-automatically, where the graphical *boxer* program (within EMAN) can choose particles similar to the reference particles that were manually chosen (Figure 5.1A). Once a particle data set has been generated the refinement procedure requires the construction of an initial model (Figure 5.1B). There are a number of methods that can be used to generate the initial model, including using a small subset of particles, that represent different views of the molecule to generate a 3D map with known symmetry. When the reconstruction is performed on asymmetrical particles (such as RNAP), there is a generic model-generating routine available in EMAN (Ludtke *et al*, 1999). This process uses reference-free class averages, where each class average represents a characteristic view of the molecule to construct the 3D model. Hopefully, the class averages will represent most of the different orientations of the particle and several class averages can be chosen to create the initial model. The quality of the initial model can influence the number of iterations the refinement loop must complete for the model to reach convergence and create the final model output (Ludtke *et al*, 1999). EMAN determines model convergence to occur when the Fourier Shell Correlation (FSC) stabilizes between subsequent rounds of iteration (Ludtke *et al*, 1999).
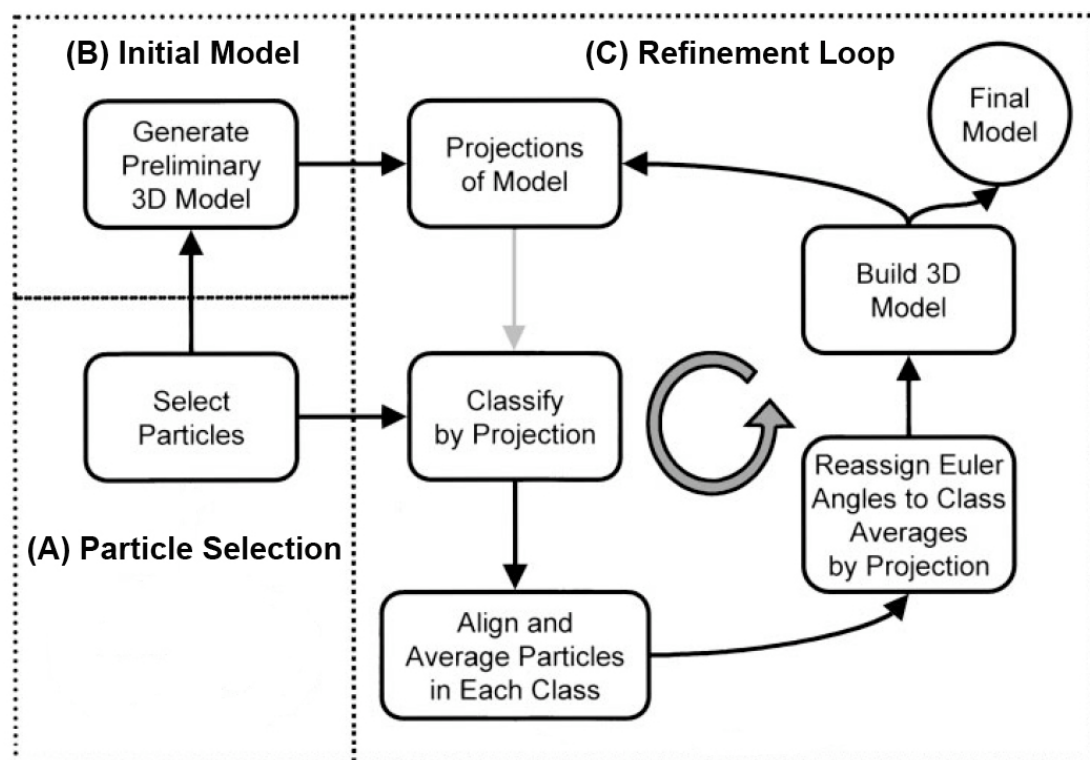
Figure 5.1: The EMAN reconstruction process, which has 3 major stages. (A) Particle selection. (B) Generation of the initial model. (C) Refinement loop. Here the initial model is refined iteratively using the particle data set until convergence is reached, which becomes the final model (Ludtke *et al*, 1999)

The refinement loop used in EMAN creates a set of projections of different orientations of the initial model. These projections are used to classify the raw particle images, commonly using the *classbymra* program. This program can rotate and align the particles to the reference projections. Particles within each projection are then aligned and averaged to create the class average (Ludtke *et al*, 1999). The class averages are assigned the Euler angles from the projections and used in this orientation to construct a new 3D model (Figure 5.1C). This model is then used as the initial model in the next round of refinement and the loop continues until a final model is created. The final model can then be assessed to determine the resolution and hopefully provide detailed structural information.

Following the reconstruction of protein complexes, the docking of known protein structures within the final electron density map could provide information on the position and site of interaction between the proteins. This was the aim of using this process to resolve the PcrA-RNAP complex structure.

5.2 Results

5.2.1 TEM and particle picking

In order to resolve the PcrA-RNAP complex structure both proteins were purified (Section 2.5.4) and then mixed *in vitro* to form the complex (Section 2.10.1). In order to correctly identify single particles of the PcrA-RNAP complex and limit the amount of unbound PcrA and RNAP present in the sample, the proteins were mixed in an approximately 1:1 ratio, with a slight excess of PcrA. Typically a larger excess of the smaller protein would be used to form the complex but as PcrA is also a large protein the unbound PcrA was seen on the micrographs. If picked particles were just PcrA it would adversely affect the 3D reconstruction. The sample was loaded onto carbon grids and stained with uranyl formate (Section 2.10.1). Negative staining with a heavy metal, such as uranyl formate increases the contrast and promotes the visibility of the particles in the micrograph (Ohi *et al*, 2004). A total of 450 micrographs of the PcrA-RNAP complex were captured (Section 2.10.2). The micrographs were assessed for particle aggregation, poor contrast and astigmatism and micrographs with these issues were discarded, the single particles were then isolated from the remaining micrographs (Figure 5.2A). This involved the manual selection of single particles that represented

the PcrA-RNAP complex in random orientations (Section 2.10.3). After initially selecting all particles from the micrographs the data set contained approximately 17 000 particles. The large data set was manually pruned at least three times to remove images falsely identified as particles, where each round of pruning removed more 'bad' particles. The size of the final PcrA-RNAP complex data set used in the 3D reconstruction was 15 458 particles and an example of some of the final particles selected is shown in Figure 5.2B.
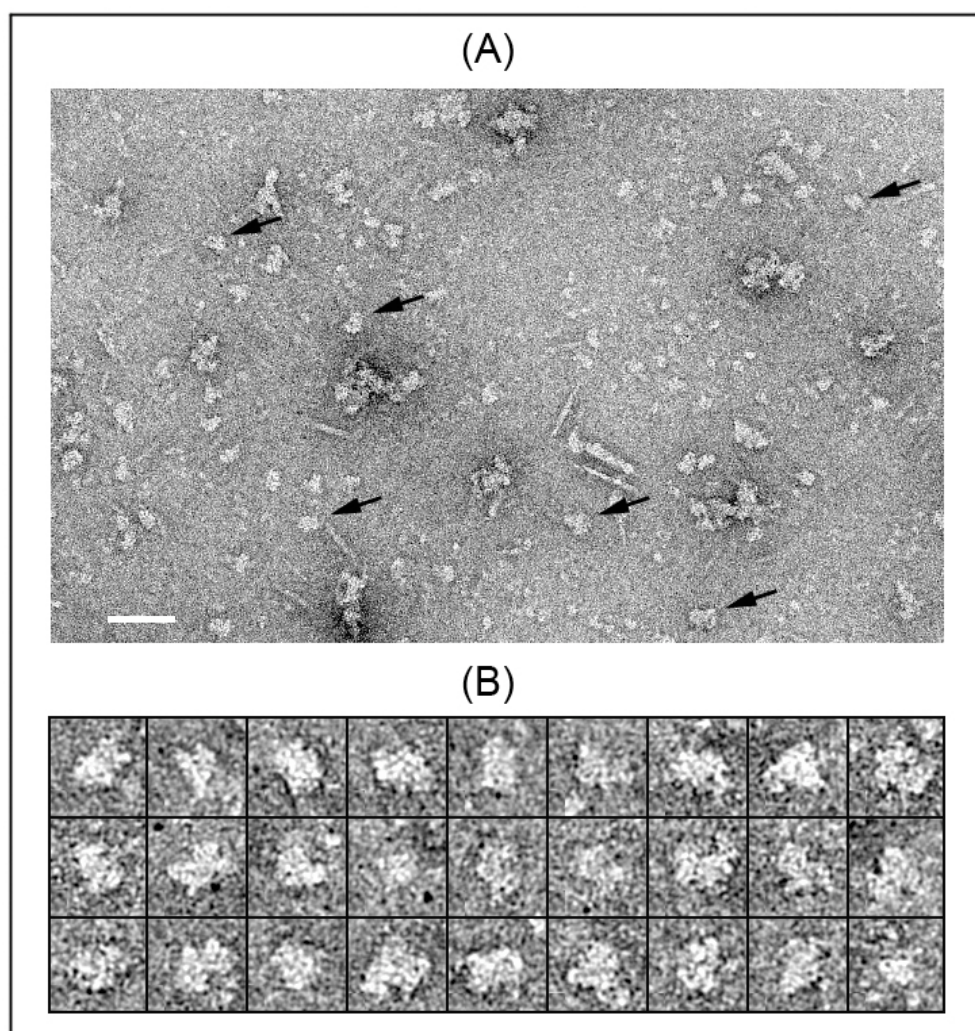
Figure 5.2: Negative staining TEM. (A) A section of a TEM micrograph with examples of individual PcrA-RNAP complexes indicated with arrows. Scale bar is 50 nm. (B) Small subset of the individual particles that were boxed out from micrographs.

5.2.2 3D Reconstruction

The data set was then subjected to 3D reconstruction using EMAN software (Section 5.1; Ludtke, 1999). The PcrA-RNAP complex is an asymmetrical molecule and is relatively small for use in single particle reconstruction, which can cause issues with the quality of the reconstruction. This 3D reconstruction from negatively stained particles was more commonly used in solving the structures of large symmetrical molecules (Ludtke *et al*, 2001). Luckily, core RNAP and RNAP in complex with transcription factor NusA, had been resolved (Yang *et al*, 2009) using this method. This work provided a basis for completing the PcrA-RNAP complex reconstruction and also provided a point of comparison between the complex and the core RNAP.

The initial model and the classification of the particle data set into differing class averages formed the basis of the reconstruction. Particles were analysed in each round of refinement and separated into different orientations, which were then used to form the 3D model. The correct classification of the particles is more difficult when no axis of symmetry is present, such as in the case with RNAP. In order to assess the quality of the model output, the distribution of the particles across the class averages could be assessed. This involved extracting the number of particles found in each of the class averages after the final iteration round of the reconstruction and plotting the results (Figure 5.3A). For example 118 class sums consisted of 14 particles in the RNAP-PcrA complex data set. This was performed for both the PcrA-RNAP complex and the RNAP core. A normal distribution of the number of particles across the class averages showed that the reconstruction was not biased to certain orientations. If certain class averages had many more particles, while others had very few this would bias the 3D

reconstruction and it would suggest that the class averages are not representative of the molecule in all of the different orientations. Without a normal distribution of particles it is much harder for the refinement process to reach convergence and achieve an accurate final model. Both the PcrA-RNAP complex (Figure 5.3A, PcrA) and RNAP core (Figure 5.3A, RNAP) reconstructions showed a normal distribution. The total number of particles used in the PcrA-RNAP complex data set (15 458) was greater than the RNAP core data set (13 801) and this accounts for higher peaks in particle numbers for the PcrA-RNAP complex.
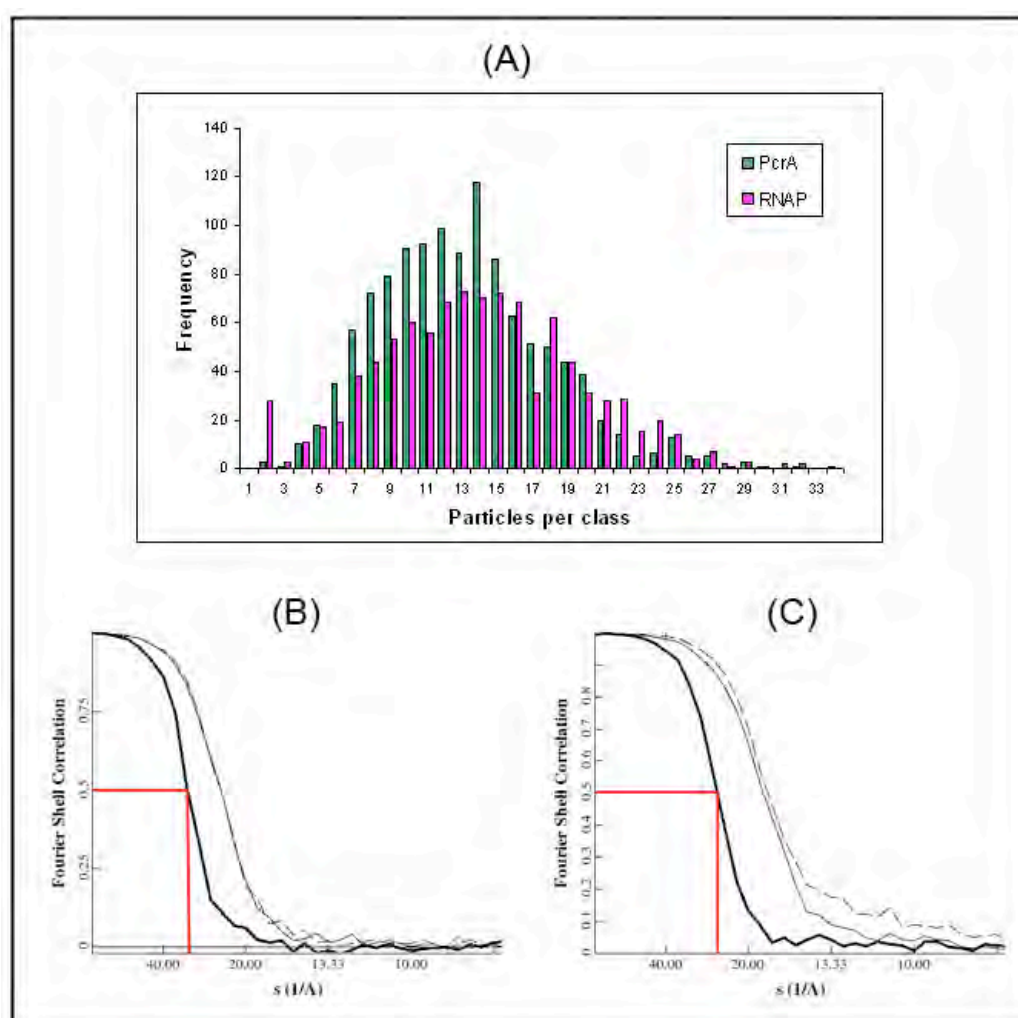
Figure 5.3: Assessment of model quality. (A) Particle distribution within the class averages. Normal distribution indicated the reconstruction was not biased to certain orientations. (B) Estimation of the PcrA-RNAP complex model resolution using the even-odd (EO) test. (C) Estimation of the RNAP core model resolution using the EO test. EO test is the dark line on the FSC graph, resolution is estimated when the curve dropped below 0.5 (red line). Grey curves are the FSC curves of the PcrA-RNAP complex and RNAP core models.

After the 3D reconstruction was completed the quality of the model could also be assessed using an even/odd (EO) test in EMAN. This test was used to estimate the resolution of the model, which separated the particles into 'even' and 'odd' half data sets and performed independent reconstructions on each half. The point at which the EO test curve drops below 0.5 on the Fourier Shell Correlation (FSC) graph is the estimation of resolution. This was shown by the red lines in the FSC graph for both the PcrA-RNAP complex (Figure 5.3B) and the RNAP core (Figure 5.3C). The resolution of the RNAP core was estimated to be approximately 25 Å while the resolution of the PcrA-RNAP complex was lower, estimated to be above 35 Å. The Fourier shell correlation (FSC) criterion, particularly the conservative 0.5 FSC threshold, is another standard quality measure that could be used determine resolution, but the choice of FSC threshold remains controversial (van Heel & Schatz, 2005). The grey curves in Figure 5.3B

Following 3D reconstruction the models were compared to the 2D reprojections and class averages at corresponding Euler angles. This was to ensure the assignment of the Euler angles was correctly performed during the 3D reconstructions. When comparing the 3D model (I) to the 2D reprojection (II) and the class averages (III) for both the RNAP core (Figure 5.4A) and the PcrA-RNAP complex (Figure 5.4B), the additional mass is seen in certain orientations, indicated with arrows.
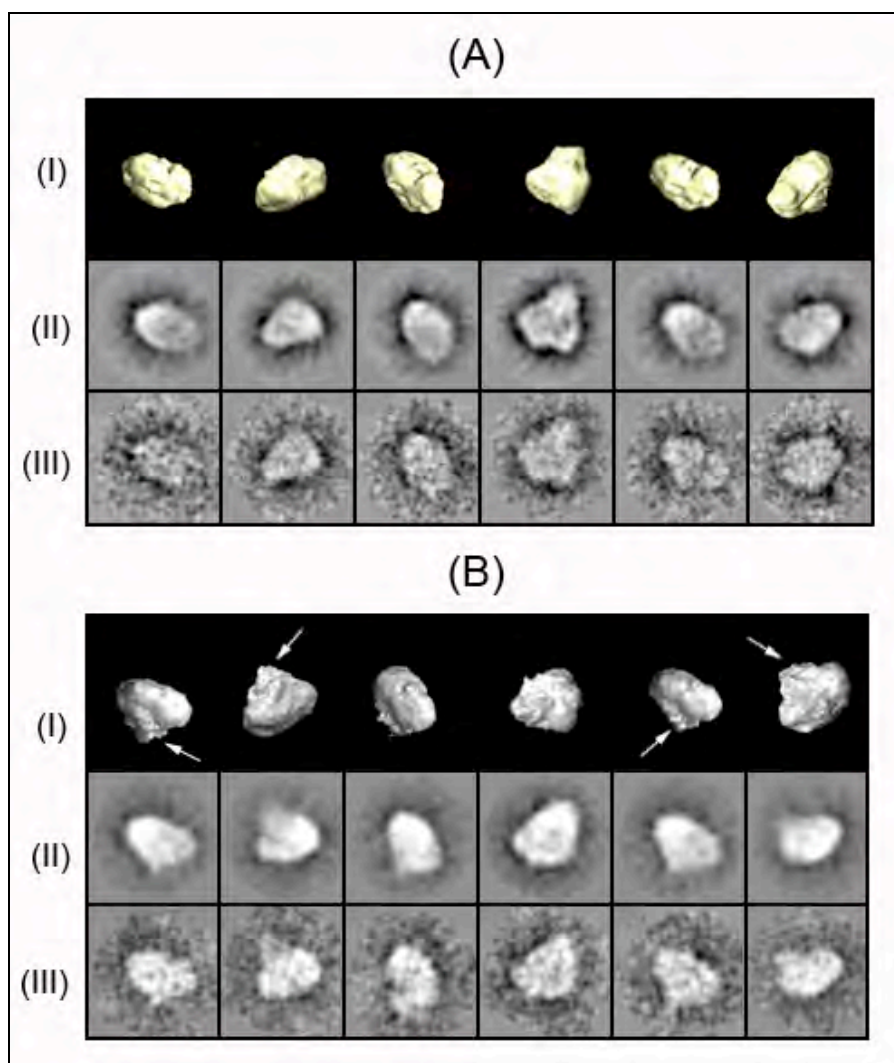
Figure 5.4: 3D reconstructions (I), 2D reprojections (II) and class averages (III) of (A) the RNAP core (Yang et al, 2009) and (B) the PcrA-RNAP complex. The additional mass attributed to PcrA is indicated by arrows on the 3D reconstructions (I).

The comparison of the 3D models of the PcrA-RNAP complex and the RNAP core were then performed in UCSF Chimera (Pettersen *et al*, 2004). The electron density maps of both the PcrA-RNAP complex (Figure 5.5A & D) and RNAP core (Figure 5.5B & E) are shown, along with the overlay of the two models (Figure 5.5C & F).
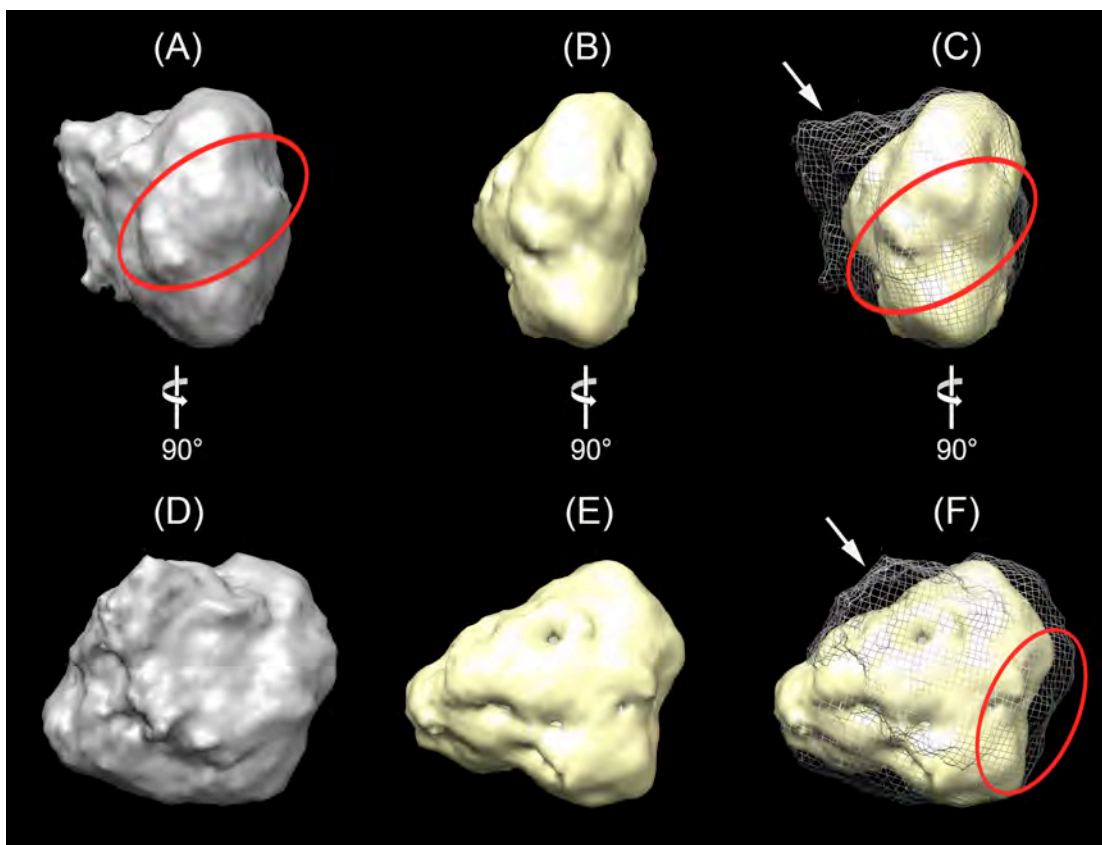
Figure 5.5: Reconstruction of the PcrA-RNAP complex. (A) PcrA-RNAP complex reconstruction. (B) RNAP core reconstruction (Yang *et al.* 2009), for comparison. (C) Overlay of the RNAP core and PcrA-RNAP complex reconstructions. Reconstructions are shown in the face-on orientation. (D) PcrA-RNAP complex reconstruction. (E) RNAP core reconstruction. (F) Overlay of the RNAP core and PcrA-RNAP complex reconstructions. Reconstructions are turned 90° into the page to show upstream side-on orientation. PcrA-RNAP complex is grey and RNAP core is yellow. The PcrA complex is represented by a mesh in panels (C) and (F), where the additional mass attributed to PcrA is indicated by an arrow. The low resolution of the ββ'claw in the PcrA-RNAP complex is indicated by a red circle.

The PcrA-RNAP complex model, in grey clearly shows an additional mass present when compared to the RNAP core structure, in yellow. This mass is seen in both the face- and side-on orientations of the model and is indicated with arrows when the two models were superimposed (Figure 5.5C & F), where the PcrA-RNAP complex structure is now shown in grey mesh. Previous work with the higher resolution RNAP core model has used docking to correctly orientate the RNAP homology structure within the 3D electron density map (Yang *et al*, 2009). This work allowed the determination that the additional mass attributed to PcrA was on the upstream face of RNAP in close agreement with the data obtained by far-Western blotting in Chapter 3.

5.3 Discussion

The aim of this work was to resolve the structure of the PcrA-RNAP complex to enhance the results of the *in vitro* interaction studies (Chapter 3), which had mapped the site of interaction between the PcrA and RNAP.

5.3.1 Analysis of the 3D reconstruction model of the PcrA-RNAP complex

Single particle reconstruction had been used to successfully resolve the structure of the RNAP core and RNAP in complex with transcription factor NusA (Yang *et al*, 2009). There were a few differences in the practical methodology used, between the PcrA-RNAP complex and the RNAP core. Firstly the PcrA-RNAP complex was stained with uranyl formate (UF), while the earlier RNAP core sample was stained with uranyl acetate (UA). UF staining was used in the PcrA-RNAP complex rather than UA because it was thought that the smaller grain of the uranyl formate might improve the quality of the micrographs (for example see Knight, 1975). Single particles were seen on the micrographs but aggregation of the particles and dark spots of uranyl formate were also present (Figure 5.2A). These areas on the micrographs could not be used to pick particles and therefore a large number of micrographs needed to be examined in order to obtain a suitably large particle data set. To ensure that the PcrA-RNAP complex was isolated, and not core RNAP or free PcrA, the two proteins were mixed at a 1:1 ratio which may have caused some sample aggregation, which frequently occurs in negatively stained samples. When comparing the class averages of the PcrA-RNAP complex (Figure 5.4AIII) to those of the RNAP core (Figure 5.4BIII) the difference in the staining can be seen. The PcrA-RNAP complex class averages looked brighter,

where the smaller particles of UF created brighter, less grainy images. Another difference between the PcrA-RNAP complex and the RNAP core was the dark ring surrounding the UA-stained RNAP core class averages and reprojections. This dark ring represents a good alignment of the individual particles within the class average. If the particles were not so well aligned and represented the molecules at slightly different orientations the outline around the class average would be less defined. This dark ring is sometimes used to visually identify good class averages in UA-stained datasets, but again due to the smaller grain size is not always seen in UF-stained datasets.

The other difference between the PcrA-RNAP complex reconstruction and the RNAP core was the resolution of the final model. Although both data sets produced a normal distribution of particles across the class averages, the resolution of the complex was much lower than the core RNAP (Figure 5.3B & C). This lowered resolution of the PcrA-RNAP complex could also be due to the UF staining. The particle picking and assigning of particles to the correct class averages are essential for the derivation of a 3D model and this was found to be much harder with the PcrA-RNAP complex than with the RNAP core. The lower resolution of the PcrA-RNAP complex can be seen when compared to the RNAP core model (Figure 5.5). Overall, the electron density map of the PcrA-RNAP complex has a smoother surface so the structural elements seen in the RNAP core model are not clearly visible. Additionally, the face-on view of the PcrA-RNAP complex (Figure 5.5A) shows the claw region between the β and β' subunits has extra mass present, indicated by the red circle. This extra mass can also be seen in the overlay of the PcrA-RNAP complex and RNAP core structures in both the face-on (Figure 5.5C) and the side-on (Figure 5.5F) orientations, again highlighted by the red circle. This additional mass is again possibly due to the lower resolution of the

final PcrA-RNAP complex 3D reconstruction. It is possible that with further work and additional rounds of refinement a 3D model of the PcrA-RNAP complex at a lower resolution would remove this additional mass. Importantly, the current PcrA-RNAP complex model showed a large additional mass present on the upstream face, which corresponds to binding site of PcrA.

5.3.2 Docking of the protein structures into the 3D model

Preliminary docking was performed to show whether the extra mass would encompass both PcrA and RNAP. This was made possible due to previous work docking the RNAP homology structure into the RNAP core electron density map (Yang *et al*, 2009). The orientation of RNAP within the PcrA-RNAP complex electron density map was determined by firstly adding the RNAP homology model in the known orientation to fit in the RNAP core 3D model (Figure 5.6A). Next, the electron density map of the PcrA-RNAP complex was superimposed onto the core model (Figure 5.6B) and the core density map was removed (Figure 5.6C). This showed that the extra mass in the PcrA-RNAP complex model is on the upstream face of RNAP.
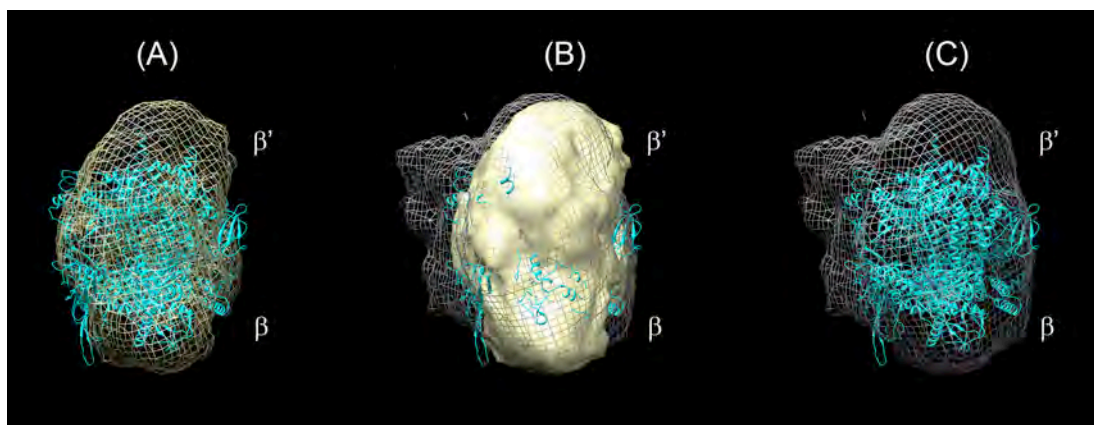
Figure 5.6: Docking of RNAP into the 3D reconstruction model. (A) RNAP core electron density map (yellow mesh) and RNAP homology model (blue ribbon). (B) PcrA-RNAP complex (grey mesh), RNAP core model (yellow surface rendered) and homology model (blue ribbon). (C) The PcrA-RNAP complex (grey mesh) and RNAP homology model (blue ribbon) in the correct orientation. RNAP is shown in the face-on orientation.

The crystal structure of PcrA (1PJR) was then fitted in the complex electron density map, along with RNAP (Figure 5.7), which showed it is feasible that PcrA could fit in the extra mass. Importantly, this model supported the *in vitro* results, which indicated that the site of interaction between the PcrA and RNAP was between both the β and β' subunits on the upstream face of RNAP.
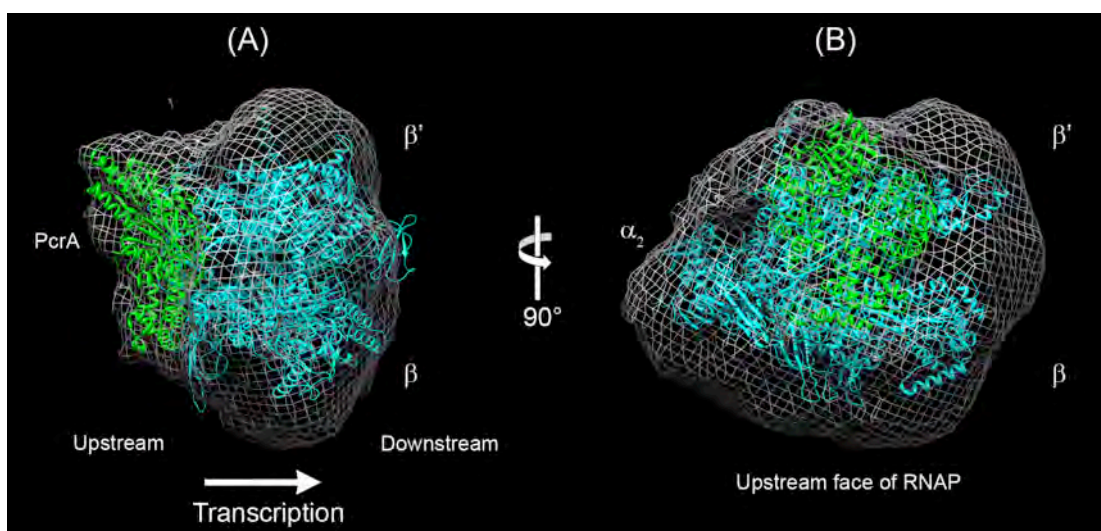
Figure 5.7 Docking of PcrA and RNAP into the 3D reconstruction model. (A) The RNAP model in the face-on orientation. (B) The RNAP model turned 90° showing the upstream face in the side-on orientation. The PcrA-RNAP complex electron density map is shown in grey mesh. RNAP homology model is cyan ribbon and PcrA is green ribbon structure.

The next step was to confirm the docking using the specific sites of interaction that were identified from far-Western blotting (Chapter 3). The interacting regions were highlighted on the docking model, as follows: PcrA $NTD_{85-189aa}$ – dark blue, $NTD_{189-310aa}$ – yellow, $β'_{1-102aa}$ – dark green and $β'_{228-310aa}$ – light pink, PcrA unstructured CTD – dark pink and $β_{1-400aa}$ – blue (Figure 5.8A). In order to show the highlighted regions of PcrA and RNAP could be interacting, the specific areas were magnified, the PcrA NTD-$β'$ subunit interaction in Figure 5.8B and the PcrA CTD-$β$ subunit interaction in Figure 5.8C. The mesh representing the 3D reconstruction model in Figure 5.8A was removed to make the images clearer.
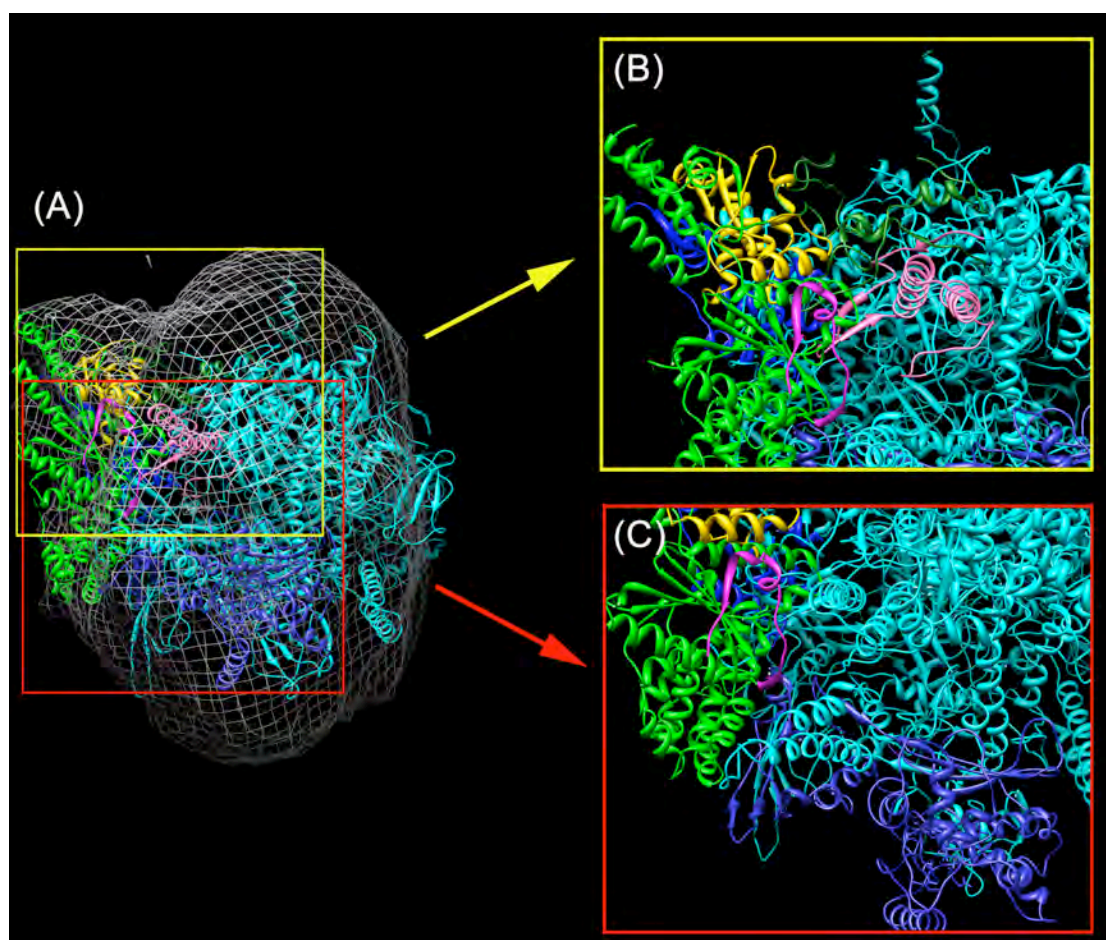
Figure 5.8: Docking of PcrA and RNAP into the 3D reconstruction model, including the interacting regions identified in the far-Western blots. (A) Complex in front-on orientation, the PcrA-RNAP complex model in grey mesh. (B) Zoomed in view of the PcrA NTD-$β'$ subunit interactions, as indicated by the yellow boxed region in A. PcrA $NTD_{85-189aa}$ – dark blue, $NTD_{189-310aa}$ – yellow. $β'_{1-102aa}$ – dark green and $β'_{228-310aa}$ – light pink. (C) Zoomed in view of the PcrA CTD-$β$ subunit interactions, as indicated by the red boxed region in A. PcrA unstructured CTD – dark pink, $β_{1-400aa}$ – blue. RNAP – cyan, PcrA – green.

When first considering the PcrA NTD-β' subunit interaction, this docking orientation showed that it was possible that the two regions of PcrA (NTD$_{85-189aa}$ – dark blue, NTD$_{189-310aa}$ – yellow) could interact with the two β' subunit fragments in dark green and pink (Figure 5.8B). Also the dark pink PcrA CTD region in this orientation looks like it may be contacting the β'$_{228-310aa}$ region in light pink (Figure 5.8B). This was found to not be the case in the far-Western blot (Figure 3.23) and it is important to note that the X-ray crystal structure of the PcrA CTD (652-739aa) has not been resolved. This modelling suggests that the PcrA CTD stretches down, away from the body of the PcrA structure to interact with the β subunit, in blue. This is particularly obvious in Figure 5.8C, where the PcrA CTD is looped and points down towards the bottom of RNAP, the addition of the unresolved end of PcrA would then be in reach of the interaction region in $\beta_{1-400aa}$.

5.4 Conclusion

In conclusion, this work successfully confirmed that the interaction occurs between PcrA and RNAP on the upstream face of RNAP. Although the resolution of the complex was too low to determine the structural elements involved in the protein-protein interaction, it was important to confirm the position of the interaction. The characterisation of this interaction permits speculation on the function of PcrA during transcription (General Discussion).

Chapter 6:

General Discussion

Chapter 6: General Discussion

6.1 Introduction

PcrA is an essential DNA helicase in Gram-positive bacteria, and has roles in plasmid replication and homologous recombination repair. An interaction between PcrA and RNAP was first identified in the examination of protein interactions during replication (Noirot-Gros *et al*, 2002a). The interaction was also identified in a yeast-two hybrid study that attempted to identify possible new transcription factors (Section 3.1.1). This work aimed to characterise the protein-protein interaction between PcrA and RNAP. The investigation of the interaction *in vitro* showed the interaction was more complex than originally expected. There were multiple sites of interaction identified between PcrA and both the β and β' subunits of RNAP (Section 3.2.2.2). Additional work then showed this interaction was conserved with a similar SF1 helicase UvrD and RNAP in *E. coli* (Section 3.2.3). Further characterisation of the interaction included mutagenic analysis of the PcrA CTD – β subunit interaction and the investigation of the effects the PcrA mutations had on binding to RNAP *in vivo* (Chapter 4). Finally, to confirm the position of the interaction between PcrA and RNAP identified using far-Western blotting, single particle reconstruction was used to reconstruct the PcrA-RNAP complex structure (Chapter 5).

6.2 Examining the PcrA-RNAP interaction

6.2.1 The PcrA RNAP interaction in *B. subtilis*

The affinity chromatography results showed the interaction between PcrA and RNAP was strong and was disrupted only after the addition of 1 M NaCl. This interaction was also shown to exist between PcrA$_{Gst}$ and RNAP. In contrast, the interaction between RNAP and GreA was disrupted by the addition of 0.4 M NaCl (Figure 3.14). Recent ChAP-chip (chromatin affinity precipitation coupled with DNA microarrays) studies have shown GreA is uniformly present throughout transcribed regions of DNA, which implies GreA is associated with RNAP through both transcription initiation and elongation (Kusuya *et al*, 2011). The fact that the interaction between PcrA and RNAP is stronger than that of RNAP and GreA suggested that the interaction is significant and required further investigation.

The far-Western blot was then used to identify the region of RNAP that interacts with PcrA. An unexpected and exciting result using the far-Western blot was the identification of two separate PcrA binding sites on RNAP (Figure 3.18). This showed the interaction was more complex than initially believed and the two binding sites between the proteins may be promoting the strong interaction. To determine if both of the PcrA sites of interaction on RNAP were utilised by a single PcrA molecule, further far-Western blotting was performed. In spite of the difficulties encountered with this section of the work (see Section 3.3.1), a model of the PcrA RNAP interaction was created using the regions identified in the far-Western blot (Section 5.3.2 & 6.2.3).

6.2.2 Mutagenic analysis of the PcrA CTD β subunit interaction

In an attempt to identify the specific amino acids involved in the interaction, a mutagenic yeast two-hybrid assay was performed on the PcrA CTD. The effect of the resulting mutations was then investigated using both *in vitro* and *in vivo* interaction assays. The *in vitro* studies showed that three of the four PcrA CTD mutations disrupted the interaction with the β subunit. The *in vivo* studies that followed aimed to examine the effect of disrupting the interaction in *B. subtilis* cells.

This work successfully identified amino acids were important in the PcrA CTD-β subunit interaction, it also showed that more work is required to understand this complex interaction. The reduced transformation efficiency caused by the introduction of the *pcrA* mutations *in vivo* suggested the disruption of that interaction was important, but further investigation of the mutant strains showed no observable phenotype. Initially it was believed the β subunit was the only site of interaction, which prompted the mutagenic analysis. The results of the tandem affinity purification on the mutant strains have shown that the PcrA interaction with the β' subunit is sufficient to form the PcrA-RNAP complex (discussed in Section 4.3). Although, there is still a possibility that the interaction with the PcrA NTD promotes the interaction between the PcrA CTD and β subunit and *in vivo* the full length mutant PcrA can then bind to both the β and the β' subunit.

In the future, when examining the sites of interaction between PcrA and RNAP, it will be important to identify important amino acids in both binding sites. One possible method that could be used is cross-linking protein interaction analysis. Cross-linking

experiments can be performed both *in vivo* and *in vitro* and therefore could also be performed on the *B. subtilis* mutant *pcrA* strains to conclusively show if the interaction was disrupted.

6.2.3 Modelling the PcrA-RNAP interaction

6.2.3.1 Single particle reconstruction of the PcrA-RNAP complex

The single particle reconstruction of the PcrA-RNAP complex was performed in order to confirm the position of the PcrA binding. The resultant 3D electron density map of the complex was a low resolution structure, as discussed previously (Section 5.3.1) but additional mass attributed to PcrA was shown on the upstream side of RNAP (Figure 5.5). Initial docking of the PcrA-RNAP complex into the 3D reconstruction model has been shown to accommodate both the RNAP and PcrA protein structures (Section 5.3.2, Figure 5.7) and confirms the binding regions identified by the far-Western blot are able to interact in this orientation (Figure 5.8). This docking was performed using the X-ray crystal structure of PcrA in the absence of DNA (PDB:1PJR), but there are additional crystal structures of PcrA available from the PDB, particularly one of PcrA in complex with DNA (PDB:3PJR). Modelling of the interaction using this PcrA structure and attempting to track the path of the DNA exiting RNAP and entering PcrA may provide additional information about the interaction and its biological relevance.

It is well established that PcrA is predominantly a 3'-5' DNA helicase and earlier reports of PcrA having bi-directional activity are currently being questioned (Section 1.6.3). Importantly, the crystal structure of *G. stearothermophilus* PcrA in complex with DNA indicates it is a 3'-5' unidirectional helicase. For that reason it is believed

that PcrA would be bound on the template DNA strand, while interacting with the upstream face of RNAP (Figure 6.1A). This would allow for PcrA to travel in the same direction as transcription and remain bound to RNAP. If PcrA was bound to the non-template strand and travelling in the 3'-5' direction, PcrA would be moving away from RNAP and would be unwinding the dsDNA exiting from RNAP (Figure 6.1B). This would mean a transient PcrA-RNAP interaction was occurring. The strength of the interaction suggests that this would be unlikely and instead the interaction would be ongoing. Additionally, the DNA unwinding presents a problem for subsequent rounds of transcription or replication, as it could promote the formation of R-loops, where an RNA-DNA hybrid forms between the RNA and the upstream DNA template strand. R-loops can block the progression of the following RNAP or replication fork.
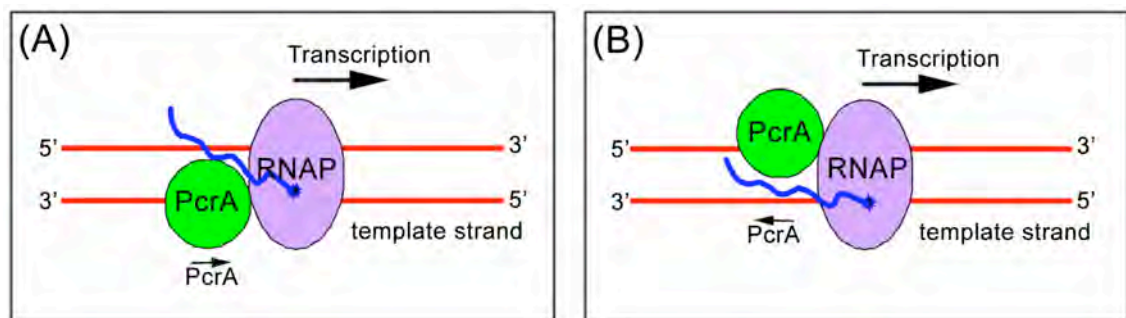
Figure 6.1: PcrA DNA binding during transcription. (A) PcrA could be bound to the template DNA strand and travelling in the same direction as transcription. (B) PcrA could be bound to the non-template strand and would then be travelling in the opposite direction of the transcription complex.

To examine the path of the DNA in the PcrA-RNAP interaction docking of the PcrA-RNAP complex was repeated. The PcrA protein structure in complex with DNA (PDB:3PJR) and the RNAP elongation complex homology model with DNA and RNA (Yang *et al*, 2009), were used in the same orientated as previously presented (Figure 5.7 & 5.8). PcrA is a highly flexible protein that undergoes conformational changes when bound to DNA and this change in the structure (Figure 6.2) affected the docking of

PcrA into the 3D electron density of the PcrA-RNAP complex reconstruction (Figure 6.3).
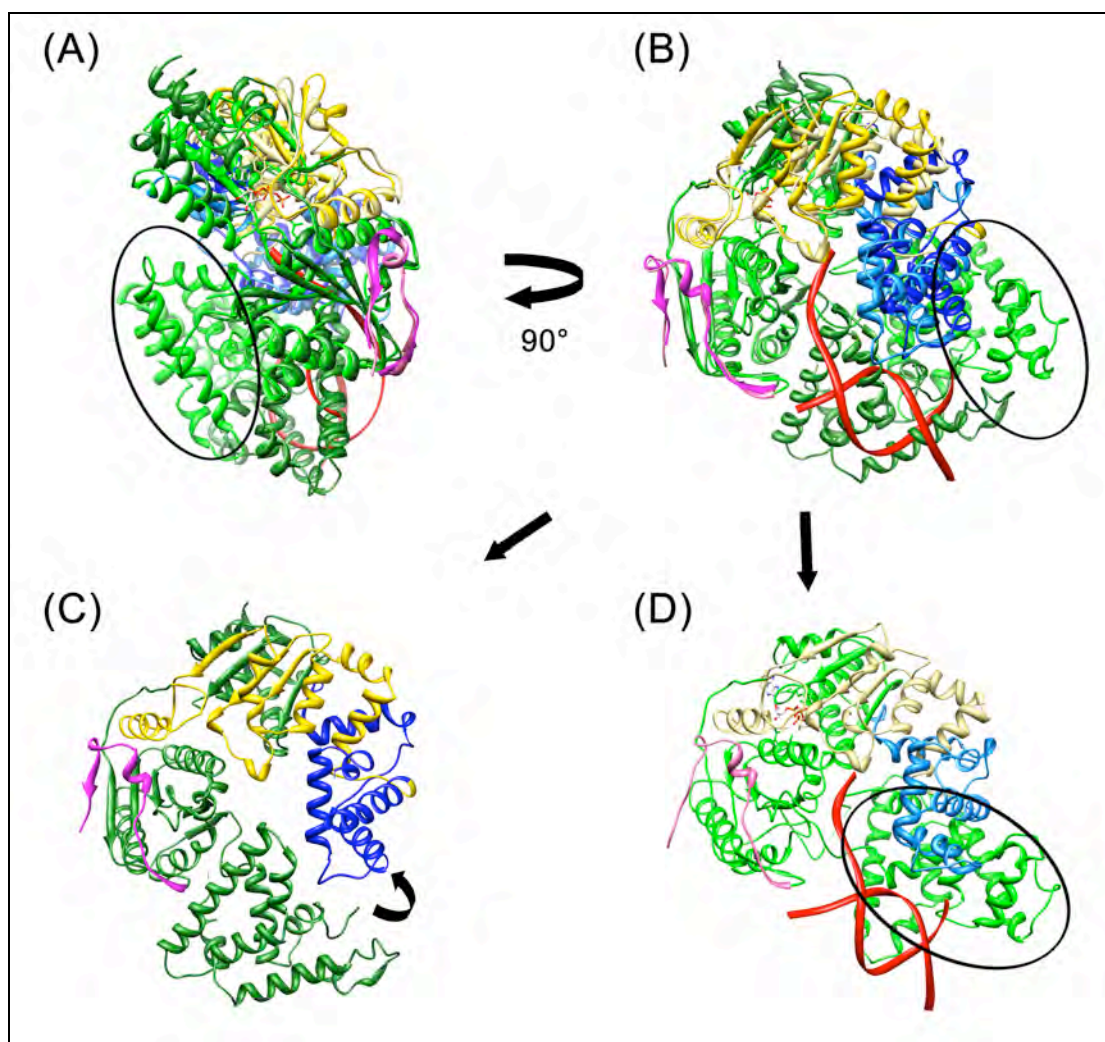
Figure 6.2: PcrA protein structure comparison between unbound (1PJR) and in complex with DNA (3PJR). (A) Overlay of the two PcrA structures, conformational change in the structure from binding DNA is circled. (B) The overlay of the two structures rotated 90° into the page to clearly show the regions that interact with RNAP (highlighted) are still well aligned. (C) Apo-PcrA in same orientation as in panel (B). (D) PcrA in complex with DNA in the same orientation as (B). Again, the change in conformation is circled. Apo-PcrA – dark green, PcrA in complex with DNA – light green, DNA – red. PcrA amino acids 85-189 – blue, 189-310 – yellow, and the region immediately adjacent to the unstructured CTD – pink. Black circle in panels (A), (B) and (D) show the region of PcrA that shifts after binding DNA, this is also indicated by the arrow in panel (C). (See also the additional pymol file provided – Appendix V: PcrA structural morph).

The two PcrA structures (with and without DNA) were aligned to identify the conformational changes that could affect the docking of PcrA-DNA complex into the 3D reconstruction model. The change in conformation is seen at the bottom of Figure 6.2A, indicated by the circle. When the structural alignment was rotated 90° it showed

the RNAP interacting regions (highlighted, blue, yellow & pink) were still well aligned (Figure 6.2B). The alignment of the RNAP interacting regions of PcrA in the two structures indicated the binding model (Figure 5.8) could be valid in the presence or absence of DNA. To more clearly see the conformational change the two structures were separated into dark green Apo-PcrA (Figure 6.2C) and light green PcrA-DNA complex (Figure 6.2D). The shift of the PcrA 2B domain upwards on binding DNA is indicated by the arrow in Figure 6.2C and is circled in Figure 6.2D. The shift can also be seen by comparing Figures 6.2C & D and can be seen in the supplementary file provided (PcrA structural morph, see Appendix V). The docking of PcrA into the 3D electron density of the complex (Figure 6.3A) showed this shifted 2B domain of PcrA in complex with DNA did not neatly fit into the additional mass (Figure 6.3B, black circle).
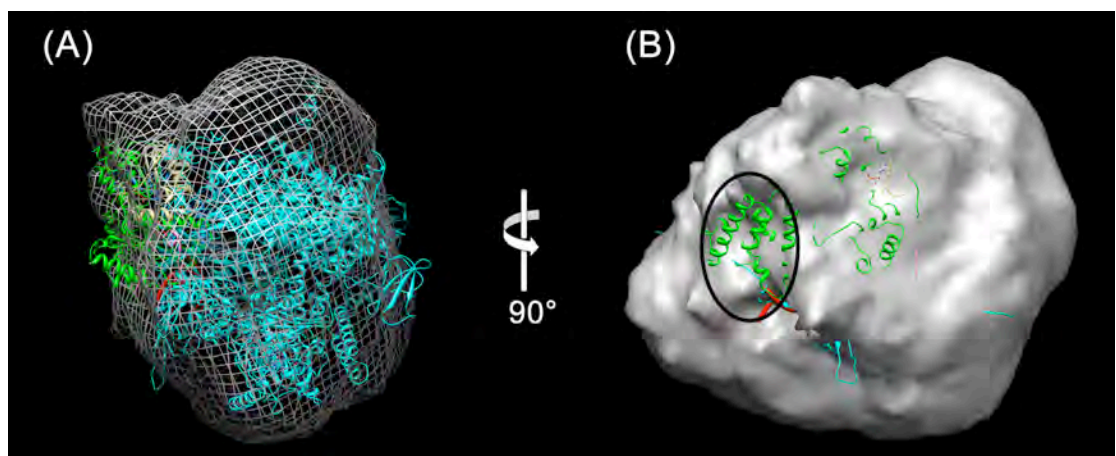
Figure 6.3: Docking of PcrA-DNA structure in the PcrA-RNAP 3D electron density. (A) Front on view of complex, with PcrA on upstream face of RNAP; the electron density is shown in mesh to permit visualisation of the PcrA and RNAP structures within the mass. (B) PcrA-RNAP complex in profile; the electron density is shown as surface rendered to highlight the region of PcrA that does not fit within the additional mass, as indicated by the black circle.

6.2.3.2 Examining the path of the DNA in the PcrA-RNAP interaction

As expected, when docking the 3PJR structure into the PcrA-RNAP complex 3D electron density the shifted 2B domain was seen to extend beyond the additional mass

attributed to PcrA. The results indicated it would be simpler to examine the path of DNA in the interaction without using the complex 3D electron density.

The path of the DNA through the PcrA-RNAP complex has been modelled in Figure 6.4. PcrA was positioned in the same orientation as in Figure 5.8. However, in this model the entry of the dsDNA (red) and the exit of the RNA (blue) from RNAP has been included (Figure 6.4A). The DNA in complex with PcrA is also shown and this orientation is looking down the middle of the DNA double helix (Figure 6.4A). The dsDNA exiting RNAP upstream is not shown in the RNAP protein structure. Therefore the schematic of the PcrA complex in the front-on orientation (Figure 6.4C) is an example of the possible path of the DNA, which indicates the dsDNA exiting RNAP would again be separated by PcrA helicase activity. This model was then rotated 90° to show the RNAP upstream face in profile (Figure 6.4B &D). In this orientation the DNA in complex with PcrA is seen more clearly. If PcrA was bound to both RNAP and DNA in this orientation, PcrA would unwind DNA in the direction opposite to which transcription was proceeding. In Figure 6.4 B and D this is shown by the separation of the dsDNA in PcrA, at the front of the complex, indicated by the arrows (Figure 6.4B & D). This is where the DNA exiting RNAP would enter PcrA. The unwinding of the upstream DNA would not occur if PcrA was acting as a DNA translocase.

Figure 6.4: Modelling of the PcrA-RNAP interaction to examine the path of DNA. (A) Ribbon protein structures of the RNAP elongation complex and PcrA in the face-on orientation. (B) Ribbon protein structures of RNAP elongation complex and PcrA in side-on orientation. (C) A simplified schematic of panel (A) with the path of the DNA through RNAP and PcrA in the face-on orientation. (D) Simplified schematic of panel (B) the path of DNA through RNAP and PcrA in the side-on orientation. Cyan – RNAP, green – PcrA, red – DNA, blue - RNA, star – RNAP active site, arrow – DNA unwinding by PcrA.
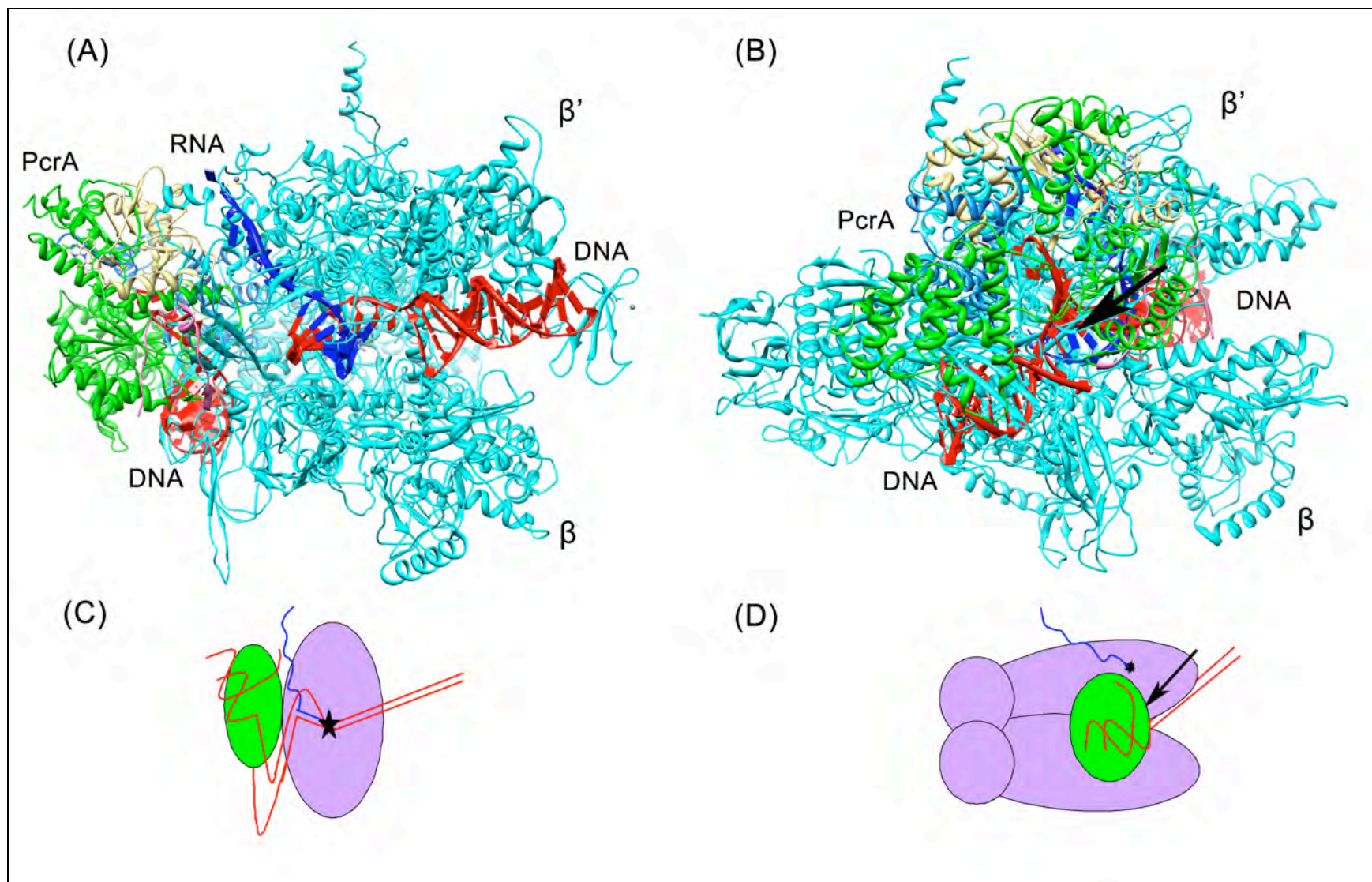
As discussed earlier (Figure 6.1) it was believed that PcrA would be travelling in the same direction as transcription, but initial modelling (Figure 6.4) of the interacting regions in the presence of DNA indicated that this was unlikely, particularly if PcrA was exhibiting helicase activity. Additional modelling of the interacting regions was performed to determine if manipulation of the structures would allow for the DNA to travel from RNAP into PcrA and also allow for PcrA to travel in the direction of transcription.

To accommodate the DNA travelling from RNAP into PcrA and having PcrA travelling and unwinding dsDNA in the direction of transcription, the modelling (Figure 6.5) changed the orientation of PcrA. In this orientation the regions of the PcrA NTD shown to interact with the β' subunit could be accommodated, but the PcrA CTD region was now situated too far away from RNAP to interact with the β subunit (Figure 6.5A & B). This modelling is inconsistent with the binding results, which has shown the interaction between the PcrA CTD and the β subunit is very strong *in vitro* (Figure 4.4B).
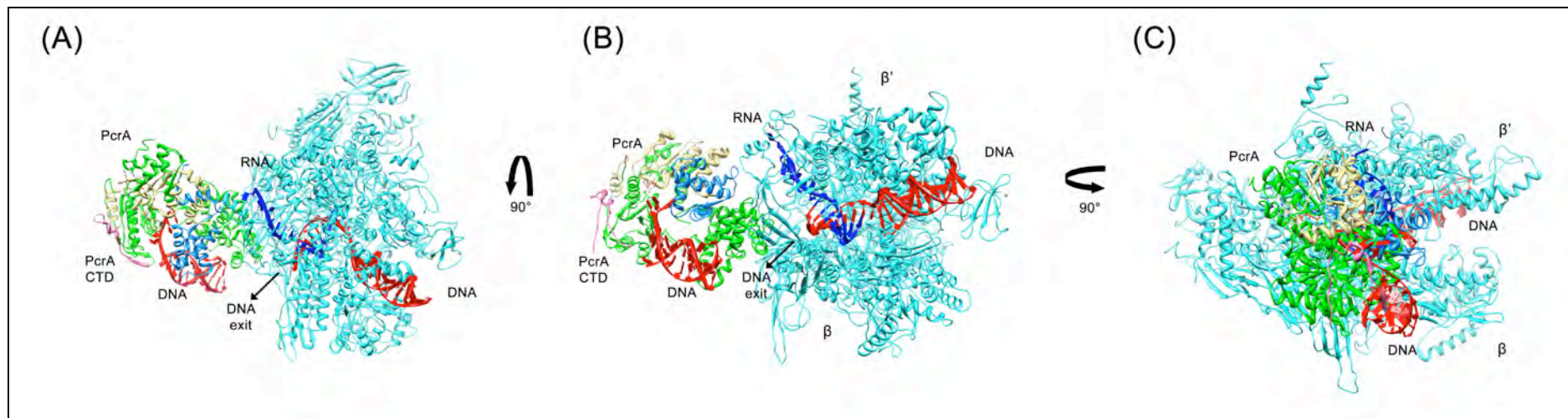
Figure 6.5: Alternative docking of PcrA RNAP and DNA. The PcrA NTD interacting regions 85-189 aa in blue and 189-310 aa in yellow could still be interacting with the β subunit but the unstructured CTD in pink is on opposite side. (A) PcrA-RNAP complex, viewing the complex from above. (B) PcrA-RNAP complex from panel (A) rotated 90° up, now showing the face-on orientation. (C) PcrA-RNAP complex from panel (B) rotated 90°, showing the upstream face in profile. Cyan – RNAP, green – PcrA, red – DNA, blue – RNA.

Although the CTD structure has not been resolved it is highly unlikely that this current model conformation would represent the PcrA binding position, because the CTD does not interact with the β subunit. Additionally, when PcrA is turned in this orientation it doesn't fit into the PcrA-RNAP complex 3D reconstruction (not shown). In this orientation the distance between the PcrA CTD (Figure 6.5, pink) and the β subunit was approximately 106 Å. The unresolved C-terminus of *G. stearothermophilus* PcrA is 72 amino acids and if this region contained a β strand it could be stretched to approximately 230 Å. If the C-terminus was α helical the C-terminal portion would be approximately 100 Å long. Although the CTD was unresolved it is expected to contain some secondary structure and therefore the stretched CTD theoretically could reach the β subunit.

The strong interaction between PcrA and RNAP has been conclusively proven throughout this work. The modelling of this interaction has shown (Figure 5.7 & 5.8) the identified binding regions of PcrA and RNAP are positioned to allow contact between the two proteins. The additional hypothetical modelling of the DNA (Figure 6.4 & 6.5) in this interaction has not provided a convincing model. This indicates PcrA may not be binding to both RNAP and DNA at the same time. Instead, PcrA could use RNAP as a chaperone or a marker for identifying DNA lesions, where PcrA firstly interacts with RNAP and subsequently binds DNA to stimulate DNA repair. Additionally PcrA may not act as a helicase, instead only possessing translocase activity, which would eliminate issues such as R-loop formation if PcrA was unwinding upstream dsDNA.

PcrA has shown to be a processive ssDNA translocase (Dillingham *et al*, 2000; Niedziela-Majka *et al*, 2007), but it hasn't been reported that PcrA can translocate on dsDNA. If PcrA was translocating on the upstream dsDNA behind RNAP during transcription it could solve problems of the direction of the travelling PcrA and would stop unwound DNA from forming R-loops.

The investigations into proteins containing helicase motifs have shown that a number of the proteins with designated helicase motors cannot unwind duplex DNA. It has been proposed that the lack of unwinding is due to ssDNA not being required to carry out the cellular process of the enzyme (Stanley *et al*, 2006). PcrA shows a lack of processive helicase activity *in vitro* (Niedziela-Majka *et al*, 2007) and this could indicate that the PcrA is a dsDNA translocase. A study into remodeler ATPases (Saha *et al*, 2002) used the separation of the helicase and translocase activity of PcrA performed by Soultanas *et al.* (2000) as a model to reason the remodel ATPase enzymes possess dsDNA and ssDNA translocation. dsDNA translocases can contact and directionally convey one strand of DNA, while the other strand is passively transported through minimal contacts that play a supplementary clamp role (Stanley *et al*, 2006). This would indicate the transition from a ssDNA translocase to a dsDNA translocase for PcrA may be possible. It is also possible that PcrA requires helicase activity for some cellular roles, like plasmid replication but in other roles, such as during transcription, only translocation is required. The other reason for PcrA possessing dsDNA translocase activity is the comparison between PcrA and Mfd (discussed in Section 6.3.3), as Mfd uses dsDNA translocation to removed stalled RNAP (Mahdi *et al*, 2003).

Although more work is required to confirm PcrA is a dsDNA translocase, it is believed that this activity could occur during the PcrA-RNAP interaction. The dsDNA translocation activity of PcrA supports the model of interaction presented (Figure 5.8) and would allow for PcrA to remove stalled RNAP (Section 6.3). The alternative could be that PcrA is translocating the ssDNA exiting RNAP, before the dsDNA is re-formed and in that way could act to remove stalled RNAP.

6.2.4 The interaction between UvrD and RNAP in *E. coli*

Another exciting result from the investigation of the PcrA RNAP interaction was the identification of the interaction between UvrD and RNAP (discussed in Section 3.3.2). The interaction between PcrA and RNAP had been reported previously (Noirot-Gros *et al*, 2002; Delumeau *et al*, 2011), but no similar result has been reported for UvrD. This was despite numerous studies of transcription complexes in Gram-negative bacteria (for example see Butland *et al*, 2005) and therefore this study has identified a novel interaction (Section 3.2.3). The interaction detected between UvrD and RNAP was also a strong interaction and it was surprising that this interaction had not been detected previously. On closer inspection of the reported *E. coli* protein network (Butland *et al*, 2005) an interaction between UvrD and the β subunit of RNAP (*rpoB*) was detected using liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Butland, *et al.* 2005, Supplementary Table 1), which would be consistent with an interaction with RNAP similar to that observed with PcrA. This result was not validated by the authors and it is possible that the interaction was discarded as background because UvrD is an abundant DNA binding protein.

There are many reports of PcrA and UvrD having overlapping functions in Gram-positive and Gram-negative bacteria and additionally that the proteins can be interchanged in certain roles with no adverse effects. The expression of *pcrA* in *E. coli* restored the viability of the double knockout *uvrD rep* mutant (Petit *et al*, 1998). Other examples of PcrA and UvrD sharing functions include the removal of the Tus replication terminator protein and the dismantling of RecA filaments (Veaute *et al*, 2005; Bidnenko *et al*, 2006; Lestini & Michel, 2007; Park *et al*, 2010). Importantly, Rep cannot dismantle RecA filaments but inhibits their formation and has no effect on preformed filaments (Myong *et al*, 2005; Veaute *et al*, 2005). This, along with the similar cellular abundance levels of UvrD and PcrA (Section 3.1.3) supports UvrD being the SF1 helicase that would interact with RNAP in Gram-negative bacteria.

6.3 Roles for PcrA during transcription

The characterisation of the PcrA-RNAP interaction has provided information on conformational constraints of the interaction, which can then be used to hypothesise likely functions of PcrA as a result of this interaction.

During transcription many events can cause elongating RNAP complexes to stall on DNA. These events can prevent the complete transcription of RNA and can also adversely affect other essential processes, such as replication fork progression. The resolution of the stalled RNAP complex is vitally important in bacterial cells and therefore there are a number of ways that stalled RNAP can be resolved. The different events that cause RNAP to stall can also influence the method used to remove RNAP. As a helicase that has been shown to remove proteins bound to DNA and now

additionally to interact with RNAP, PcrA may function to aid the removal of blocked RNAP.

6.3.1 Removing stalled RNAP at DNA-protein cross-links

DNA-protein cross-links (DPCs) can be induced by chemicals and/or radiation and can occur on DNA with different DNA-binding proteins *in vivo*. When DPCs were artificially created *in vitro* they were shown to block transcription elongation by *E. coli* RNAP, but the effect of DPCs on transcription *in vivo* are still unknown (Som & Friedman, 1994). Transcription complexes blocked by DNA damage sometimes can be resolved by Mfd (Selby & Sancar, 1995), but Mfd cannot act on RNAP stalled by DPCs (Kuo *et al*, 2010). The transfer-messenger RNA (tmRNA) translational quality control system was implicated in the processing of transcription/translation complexes stalled at DPCs in *E. coli* (Kuo *et al*, 2010). The tmRNA pathway's main function is to allow the completion of mRNA translation lacking a stop codon and the induction of the degradation of aberrant mRNA (Yamamoto *et al*, 2003; Mehta *et al*, 2006). The tmRNA pathway is responsible for releasing the ribosomes from the mRNA at the stalled transcription complex but the RNAP must also be removed from the DNA and currently the mechanism for this is unknown. PcrA may be a better candidate than Mfd to remove RNAP stalled at DPCs. This work has shown that there is a strong interaction between PcrA and RNAP (Section 3.2.2.1) and the far-Western blotting had identified one of the PcrA binding sites to be the same region of the β subunit of RNAP as the Mfd binding site (Section 3.2.2.2). The association between RNAP and PcrA is also tighter than that with Mfd *in vivo,* which was shown by the isolation of RNAP complexes. PcrA was identified in one hundred percent of isolated RNAP complexes

but Mfd was detected in only thirty eight percent of samples (Delumeau *et al*, 2011). Other investigations into the protein-protein interactions of the transcription complex have not identified the Mfd-RNAP interaction (Butland *et al*, 2005; Verma *et al*, 2007; Lee *et al*, 2008).

6.3.2 Removing stalled RNAP at replication-transcription collision sites

When transcription complexes collide with replication forks it can result in replication arrest, which can lead to genomic instability and ultimately cell death (Mirkin & Mirkin, 2007; Pomerantz & O'Donnell, 2010). There are two types of replication-transcription collisions which occur *in vivo*. Firstly, co-directional collisions, where RNAP is travelling in the same direction as the replication fork and secondly, head-on collisions, where the complexes are travelling in opposite directions (Figure 6.6A, Pomerantz & O'Donnell, 2010). To overcome this there are mechanisms that can aid the progression of the replication fork past the transcription complex (Pomerantz & O'Donnell, 2010). One mechanism utilises transcription factors, such as GreA and Mfd to process stalled RNAP complexes (Selby & Sancar, 1993a; Rudolph *et al*, 2007), which if not removed could form a RNAP array (Figure 6.6B). An RNAP array is a number of stalled RNAP complexes on the same region of DNA, and the replication fork cannot progress through this RNAP array, leading to replication fork arrest (Figure 6.6B, Pomerantz & O'Donnell, 2010). Recently, helicases such as UvrD or Rep, in conjunction with DinG in *E. coli* have also been shown to promote the removal of the stalled RNAP complexes and to promote the progression of the replication fork through protein roadblocks (Figure 6.6C & D, Guy & Roten, 2004; Heller & Marians, 2007; Boubakri *et al*, 2010).
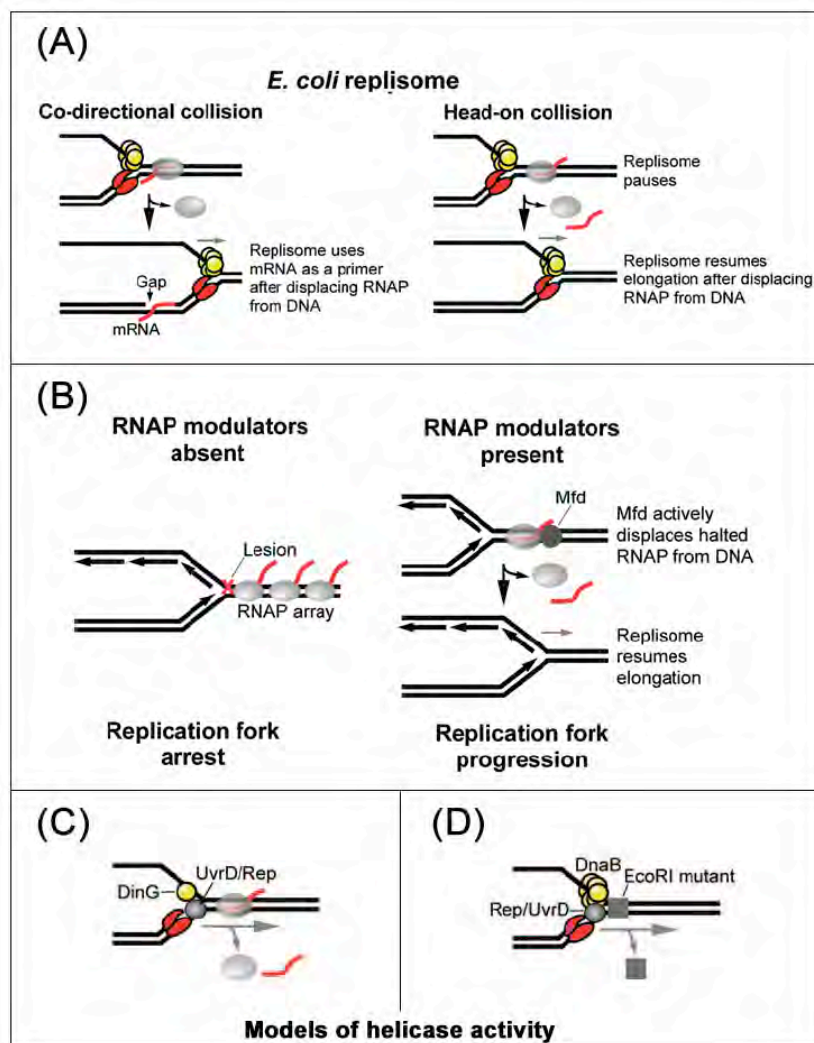
Figure 6.6: Effects of transcription-replication collisions. (A) Co-directional and head-on models of replisome-RNAP collisions performed *in vitro*. (B) Modulators of RNAP activity (such as Mfd) prevent an RNAP array from forming, which prevents a head-on collision with the replication fork that can result in replication fork arrest. (C & D) DNA helicases can promote replication through transcription complexes. (C) In *E. coli* DinG acts on the lagging strand with either UvrD or Rep acting on the leading strand to promote replication past the transcription complex. (D) Rep or UvrD have also been shown to act on the leading strand opposite DnaB to promote progression through a protein roadblock (adapted from (Pomerantz & O'Donnell, 2010).

The majority of the experimental work investigating the effects of the transcription-replication complex collisions has been carried out in *E. coli*. It has been established that PcrA is a SF1 homologue of UvrD in Gram-positive bacteria, which has previously been substituted for UvrD in promoting replication through protein roadblocks (Bidnenko *et al*, 2006). Additional observations also suggest that the SF1 helicases may not have to interact directly with the replisome to remove the protein roadblocks ahead

of the replication fork (Pomerantz & O'Donnell, 2010). Although this requires further investigation, the interaction between both PcrA and UvrD and RNAP carried out in this work could support the removal of RNAP, without interacting with the replisome.

6.3.3 Removing stalled RNAP at DNA lesions

Mfd uses its translocase activity to remove RNAP that has stalled at DNA lesions and then acts to recruit the components of the nucleotide excision repair pathway to the site of the DNA damage (Selby & Sancar, 1993a; Selby & Sancar, 1993b). The interaction between Mfd and RNAP promotes the transcription-coupled repair activity of Mfd and is required for Mfd to recognise DNA damage (Westblade *et al*, 2010).

The role of Mfd may also be performed by PcrA, as previously discussed (Section 4.3). This is thought to be a possible role for PcrA due to the similar binding site of PcrA and Mfd on RNAP (Section 4.2.2.2, Deaconescu *et al*, 2006; Westblade *et al*, 2010). Another indication that PcrA may carry out transcription-coupled repair is the interaction between PcrA and nucleotide excision repair (NER) proteins, UvrB and C (Section 1.6.5; Manelyte *et al*, 2009).

The RNAP β-subunit region of interaction with Mfd was identified by yeast and bacterial two-hybrid analyses (Deaconescu *et al*, 2006). The X-ray crystal structure of the Mfd-RNAP interacting domain (RID) in complex with the RNAP β1 domain then provided details of this protein interaction and identified specific amino acids involved in the interaction (Westblade *et al*, 2010). The structural details of the protein complex can provide a basis to examine the function of Mfd, as a result of the interaction

(Westblade *et al*, 2010). The characterisation of the PcrA RNAP interaction performed in this work aimed to provide information which could be useful in the investigation of the function of PcrA in transcription. It would be interesting to investigate if the β subunit amino acids identified in the Mfd interaction are also involved in the PcrA interaction, which could be done using similar mutagenic analysis as utilised in Chapter 4. If the amino acids involved in the PcrA and Mfd interaction with β subunit were shared, it would provide additional support to PcrA and Mfd having overlapping functions in removing RNAP stalled at sites of DNA lesions.

6.3.4 Model of PcrA roles

Due to the upstream binding position of PcrA on RNAP, with respect to the direction of transcription it is believed PcrA could remove stalled RNAP. As discussed events that can stall RNAP include DNA-protein cross-linking (DPCs, Figure 6.7A) and DNA lesions (Figure 6.7B). The removal and repair of DPCs may be performed by the nucleotide excision repair (NER) pathway, which has been implicated in the excision of the DPC-containing DNA by UvrABC nuclease (Minko *et al*, 2002), followed by DNA repair. If PcrA removed RNAP stalled at DPCs (Figure 6.7A) it could recruit the UvrABC through interactions with UvrB and UvrC. PcrA would have the same role acting on RNAP stalled at DNA lesions (Figure 6.7B). PcrA could also prevent adverse affects of replication-transcription collisions by removing RNAP (Figure 6.7C), as already discussed (Section 6.3.2).
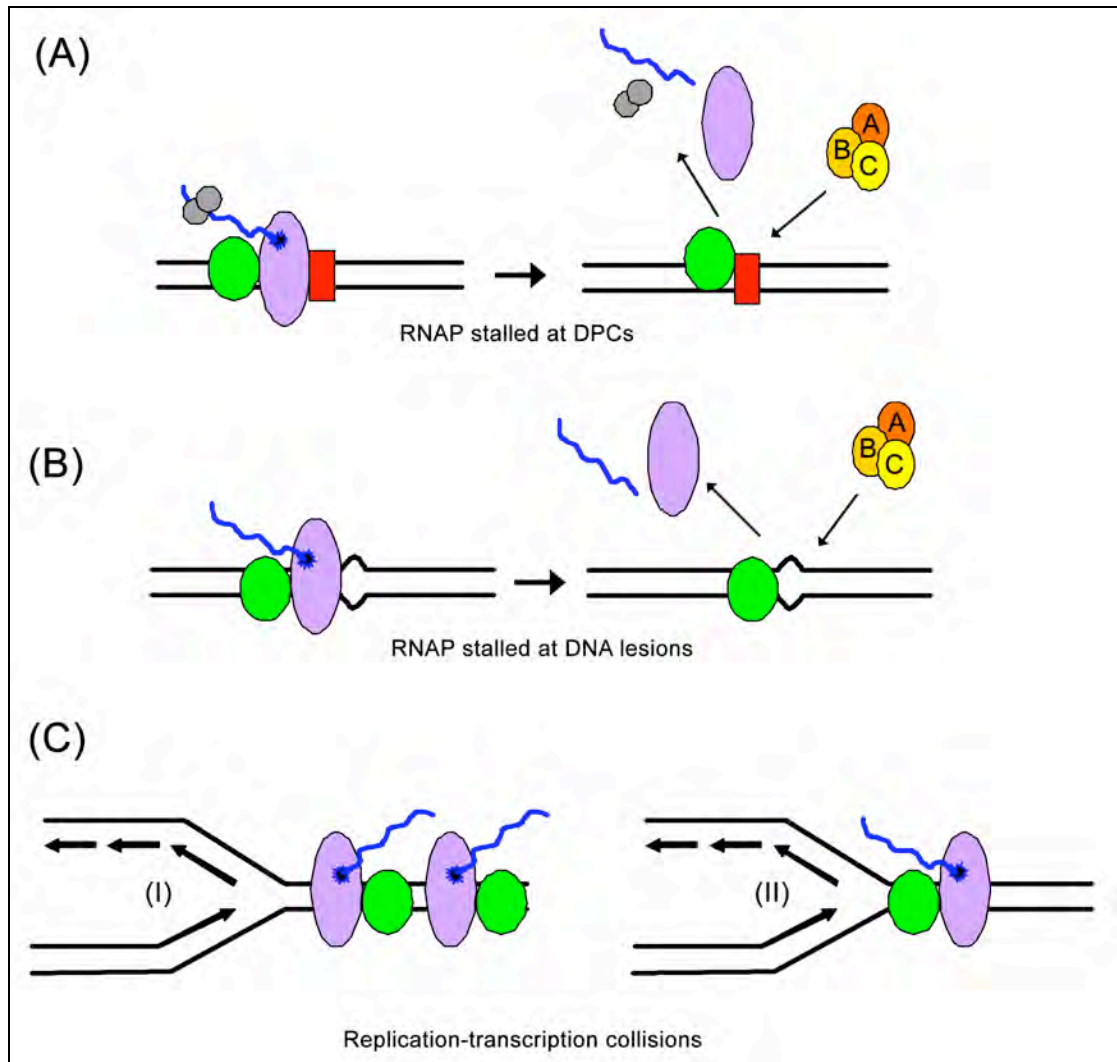
Figure 6.7: Possible roles of PcrA during transcription. PcrA could be removing stalled RNAP. Events that may cause RNAP to stall include (A) DPCs (DNA-protein cross-links) and (B) DNA lesions. PcrA could remove RNAP and recruit NER enzyme UvrABC. (C) The collision between replication and transcription could also be prevented by PcrA removing transcription complexes (I) RNAP transcribing in opposite direction to replication (II) RNAP transcribing the same direction as replication. RNAP – purple oval, PcrA – green circle, DNA – black, RNA – blue. DPCs – red rectangle, ribosome – grey circles and UvrABC nuclease – orange and yellow circles.

The removal of stalled RNAP can also be carried out by other transcription factors and even cooperation between subsequent transcribing RNAP (Epshtein & Nudler, 2003; Epshtein *et al*, 2003). There is a level of redundancy in essential cell processes and therefore it is possible PcrA can also function to remove stalled RNAP.

6.4 Conclusions

The characterisation of the interaction between PcrA and RNAP has provided important information on the position of the interaction and lead to the proposal of roles for PcrA during transcription. The work has shown there is a strong interaction between PcrA and RNAP and has identified multiple binding sites on both RNAP and PcrA. The presence of multiple binding sites indicated the interaction between the PcrA and RNAP was a complex interaction. Mutagenic analysis showed that the disruption of one binding site still allowed for the complex to form *in vivo* in *B. subtilis*. The PcrA-RNAP structure was examined using single particle reconstruction, which supported the sites of interactions identified throughout this work. The interaction is stronger than that between RNAP and important transcription factors and is also conserved between UvrD and RNAP in the Gram-negative *E. coli*. This indicates the interaction has a functional significance during transcription and the investigation of the possible roles of PcrA would be the exciting progression of this work in the future.

References

References

Abdelhaleem M. (2010) Helicases: an overview. *Methods Mol Biol* 587: 1-12.

Achberger E. C., Hilton M. D. & Whiteley H. R. (1982) The effect of the delta subunit on the interaction of *Bacillus subtilis* RNA polymerase with bases in a SP82 early gene promoter. *Nucleic Acids Res* 10: 2893-2910.

Alonso J. C., Tailor R. H. & Luder G. (1988) Characterization of recombination-deficient mutants of *Bacillus subtilis*. *J Bacteriol* 170: 3001-3007.

Amundsen S. K. & Smith G. R. (2003) Interchangeable parts of the *Escherichia coli* recombination machinery. *Cell* 112: 741-744.

Anand S. P., Chattopadhyay A. & Khan S. A. (2005) The PcrA3 mutant binds DNA and interacts with the RepC initiator protein of plasmid pT181 but is defective in its DNA helicase and unwinding activities. *Plasmid* 54: 104-113.

Anand S. P. & Khan S. A. (2004) Structure-specific DNA binding and bipolar helicase activities of PcrA. *Nucleic Acids Res* 32: 3190-3197.

Anand S. P., Mitra P., Naqvi A. & Khan S. A. (2004) *Bacillus anthracis* and *Bacillus cereus* PcrA helicases can support DNA unwinding and in vitro rolling-circle replication of plasmid pT181 of *Staphylococcus aureus*. *J Bacteriol* 186: 2195-2199.

Anand S. P., Zheng H., Bianco P. R., Leuba S. H. & Khan S. A. (2007) DNA helicase activity of PcrA is not required for the displacement of RecA protein from DNA or inhibition of RecA-mediated strand exchange. *J Bacteriol* 189: 4502-4509.

Arthur H. M. & Eastlake P. B. (1983) Transcriptional control of the uvrD gene of *Escherichia coli*. *Gene* 25: 309-316.

Artsimovitch I. & Landick R. (2002) The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell* 109: 193-203.

Ayora S., Rojo F., Ogasawara N., Nakai S. & Alonso J. C. (1996) The Mfd protein of *Bacillus subtilis* 168 is involved in both transcription-coupled DNA repair and DNA recombination. *J Mol Biol* 256: 301-318.

Bai L., Santangelo T. J. & Wang M. D. (2006) Single-molecule analysis of RNA polymerase transcription. *Annu Rev Biophys Biomol Struct* 35: 343-360.

Banerjee S., Chalissery J., Bandey I. & Sen R. (2006) Rho-dependent transcription termination: more questions than answers. *J Microbiol* 44: 11-22.

Barker M. M. & Gourse R. L. (2001) Regulation of rRNA transcription correlates with nucleoside triphosphate sensing. *J Bacteriol* 183: 6315-6323.

Belogurov G. A., Sevostyanova A., Svetlov V. & Artsimovitch I. (2010) Functional regions of the N-terminal domain of the antiterminator RfaH. *Mol Microbiol* 76: 286-301.

Bianco P. R., Tracy R. B. & Kowalczykowski S. C. (1998) DNA strand exchange proteins: a biochemical and physical comparison. *Front Biosci* 3: D570-603.

Bidnenko V., Lestini R. & Michel B. (2006) The *Escherichia coli* UvrD helicase is essential for Tus removal during recombination-dependent replication restart from Ter sites. *Mol Microbiol* 62: 382-396.

Bird L. E., Brannigan J. A., Subramanya H. S. & Wigley D. B. (1998) Characterisation of *Bacillus stearothermophilus* PcrA helicase: evidence against an active rolling mechanism. *Nucleic Acids Res* 26: 2686-2693.

Bork P. & Koonin E. V. (1993) An expanding family of helicases within the 'DEAD/H' superfamily. *Nucleic Acids Res* 21: 751-752.

Borukhov S., Lee J. & Laptenko O. (2005) Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol Microbiol* 55: 1315-1324.

Borukhov S. & Nudler E. (2003) RNA polymerase holoenzyme: structure, function and biological implications. *Curr Opin Microbiol* 6: 93-100.

Borukhov S. & Nudler E. (2008) RNA polymerase: the vehicle of transcription. *Trends Microbiol* 16: 126-134.

Borukhov S., Sagitov V. & Goldfarb A. (1993) Transcript cleavage factors from *E. coli*. *Cell* 72: 459-466.

Boubakri H., de Septenville A. L., Viguera E. & Michel B. (2010) The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units *in vivo*. *EMBO J* 29: 145-157.

Brendel V., Hamm G. H. & Trifonov E. N. (1986) Terminators of transcription with RNA polymerase from *Escherichia coli*: what they look like and how to find them. *J Biomol Struct Dyn* 3: 705-723.

Brendza K. M., Cheng W., Fischer C. J., Chesnik M. A., Niedziela-Majka A. & Lohman T. M. (2005) Autoinhibition of *Escherichia coli* Rep monomer helicase activity by its 2B subdomain. *Proc Natl Acad Sci U S A* 102: 10076-10081.

Butland G., Peregrin-Alvarez J. M., Li J., Yang W., Yang X., Canadien V., Starostine A., Richards D., Beattie B., Krogan N., Davey M., Parkinson J., Greenblatt J. & Emili A. (2005) Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* 433: 531-537.

Caramori T. & Galizzi A. (1998) The UP element of the promoter for the flagellin gene, hag, stimulates transcription from both SigD- and SigA-dependent promoters in *Bacillus subtilis*. *Mol Gen Genet* 258: 385-388.

Cardinale C. J., Washburn R. S., Tadigotla V. R., Brown L. M., Gottesman M. E. & Nudler E. (2008) Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science* 320: 935-938.

Chang T. L., Naqvi A., Anand S. P., Kramer M. G., Munshi R. & Khan S. A. (2002) Biochemical characterization of the *Staphylococcus aureus* PcrA helicase and its role in plasmid rolling circle replication. *J Biol Chem* 277: 45880-45886.

Chaudhuri R. R., Allen A. G., Owen P. J., Shalom G., Stone K., Harrison M., Burgis T. A., Lockyer M., Garcia-Lara J., Foster S. J., Pleasance S. J., Peters S. E., Maskell D. J.

& Charles I. G. (2009) Comprehensive identification of essential *Staphylococcus aureus* genes using Transposon-Mediated Differential Hybridisation (TMDH). *BMC genomics* 10: 291.

Cheng W., Brendza K. M., Gauss G. H., Korolev S., Waksman G. & Lohman T. M. (2002) The 2B domain of the *Escherichia coli* Rep protein is not required for DNA helicase activity. *Proc Natl Acad Sci U S A* 99: 16006-16011.

Cheng W., Hsieh J., Brendza K. M. & Lohman T. M. (2001) E. coli Rep oligomers are required to initiate DNA unwinding in vitro. *J Mol Biol* 310: 327-350.

Ciampi M. S. (2006) Rho-dependent terminators and transcription termination. *Microbiology* 152: 2515-2528.

Costes A., Lecointe F., McGovern S., Quevillon-Cheruel S. & Polard P. (2010) The C-terminal domain of the bacterial SSB protein acts as a DNA maintenance hub at active chromosome replication forks. *PLoS Genet* 6: e1001238.

Cramer P. (2002) Multisubunit RNA polymerases. *Curr Opin Struct Biol* 12: 89-97.

Dao V. & Modrich P. (1998) Mismatch-, MutS-, MutL-, and helicase II-dependent unwinding from the single-strand break of an incised heteroduplex. *J Biol Chem* 273: 9202-9207.

Davies K. M., Dedman A. J., van Horck S. & Lewis P. J. (2005) The NusA:RNA polymerase ratio is increased at sites of rRNA synthesis in *Bacillus subtilis*. *Mol Microbiol* 57: 366-379.

Davis S. E., Mooney, R.A., Kanin, E.I., Grass, J., Landick, R. & Ansari, A.Z. (2011) Mapping *E. coli* RNA polymerase and associated trasncription factors and identifying promoters genome-wide. *Methods Enzymol* 498: 449-471.

Deaconescu A. M., Chambers A. L., Smith A. J., Nickels B. E., Hochschild A., Savery N. J. & Darst S. A. (2006) Structural basis for bacterial transcription-coupled DNA repair. *Cell* 124: 507-520.

Delumeau O., Lecointe F., Muntel J., Guillot A., Guedon E., Monnet V., Hecker M., Becher D., Polard P. & Noirot P. (2011) The dynamic protein partnership of RNA polymerase in *Bacillus subtilis*. *Proteomics* 11: 2992-3001.

DeVito J. & Das A. (1994) Control of transcription processivity in phage lambda: Nus factors strengthen the termination-resistant state of RNA polymerase induced by N antiterminator. *Proc Natl Acad Sci U S A* 91: 8660-8664.

Dillingham M. S., Wigley D. B. & Webb M. R. (2000) Demonstration of unidirectional single-stranded DNA translocation by PcrA helicase: measurement of step size and translocation speed. *Biochemistry* 39: 205-212.

Dillingham M. S., Wigley D. B. & Webb M. R. (2002) Direct measurement of single-stranded DNA translocation by PcrA helicase using the fluorescent base analogue 2-aminopurine. *Biochemistry* 41: 643-651.

Dittrich M. & Schulten K. (2006) PcrA helicase, a prototype ATP-driven molecular motor. *Structure* 14: 1345-1353.

Doherty G. P. (2007) The composition of transcription complexes in *Bacillus subtilis*. University of Newcastle, Australia.

Driscoll R. & Cimprich K. A. (2009) HARPing on about the DNA damage response during replication. *Genes Dev* 23: 2359-2365.

Ebright R. H. (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol* 304: 687-698.

Eick D., Wedel A. & Heumann H. (1994) From initiation to elongation: comparison of transcription by prokaryotic and eukaryotic RNA polymerases. *Trends Genet* 10: 292-296.

Eisen J. A. & Hanawalt P. C. (1999) A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res* 435: 171-213.

Eisen J. A., Sweder K. S. & Hanawalt P. C. (1995) Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Res* 23: 2715-2723.

Epshtein V., Cardinale C. J., Ruckenstein A. E., Borukhov S. & Nudler E. (2007) An allosteric path to transcription termination. *Mol Cell* 28: 991-1001.

Epshtein V., Dutta D., Wade J. & Nudler E. (2010) An allosteric mechanism of Rho-dependent transcription termination. *Nature* 463: 245-249.

Epshtein V. & Nudler E. (2003) Cooperation between RNA polymerase molecules in transcription elongation. *Science* 300: 801-805.

Epshtein V., Toulme F., Rahmouni A. R., Borukhov S. & Nudler E. (2003) Transcription through the roadblocks: the role of RNA polymerase cooperation. *EMBO J* 22: 4719-4727.

Estrem S. T., Ross W., Gaal T., Chen Z. W., Niu W., Ebright R. H. & Gourse R. L. (1999) Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes Dev* 13: 2134-2147.

Evans E., Moggs J. G., Hwang J. R., Egly J. M. & Wood R. D. (1997) Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *EMBO J* 16: 6559-6573.

Fernandez S., Ayora S. & Alonso J. C. (2000) *Bacillus subtilis* homologous recombination: genes and products. *Res Microbiol* 151: 481-486.

Fish R. N. & Kane C. M. (2002) Promoting elongation with transcript cleavage stimulatory factors. *Biochim Biophys Acta* 1577: 287-307.

Flores M. J., Bidnenko V. & Michel B. (2004) The DNA repair helicase UvrD is essential for replication fork reversal in replication mutants. *EMBO Rep* 5: 983-988.

Flores M. J., Sanchez N. & Michel B. (2005) A fork-clearing role for UvrD. *Mol Microbiol* 57: 1664-1675.

Fulop V. & Jones D. T. (1999) Beta propellers: structural rigidity and functional diversity. *Curr Opin Struct Biol* 9: 715-721.

Goosen N. & Moolenaar G. F. (2001) Role of ATP hydrolysis by UvrA and UvrB during nucleotide excision repair. *Res Microbiol* 152: 401-409.

Gross C. A., Chan C., Dombroski A., Gruber T., Sharp M., Tupy J. & Young B. (1998) The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harb Symp Quant Biol* 63: 141-155.

Gusarov I. & Nudler E. (1999) The mechanism of intrinsic transcription termination. *Mol Cell* 3: 495-504.

Gusarov I. & Nudler E. (2001) Control of intrinsic transcription termination by N and NusA: the basic mechanisms. *Cell* 107: 437-449.

Guy L. & Roten C. A. (2004) Genometric analyses of the organization of circular chromosomes: a universal pressure determines the direction of ribosomal RNA genes transcription relative to chromosome replication. *Gene* 340: 45-52.

Ha K. S., Toulokhonov I., Vassylyev D. G. & Landick R. (2010) The NusA N-terminal domain is necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase. *J Mol Biol* 401: 708-725.

Hall M. C. & Matson S. W. (1997) Mutation of a highly conserved arginine in motif IV of *Escherichia coli* DNA helicase II results in an ATP-binding defect. *J Biol Chem* 272: 18614-18620.

Hall M. C. & Matson S. W. (1999) Helicase motifs: the engine that powers DNA unwinding. *Mol Microbiol* 34: 867-877.

Harper S. & Speicher D. W. (2011) Purification of proteins fused to glutathione S-transferase. *Methods Mol Biol* 681: 259-280.

Harriott K. (2006) The characterisation of the interaction between PcrA and the β subunit of RNA polymerase. University of Newcastle, Australia.

Hawley D. K. & McClure W. R. (1983) Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucleic Acids Res* 11: 2237-2255.

Heller R. C. & Marians K. J. (2005) Unwinding of the nascent lagging strand by Rep and PriA enables the direct restart of stalled replication forks. *J Biol Chem* 2880: 24143-34151.

Heller R. C. & Marians K. J. (2007) Non-replicative helicases at the replication fork. *DNA repair (Amst)* 6: 945-952.

Helmann J. D. (2003) Purification of *Bacillus subtilis* RNA polymerase and associated factors. *Methods Enzymol* 370: 10-24.

Hickman A. B. & Dyda F. (2005) Binding and unwinding: SF3 viral helicases. *Curr Opin Struct Biol* 15: 77-85.

Hiller-Sturmehofel S., Sobin J. & Mayfield R. D. (2008) Proteomic approaches for studying alcoholism and alcohol-induced organ damage. *Alcohol Research and Health* 31: 36-48.

Ho S. N., Hunt H. D., Horton R. M., Pullen J. K. & Pease L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77: 51-59.

Holm L. & Rosenstrom P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38: W545-549.

Horwitz R. J., Li J. & Greenblatt J. (1987) An elongation control particle containing the N gene transcriptional antitermination protein of bacteriophage lambda. *Cell* 51: 631-641.

Iben S., Tschochner H., Bier M., Hoogstraten D., Hozak P., Egly J. M. & Grummt I. (2002) TFIIH plays an essential role in RNA polymerase I transcription. *Cell* 109: 297-306.

Iordanescu S. (1993) Characterization of the *Staphylococcus aureus* chromosomal gene pcrA, identified by mutations affecting plasmid pT181 replication. *Mole Gen Genet* 241: 185-192.

Ish-Horowicz D. & Burke J. F. (1981) Rapid and efficient cosmid cloning. *Nucleic Acids Res* 9: 2989-2998.

Jankowsky E. & Fairman-Williams M. E. (2010) An introduction to RNA helicases: Superfamilies, families and major themes. In: *RNA helicases* (ed. E. Jankowsky) pp. 1-31. RSC Biomolecular Sciences 19.

Johnston E. B. (2010) The RNA polymerase-σ interaction as a target for novel antimicrobials, University of Newcastle, Australia.

Johnston E. B., Lewis P. J. & Griffith R. (2009) The interaction of *Bacillus subtilis* sigmaA with RNA polymerase. *Protein Sci* 18: 2287-2297.

Juang Y. L. & Helmann J. D. (1994) The delta subunit of *Bacillus subtilis* RNA polymerase. An allosteric effector of the initiation and core-recycling phases of transcription. *J Mol Biol* 239: 1-14.

Kapanidis A. N., Margeat E., Ho S. O., Kortkhonjia E., Weiss S. & Ebright R. H. (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* 314: 1144-1147.

Knight D. P. (1975) Negative staining of rat tail tendon collagen fibrils with uranyl formate. *Tissue Cell* 7: 651-654.

Komissarova N. & Kashlev M. (1997) Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc Natl Acad Sci U S A* 94: 1755-1760.

Kornberg A., Scott J. F. & Bertsch L. L. (1978) ATP utilization by rep protein in the catalytic separation of DNA strands at a replicating fork. *J Biol Chem* 253: 3298-3304.

Korolev S., Hsieh J., Gauss G. H., Lohman T. M. & Waksman G. (1997) Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. *Cell* 90: 635-647.

Krasny L. & Gourse R. L. (2004) An alternative strategy for bacterial ribosome synthesis: Bacillus subtilis rRNA transcription regulation. *EMBO J* 23: 4473-4483.

Kunalan S. (2009) Mapping the binding sites for transcription factors NusG and Rho on RNA polymerase. University of Newcastle, Australia.

Kunst F. & Rapoport G. (1995) Salt stress is an environmental signal affecting degradative enzyme synthesis in Bacillus subtilis. *J Bacteriol* 177: 2403-2407.

Kuo H. K., Krasich R., Bhagwat A. S. & Kreuzer K. N. (2010) Importance of the tmRNA system for cell survival when transcription is blocked by DNA-protein cross-links. *Mol Microbiol* 78: 686-700.

Kushner S. R. (1978) *Genetic Engineering*. North Holland Publishers, Amsterdam.

Kusuya Y., Kurokawa K., Ishikawa S., Ogasawara N. & Oshima T. (2011) Transcription factor GreA contributes to resolving promoter-proximal pausing of RNA polymerase in *Bacillus subtilis* cells. *J Bacteriol* 193: 3090-3099.

Kuzminov A. & Stahl F. W. (1999) Double-strand end repair via the RecBC pathway in *Escherichia coli* primes DNA replication. *Genes Dev* 13: 345-356.

Landick R. (1999) Shifting RNA polymerase into overdrive. *Science* 284: 598-599.

Laptenko O., Lee J., Lomakin I. & Borukhov S. (2003) Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J* 22: 6322-6334.

Lederberg E. M. & Cohen S. N. (1974) Transformation of *Salmonella typhimurium* by plasmid deoxyribonucleic acid. *J Bacteriol* 119: 1072-1074.

Lee D. J., Busby S. J., Westblade L. F. & Chait B. T. (2008) Affinity isolation and I-DIRT mass spectrometric analysis of the *Escherichia coli* O157:H7 Sakai RNA polymerase complex. *J Bacteriol* 190: 1284-1289.

Lestini R. & Michel B. (2007) UvrD controls the access of recombination proteins to blocked replication forks. *EMBO J* 26: 3804-3814.

Lewis P. J. & Marston A. L. (1999) GFP vectors for controlled expression and dual labelling of protein fusions in *Bacillus subtilis*. *Gene* 227: 101-110.

Lichty J. J., Malecki J. L., Agnew H. D., Michelson-Horowitz D. J. & Tan S. (2005) Comparison of affinity tags for protein purification. *Protein Expr Purif* 41: 98-105.

Lin Y. C., Choi W. S. & Gralla J. D. (2005) TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. *Nat Struct Mol Biol* 12: 603-607.

Lohman T. M. & Bjornson K. P. (1996) Mechanisms of helicase-catalyzed DNA unwinding. *Annu Rev Biochem* 65: 169-214.

Lopez de Saro F. J., Woody A. Y. & Helmann J. D. (1995) Structural analysis of the *Bacillus subtilis* delta factor: a protein polyanion which displaces RNA from RNA polymerase. *J Mol Biol* 252: 189-202.

Lopez de Saro F. J., Yoshikawa N. & Helmann J. D. (1999) Expression, abundance, and RNA polymerase binding properties of the delta factor of *Bacillus subtilis*. *J Biol Chem* 274: 15953-15958.

Ludtke S. J. (2010) 3-D structures of macromolecules using single-particle analysis in EMAN. *Methods Mol Biol* 673: 157-173.

Ludtke S. J., Baldwin P. R. & Chiu W. (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128: 82-97.

Ludtke S. J., Jakana J., Song J. L., Chuang D. T. & Chiu W. (2001) A 11.5 A single particle reconstruction of GroEL using EMAN. *J Mol Biol* 314: 253-262.

Mahdi A. A., Briggs G. S., Sharples G. J., Wen Q. & Lloyd R. G. (2003) A model for dsDNA translocation revealed by a structural motif common to RecG and Mfd proteins. *EMBO J* 22: 724-734.

Malta E., Moolenaar G. F. & Goosen N. (2006) Base flipping in nucleotide excision repair. *J Biol Chem* 281: 2184-2194.

Maluf N. K., Fischer C. J. & Lohman T. M. (2003) A Dimer of Escherichia coli UvrD is the active form of the helicase in vitro. *J Mol Biol* 325: 913-935.

Manelyte L., Guy C. P., Smith R. M., Dillingham M. S., McGlynn P. & Savery N. J. (2009) The unstructured C-terminal extension of UvrD interacts with UvrB, but is dispensable for nucleotide excision repair. *DNA repair (Amst)* 8: 1300-1310.

Marians K. J. (1997) Helicase structures: a new twist on DNA unwinding. *Structure* 5: 1129-1134.

Marintcheva B. & Weller S. K. (2003) Helicase motif Ia is involved in single-strand DNA-binding and helicase activities of the herpes simplex virus type 1 origin-binding protein, UL9. *J Virol* 77: 2477-2488.

Matson S. W. & George J. W. (1987) DNA helicase II of *Escherichia coli*. Characterization of the single-stranded DNA-dependent NTPase and helicase activities. *J Biol Chem* 262: 2066-2076.

Matson S. W. & Kaiser-Rogers K. A. (1990) DNA helicases. *Annu Rev Biochem* 59: 289-329.

Matsushima Y., Farr C. L., Fan L. & Kaguni L. S. (2008) Physiological and biochemical defects in carboxyl-terminal mutants of mitochondrial DNA helicase. *J Biol Chem* 283: 23964-23971.

McGlynn P. (2011) Helicases that underpin replication of protein-bound DNA in *Escherichia coli. Biochemical Society transaction* 39: 606-610.

McKinley B. A. & Sukhodolets M. V. (2007) Escherichia coli RNA polymerase-associated SWI/SNF protein RapA: evidence for RNA-directed binding and remodeling activity. *Nucleic Acids Res* 35: 7044-7060.

Mehta P., Richards J. & Karzai A. W. (2006) tmRNA determinants required for facilitating nonstop mRNA decay. *RNA* 12: 2187-2198.

Mendonca V. M., Kaiser-Rogers K. & Matson S. W. (1993) Double helicase II (uvrD)-helicase IV (helD) deletion mutants are defective in the recombination pathways of Escherichia coli. *J Bacteriol* 175: 4641-4651.

Mendonca V. M., Klepin H. D. & Matson S. W. (1995) DNA helicases in recombination and repair: construction of a delta uvrD delta helD delta recQ mutant deficient in recombination and repair. *J Bacteriol* 177: 1326-1335.

Michel B. (2000) Replication fork arrest and DNA recombination. *Trends Biochem Sci* 25: 173-178.

Minko I. G., Zou Y. & Lloyd R. S. (2002) Incision of DNA-protein crosslinks by UvrABC nuclease suggests a potential repair pathway involving nucleotide excision repair. *Proc Natl Acad Sci U S A* 99: 1905-1909.

Mirkin E. V. & Mirkin S. M. (2007) Replication fork stalling at natural impediments. *Microbiol Mol Biol Rev* 71: 13-35.

Miroux B. & 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. *J Mol Biol* 260: 289-298.

Mooney R. A., Davis S. E., Peters J. M., Rowland J. L., Ansari A. Z. & Landick R. (2009) Regulator trafficking on bacterial transcription units *in vivo*. *Mol Cell* 33: 97-108.

Moreland R. J., Tirode F., Yan Q., Conaway J. W., Egly J. M. & Conaway R. C. (1999) A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II. *J Biol Chem* 274: 22127-22130.

Morgan E. A. (1986) Antitermination mechanisms in rRNA operons of *Escherichia coli*. *J Bacteriol* 168: 1-5.

Murakami K. S. & Darst S. A. (2003) Bacterial RNA polymerases: the wholo story. *Curr Opin Struct Biol* 13: 31-39.

Myong S., Rasnik I., Joo C., Lohman T. M. & Ha T. (2005) Repetitive shuttling of a motor protein on DNA. *Nature* 437: 1321-1325.

Naqvi A., Tinsley E. & Khan S. A. (2003) Purification and characterization of the PcrA helicase of *Bacillus anthracis*. *J Bacteriol* 185: 6633-6639.

Neylon C., Brown S. E., Kralicek A. V., Miles C. S., Love C. A. & Dixon N. E. (2000) Interaction of the *Escherichia coli* replication terminator protein (Tus) with DNA: a model derived from DNA-binding studies of mutant proteins by surface plasmon resonance. *Biochemistry* 39: 11989-11999.

Niedziela-Majka A., Chesnik M. A., Tomko E. J. & Lohman T. M. (2007) *Bacillus stearothermophilus* PcrA monomer is a single-stranded DNA translocase but not a processive helicase in vitro. *J Biol Chem* 282: 27076-27085.

Noirot-Gros M. F., Dervyn E., Wu L. J., Mervelet P., Errington J., Ehrlich S. D. & Noirot P. (2002a) An expanded view of bacterial DNA replication. *Proc Natl Acad Sci U S A* 99: 8342-8347.

Noirot-Gros M. F., Soultanas P., Wigley D. B., Ehrlich S. D., Noirot P. & Petit M. A. (2002b) The beta-propeller protein YxaL increases the processivity of the PcrA helicase. *Mol Genet Genomics* 267: 391-400.

Nudler E. (1999) Transcription elongation: structural basis and mechanisms. *J Mol Biol* 288: 1-12.

Ohi M., Li Y., Cheng Y. & Walz T. (2004) Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biol Proced Online* 6: 23-34.

Okubo S. & Romig W. R. (1966) Impaired transformability of Bacillus subtilis mutant sensitive to mitomycin C and ultraviolet radiation. *J Mol Biol* 15: 440-454.

Park J., Myong S., Niedziela-Majka A., Lee K. S., Yu J., Lohman T. M. & Ha T. (2010) PcrA helicase dismantles RecA filaments by reeling in DNA in uniform steps. *Cell* 142: 544-555.

Park J. S., Marr M. T. & Roberts J. W. (2002) *E. coli* Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell* 109: 757-767.

Paul B. J., Barker M. M., Ross W., Schneider D. A., Webb C., Foster J. W. & Gourse R. L. (2004) DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell* 118: 311-322.

Peters J. M., Mooney R. A., Kuan P. F., Rowland J. L., Keles S. & Landick R. (2009) Rho directs widespread termination of intragenic and stable RNA transcription. *Proc Natl Acad Sci U S A* 106: 15406-15411.

Petit M. A., Dervyn E., Rose M., Entian K. D., McGovern S., Ehrlich S. D. & Bruand C. (1998) PcrA is an essential DNA helicase of *Bacillus subtilis* fulfilling functions both in repair and rolling-circle replication. *Mol Microbiol* 29: 261-273.

Petit M. A. & Ehrlich D. (2002) Essential bacterial helicases that counteract the toxicity of recombination proteins. *EMBO J* 21: 3137-3147.

Pettersen E. F., Goddard T. D., Huang C. C., Couch G. S., Greenblatt D. M., Meng E. C. & Ferrin T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605-1612.

Pomerantz R. T. & O'Donnell M. (2007) Replisome mechanics: insights into a twin DNA polymerase machine. *Trends Microbiol* 15: 156-164.

Pomerantz R. T. & O'Donnell M. (2010a) Direct restart of a replication fork stalled by a head-on RNA polymerase. *Science* 327: 590-592.

Pomerantz R. T. & O'Donnell M. (2010b) What happens when replication and transcription complexes collide? *Cell Cycle* 9: 2537-2543.

Prasch S., Jurk M., Washburn R. S., Gottesman M. E., Wohrl B. M. & Rosch P. (2009) RNA-binding specificity of *E. coli* NusA. *Nucleic Acids Res* 37: 4736-4742.

Rasnik I., Myong S., Cheng W., Lohman T. M. & Ha T. (2004) DNA-binding orientation and domain conformation of the E. coli rep helicase monomer bound to a partial duplex junction: single-molecule studies of fluorescently labeled enzymes. *J Mol Biol* 336: 395-408.

Richardson J. P. (2002) Rho-dependent termination and ATPases in transcript termination. *Biochim Biophys Acta* 1577: 251-260.

Richardson J. P., Greenblatt, J (1996) Control of RNA chain elongation and termination in *Escherichia coli* and *Samonella typhimurium* In: *Cellular and Molecular Biology* (ed. F. Neidhardt, Ingraham, JL, Low, KB, Magasanik, B, Schaecher, M and Umbarger HE) pp. 822-848. American Society for Microbiology, Washington DC.

Roberts J. & Park J. S. (2004) Mfd, the bacterial transcription repair coupling factor: translocation, repair and termination. *Curr Opin Microbiol* 7: 120-125.

Rocha E. P., Cornet E. & Michel B. (2005) Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLoS Genet* 1: e15.

Rudolph C. J., Dhillon P., Moore T. & Lloyd R. G. (2007) Avoiding and resolving conflicts between DNA replication and transcription. *DNA repair (Amst)* 6: 981-993.

Saha A., Wittmeyer J. & Cairns B. R. (2002) Chromatin remodeling by RSC involves ATP-dependent DNA translocation. *Genes Dev* 16: 2120-2134.

Scott J. F. & Kornberg A. (1978) Purification of the rep protein of Escherichia coli. An ATPase which separates duplex DNA strands in advance of replication. *J Biol Chem* 253: 3292-3297.

Selby C. P. & Sancar A. (1993a) Molecular mechanism of transcription-repair coupling. *Science* 260: 53-58.

Selby C. P. & Sancar A. (1993b) Transcription-repair coupling and mutation frequency decline. *J Bacteriol* 175: 7509-7514.

Selby C. P. & Sancar A. (1995) Structure and function of transcription-repair coupling factor. II. Catalytic properties. *J Biol Chem* 270: 4890-4895.

Serebriiskii I., Khazak V. & Golemis E. A. (1999) A two-hybrid dual bait system to discriminate specificity of protein interactions. *J Biol Chem* 274: 17080-17087.

Sevostyanova A., Feklistov A., Barinova N., Heyduk E., Bass I., Klimasauskas S., Heyduk T. & Kulbachinskiy A. (2007) Specific recognition of the -10 promoter element by the free RNA polymerase sigma subunit. *J Biol Chem* 282: 22033-22039.

Sevostyanova A., Svetlov V., Vassylyev D. G. & Artsimovitch I. (2008) The elongation factor RfaH and the initiation factor sigma bind to the same site on the transcription elongation complex. *Proc Natl Acad Sci U S A* 105: 865-870.

Shereda R. D., Kozlov A. G., Lohman T. M., Cox M. M. & Keck J. L. (2008) SSB as an organizer/mobilizer of genome maintenance complexes. *Crit Rev Biochem Mol Biol* 43: 289-318.

Singleton M. R., Dillingham M. S. & Wigley D. B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* 76: 23-50.

Smith A. J. & Savery N. J. (2005) RNA polymerase mutants defective in the initiation of transcription-coupled DNA repair. *Nucleic Acids Res* 33: 755-764.

Som S. & Friedman S. (1994) Inhibition of transcription in vitro by binding of DNA (cytosine-5)-methylases to DNA templates containing cytosine analogs. *J Biol Chem* 269: 25986-25991.

Soultanas P., Dillingham M. S., Papadopoulos F., Phillips S. E., Thomas C. D. & Wigley D. B. (1999) Plasmid replication initiator protein RepD increases the processivity of PcrA DNA helicase. *Nucleic Acids Res* 27: 1421-1428.

Soultanas P., Dillingham M. S., Wiley P., Webb M. R. & Wigley D. B. (2000) Uncoupling DNA translocation and helicase activity in PcrA: direct evidence for an active mechanism. *EMBO J* 19: 3799-3810.

Soultanas P. & Wigley D. B. (2000) DNA helicases: 'inching forward'. *Curr Opin Struct Biol* 10: 124-128.

Soultanas P. & Wigley D. B. (2001) Unwinding the 'Gordian knot' of helicase action. *Trends Biochem Sci* 26: 47-54.

Stanley L. K., Seidel R., van der Scheer C., Dekker N. H., Szczelkun M. D. & Dekker C. (2006) When a helicase is not a helicase: dsDNA tracking by the motor protein EcoR124I. *EMBO J* 25: 2230-2239.

Studier F. W. & Moffatt B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 189: 113-130.

Subramanya H. S., Bird L. E., Brannigan J. A. & Wigley D. B. (1996) Crystal structure of a DExx box DNA helicase. *Nature* 384: 379-383.

Sukhodolets M. V., Cabrera J. E., Zhi H. & Jin D. J. (2001) RapA, a bacterial homolog of SWI2/SNF2, stimulates RNA polymerase recycling in transcription. *Genes Dev* 15: 3330-3341.

Sukhodolets M. V. & Jin D. J. (1998) RapA, a novel RNA polymerase-associated protein, is a bacterial homolog of SWI2/SNF2. *J Biol Chem* 273: 7018-7023.

Sukhodolets M. V. & Jin D. J. (2000) Interaction between RNA polymerase and RapA, a bacterial homolog of the SWI/SNF protein family. *J Biol Chem* 275: 22090-22097.

Svejstrup J. Q., Vichi P. & Egly J. M. (1996) The multiple roles of transcription/repair factor TFIIH. *Trends Biochem Sci* 21: 346-350.

Tachino A., Igarashi S. & Sode K. (2007) W-motif exchange between beta-propeller proteins. *Protein J* 26: 153-158.

Tehranchi A. K., Blankschien M. D., Zhang Y., Halliday J. A., Srivatsan A., Peng J., Herman C. & Wang J. D. (2010) The transcription factor DksA prevents conflicts between DNA replication and transcription machinery. *Cell* 141: 595-605.

Toseland C. P., Martinez-Senac M. M., Slatter A. F. & Webb M. R. (2009) The ATPase cycle of PcrA helicase and its coupling to translocation on DNA. *J Mol Biol* 392: 1020-1032.

Truglio J. J., Croteau D. L., Van Houten B. & Kisker C. (2006) Prokaryotic nucleotide excision repair: the UvrABC system. *Chem Rev* 106: 233-252.

van Brabant A. J., Stan R. & Ellis N. A. (2000) DNA helicases, genomic instability, and human genetic disease. *Annu Rev Genomics Hum Genet* 1: 409-459.

van Heel M. & Schatz M. (2005) Fourier shell correlation threshold criteria. *J Struct Biol* 151: 250-262.

Van Houten B., Croteau D. L., DellaVecchia M. J., Wang H. & Kisker C. (2005) 'Close-fitting sleeves': DNA damage recognition by the UvrABC nuclease system. *Mutat Res* 577: 92-117.

Vassylyev D. G., Sekine S., Laptenko O., Lee J., Vassylyeva M. N., Borukhov S. & Yokoyama S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature* 417: 712-719.

Vassylyev D. G., Vassylyeva M. N., Perederina A., Tahirov T. H. & Artsimovitch I. (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* 448: 157-162.

Veaute X., Delmas S., Selva M., Jeusset J., Le Cam E., Matic I., Fabre F. & Petit M. A. (2005) UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in *Escherichia coli*. *EMBO J* 24: 180-189.

Velankar S. S., Soultanas P., Dillingham M. S., Subramanya H. S. & Wigley D. B. (1999) Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* 97: 75-84.

Verma S., Xiong Y., Mayer M. U. & Squier T. C. (2007) Remodeling of the bacterial RNA polymerase supramolecular complex in response to environmental conditions. *Biochemistry* 46: 3023-3035.

von Hippel P. H. (1998) An integrated model of the transcription complex in elongation, termination, and editing. *Science* 281: 660-665.

Weeda G., van Ham R. C., Vermeulen W., Bootsma D., van der Eb A. J. & Hoeijmakers J. H. (1990) A presumed DNA helicase encoded by ERCC-3 is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome. *Cell* 62: 777-791.

Westblade L. F., Campbell E. A., Pukhrambam C., Padovan J. C., Nickels B. E., Lamour V. & Darst S. A. (2010) Structural basis for the bacterial transcription-repair coupling factor/RNA polymerase interaction. *Nucleic Acids Res* 38: 8357-8369.

Wigneshweraraj S. R., Burrows P. C., Severinov K. & Buck M. (2005) Stable DNA opening within open promoter complexes is mediated by the RNA polymerase beta'-jaw domain. *J Biol Chem* 280: 36176-36184.

Withers R. (2007) The investigation of the interaction between the C-terminus of PcrA and the N-terminal region of the β subunit of RNAP. University of Newcastle, Australia.

Xu J. & Zhang Y. (2010) How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* 26: 889-895.

Yamamoto Y., Sunohara T., Jojima K., Inada T. & Aiba H. (2003) SsrA-mediated trans-translation plays a role in mRNA quality control by facilitating degradation of truncated mRNAs. *RNA* 9: 408-418.

Yancey J. E. & Matson S. W. (1991) The DNA unwinding reaction catalyzed by Rep protein is facilitated by an RHSP-DNA interaction. *Nucleic Acids Res* 19: 3943-3951.

Yang X., Doherty G. P. & Lewis P. J. (2008a) Tandem affinity purification vectors for use in gram positive bacteria. *Plasmid* 59: 54-62.

Yang X. & Lewis P. J. (2008) Overproduction and purification of recombinant Bacillus subtilis RNA polymerase. *Protein Expr Purif* 59: 86-93.

Yang X., Molimau S., Doherty G. P., Johnston E. B., Marles-Wright J., Rothnagel R., Hankamer B., Lewis R. J. & Lewis P. J. (2009) The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA. *EMBO Rep* 10: 997-1002.

Yang Y., Dou S. X., Ren H., Wang P. Y., Zhang X. D., Qian M., Pan B. Y. & Xi X. G. (2008b) Evidence for a functional dimeric form of the PcrA helicase in DNA unwinding. *Nucleic Acids Res* 36: 1976-1989.

Yarranton G. T. & Gefter M. L. (1979) Enzyme-catalyzed DNA unwinding: studies on *Escherichia coli* rep protein. *Proc Natl Acad Sci U S A* 76: 1658-1662.

Yeeles J. T. & Dillingham M. S. (2010) The processing of double-stranded DNA breaks for recombinational repair by helicase-nuclease complexes. *DNA Repair (Amst)* 9: 276-285.

Young B. A., Gruber T. M. & Gross C. A. (2002) Views of transcription initiation. *Cell* 109: 417-420.

Yu J., Ha T. & Schulten K. (2006) Structure-based model of the stepping motor of PcrA helicase. *Biophys J* 91: 2097-2114.

Yu J., Ha T. & Schulten K. (2007) How directional translocation is regulated in a DNA helicase motor. *Biophys J* 93: 3783-3797.

Yuan J., Ghosal G. & Chen J. (2009) The annealing helicase HARP protects stalled replication forks. *Genes Dev* 23: 2394-2399.

Yusufzai T. & Kadonaga J. T. (2008) HARP is an ATP-driven annealing helicase. *Science* 322: 748-750.

Zhang G., Campbell E. A., Minakhin L., Richter C., Severinov K. & Darst S. A. (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 A resolution. *Cell* 98: 811-824.

Zhang W., Dillingham M. S., Thomas C. D., Allen S., Roberts C. J. & Soultanas P. (2007) Directional loading and stimulation of PcrA helicase by the replication initiator protein RepD. *J Mol Biol* 371: 336-348.

Zurita M. & Merino C. (2003) The transcriptional complexity of the TFIIH complex. *Trends Genet* 19: 578-584.

Appendices

Appendix I: PcrA protein sequence alignment

SUBT – *Bacillus subtilis*

GEOB – *Geobacillus stearothermophilus*

LIST – *Listeria innocus*

STAPS – *Staphylcoccus saprophyticus*

STAPA – *Staphylcoccus aureus*

LACT – *Lactobacillus sakei*

STREP – *Streptococcus pneumoniae*

CLOST – *Clostridium perfringes*

CORYN – *Corynebacterium diphtheriae*

SERR – *Serratia marcescens*

```
SUBT     1 -------------------------MNYISNQLLSGLNPVQQEAVKTTDGPLLLMAGAGSGKTRVLTHRIAYLMAEKHVAPWNILAITFTNKAAREMKERVESILGPGADDI
GEOB     1 -------------------------MNFLSEKLLAHLNKEQQYEAATTTEGPLLIMAGAGSGKTRVLTHRIAYLMAEKQVAPWNILAITFTNKAAREMKERVQALLGGAAEDV
LIST     1 ------------------------MNAKELVDGLNPEQRKAVESTEGPLLIMAGAGSGKTRVLTHRIAYLVRERGVNPYNILAITFTNKAAREMKSRIGNLMGGEAESI
STAPS    1 -----------------------MNALVNKMNDEQSQAVRTTEGPLLIMAGAGSGKTRVLTHRIAYLVRERGVNPYNILAITFTNKAAREMKSRIGNLMGGEAESI
STAPA    1 ------------------------MNTLLNHMNTEQSEAVKTTEGPLLIMAGAGSGKTRVLTHRIAYLLDEKDVSPYNVLAITFTNKAAREMKERVQKLVGDQAEVI
LACT     1 ------------------------MSEQELLKGMNPQAEAVALTTEGPLIMAGAGSGKTRVLTHRVAYLHVLPWRILAITFTNKAAREMRERVGNLLGEGAQDV
STREP    1 ------------------------MNALLNGMNDRQAEAVQTTEGPLLIMAGAGSGKTRVLTHRIAYLIDEKLVNPWNILAITFTNKAAREMKERAYSLN-PATQDC
CLOST    1 ----------------------MDLKSLLNKEQYEAATTIDGQVLILAGAGSGKTRVLTHRIAYMIEN-DIKPYNILAITFTNKAAGEMRERVKSLVGEVANNM
CORYN    1 MASQANPFLRQAQP----------VSQYTGADLTEGLNPQQKAAVEHYGSPLLIVAGAGSGKTSVLTRRIAYLMRVRGVQPQNILAITFTNKAAAEMRERVSSLVGPVAQRM
SERR     1 ------------------------MDVSDLLDSLNEKQREAVAAPRSNLLVLAGAGSGKTRVLVHRIAWLLSVENCSPYSIMAVTFTNKAAAEMRHRIEHLIGTSQGGM

SUBT    88 WISTFHSMCVRILRRDIDRIG-INRNFSILDTADQLSVIKGILKERNLDPKKFDPRSILGTIS-AKNELTEPEEFSKVAG---GYYDQVVSDVYADYQKKLLKNQSLDFDDL
GEOB    88 WISTFHSMCVRMLRRDIDRIG-INRNFSILDPTDQLSVIKAILKDKNIDPKKFEPRTILATISGAKNDLLTPEQEAKRAS---TYYEKIVGDVYQEYQQRLLRNHSLDFDDL
LIST    86 WISTFHSMCVRILRRDIDRIG-YERNFTILDGSDQLSVIKGILKEKNVDPKKFEPRGILASISNAKNELITASEYIKEAS---GFYDKMVGEVYEKYEKKLKKNQALDFDDL
STAPS   84 WMSTFHSMCVRILRRDADRIG-IERNFTIIDPTDQKSVIKDVLKNENIDSKKFEPRMFIGAISNLKNELKTPEDAQKEAN---DYHEQMVATVYKGYQRQLSRNEALDFDDL
STAPA   84 WMSTFHSMCVRILRRDADRIG-IERNFTIIDPTDQKSVIKDVLKNENIDSKKFEPRMFIGAISNLKNELKTPADAQKEAT---DYHSQMVATVYSGYQRQLSRNEALDFDDL
LACT    86 WVSTFHALCVRILRREIEQIG-YNRAFTIADPSEQLTMRHVCRDLNIDTKKYEPKAILGTISKAKNALQTPADYQGLAN---GPFEKMAAKCYTEYQKRLHQNQALDFDDL
STREP   83 LIATFHSMCVRILRRDADHIG-YNRNFTIVDPGEQRTLMKRILKQLNLDPKKWNERTILGTISNAKNDLIDDVAYAAQAG---DMYTQIVAQCYTAYQKELRQSESVDFDDL
CLOST   82 WISTFHSSCVRILRREIDKLG-YSKDFTIYDSDQKTLIKLVMKELNINEKEITDLEILGTIGKAKDNIQSAQSEKKENED-NFRKNKIADAYLLYQKKLKENNALDFDDL
CORYN  103 WVSTFHSTCVRILREQAQLVPGLNTNFTIYDSDDSKRLLSMISKDFSLDIKKFTPRLLATGISNLKNELIDADQACADAESTKNPYETTIASVFAEYQKRLRASNAVDFDDL
SERR    86 WIGTFHGLAHRLLRAHHLEAN-LPQDFQILDSDDLRLLKRIIKALNVDEKQWPPRQAMWYINGKKDEGLRPQHVETYNN----PVEATWLRIYQAYQEACDRAGLVDFAEL

SUBT   195 IMTTIKLFDRVPEVLEFYQRKFQYIHVDEYQDTNRAQYMLVKQLAER--------------FQNLCVVGDSDQSIYRWRGADITNILSFEKDYPNASVILLEQNYRSTKRILR
GEOB   196 IMTTIHLFDRVPDVLQYYQKFQYIHIDEYQDTNRAQYTLVKKLADR--------------FQNICAVGDADQSIYRWRGADIRNILSFERDYPNAKVILLEQNYRSTKRILQ
LIST   194 IMVTIQLFERVPDVLEYYQRKFQYIHVDEYQDTNHAQYLVKLLASR--------------FKNLCVVGDSDQSIYGWRGADISNIMSFEKDYPSAKTILLEENYRSTKRILE
STAPS  192 IMVTIRLFERVPDVLDYYQNKFQYIHVDEYQDTNKAQYTLVKLLAAK--------------FKNLCVVGDSDQSIYGWRGADIQNILSFEEDYPEAKTIFLEQNYRSTKNILN
STAPA  192 IMTTINLFERVPEVLEYYQNKFQYIHVDEYQDTNKAQYTLVKLLASK--------------FKNLCVVGDSDQSIYGWRGADIQNILSFEKDYPEANTIFLEQNYRSTKTIILN
LACT   194 IMQTIRLFRENKETLAHYQEKFQYIHVDEYQDTNEAQYTLVNMLAEK--------------YQNLCVVGDADQSIYGWRGADISNIMNFEHDYPDAKSVMLEQNYRSTKTILE
STREP  191 IMLTLRLFDQNPDVLTYYQQKFQYIHVDEYQDTNHAQYQLVKLLASR--------------FKNICVVGDADQSIYGWRGADMQNILDFEKDYPKAKVVLLEENYRSTKTILQ
CLOST  191 IVKTVELFRKHNDVLEFYQNKFKYIMIDEYQDTNKAQYEFAKLLAAK--------------HKNICVVGDDDQCIYAWRGADIRNILDFEKDYPEAKVVKLEQNYRSKGNILE
CORYN  215 IGETVRIFREHPPVTEYYRRFRHVLIDEYQDTNHAQYALIHELVGSG----------PDASELCVVGDSDQSIYAFRGATIRNIEEFERDYPNAETILLEQNYRSTQTIILN
SERR   193 LLRAHELWLNKPHILNHYRERFTNVLVDEFQDTNSIQYAWIRLLAGG------------NSNVMIVGDDDQSIYGWRGAQVENIQRFLKDFPGAETIRLEQNYRSTSNILK

SUBT   294 AANEVILWTEN-----------DEGIKISYYRGDNEFGEGQFVAGKIHQLHS-GKRKLSDIAILYRTNAQSRVIEETLLKAGLNYNIVGGTKFYDRKEIKDILAYLRLVSNPD
GEOB   295 AANEVIAHVNRKPKRLWTENPEGKPIVYYEAMNEVDEAQFVARRIQEAVESGERRYRDFAVLYRTNAQSRVIEEMFLKANIPYQIVGGLKFYDRKEIKDILAYLRVIANPD
LIST   293 AANRVIEN----------------------------------------------------QSRVMEEYFMKSNMAYTMVGGTKFYDRKEIKDILAYLRLISNNE
STAPS  291 AANEVIKNNSERKPKGLWTGNTSGDKIHYYEATTERDEAEYVVREIMKHQR-NGKKYQDMAILYRTNAQSRVLEETFMKSNLPYTMVGGQKFYDRKEIKDLLSYLRIIANSN
STAPA  291 AANEVIKNNSERKPKGLWTANTNGEKIHYYEAMTERDEAEFVIREIMKHQR-NGKKYQDMAILYRTNAQSRVLEETFMKSNMPYTMVGGQKFYDRKEIKDLLSYLRIIANSN
LACT   293 AANGVIKNNKNRNTFRKPELWTQNEEGQPITYYRGQSERDESLFVITKIQEEMQKNHRKYGDFAVLYRTNAQSRTIEEAFVKANMPYKMVGGHKFYDRKEIKDILAYLRLVVNAA
STREP  290 AANEVIKNNKNRRPKNLWTQNADGEQIVYYRADDELEAFVARTIDELSRSQNFLHKDFAVLYRTNAQSRTIEEALLKSNIPYTMVGGTKFYSRKEIRDIIAYLNLIANLS
CLOST  290 AANKVIKNNAERKFKILRTEQVGEGEKISIHRAYSDKMEAAFVVSEIEKLKNDNGYKYKDFAVLYRTNAQSRNFEEALRRAGIPYRIFGGLKFYDRKEIKDILAYLKIIVNPK
CORYN  317 AANSVIAQNENRREKKLWTALGSGDPIVGYVADNEHDEARFIANEVDALADKG-VAYSDIAVMYRTNNSSRAVEDVFMRTGIPYKVGGTRFYERKEIRDIIAYLRLIDNPD
SERR   292 AANTLIANNDGRMGKNLWTEGGEGEPISLYCAFNELDEARFVVNRIKTWQD-NGGALNDCAILYRSNAQSRVLEEALLQTAMPYRIYGGQRFFERQEIKDALAYLRLISNRN
```

```
SUBT   395 DDISFTRIVNVPKRGVGATSLEKIASYAAINGLSFFQAIQQVDFIG---VSAKAANALDSFRQMIENLTNMQD--------YLSITELTEEILDKTEYREMLKAEKSIEAQS
GEOB   407 DDLSLLRIINVPKRGIGATIDKLVRYAAEHELSLFEALGELEMIG---SAKTAGALAAFRGQLEQWTQLQE--------YVPVTELVEEVLDQSGYREMLKAERTIEAQS
LIST   345 DDISLTRIINVPKRGVGPGTLEKLNNVAAAYDLSLFEVLNRIELAG---ISAKISKDLVAFHDLVRGFTQMQD--------FLSVTELVEEILEKTGYRAMLKNERTIEAQT
STAPS  402 DDISLRRVINIPKRGIGPSSVDKIQAYAAQNDLSMFDAIAEVDFIG---LSKKVTQECISFYDVMQNLIKQQE--------FLEITEIVEEVLTKTGYRDMLEREQTLESRS
STAPA  402 DDISLQRIINVPKRGIGPSSVEKVQNYALQNNISMFDALGEADFIG---LSKKVTQECLNFYELIQSLIKEQE--------FLEIHEIVDEVLQKSGYREMLERENTLESRS
LACT   405 DSMSFERIVNSPKRGIGPGTLEKLGQFADDHQWSLLEAAQNAELST---IPARSRNTLIDFGNVIGDLAKMRE--------FLNVTDLTEQLLDKTGYKKALEVENTLESQT
STREP  402 DNISFERIINEPKRGIGLGTVEKIRDFANLQNMSLLDASANIMLSG---IKGKAAQSIWDFANMMLDLREQLD--------HLSITELVESVLEKTGYVDILNAQATLESKA
CLOST  402 DSISLRRIINVPKRSIGDTTVEKLQNHATEIDDTLYNVLLDVDYVPG--LTARSINPIKKFTDMMEEIMVMSE--------QLSVSQLIEYVLEKTGYLKSLKDSKLIEDQS
CORYN  428 DAVSLRRIINTPRRGIGDKAVAFLSLHADNHNVGFHKALLDATEDKVSMLGARGRNAVTGFVDMMKGIREEAANKVNEVTGMPDIGDMVSYILDVTGYKAELEKSNDPQDGA
SERR   403 DDAAFERVVNTPTRGIGDRTLDVVRQAARDRQLTLWQATRELMQDK--VLAGRAASALQRFIELVESLAHETAD--------MPLHVQTDRVIRDSGLFIMYEQEKGEKGQA

SUBT   496 RLENIDEFLSVTKNFEQ--KSEDKT-------------------------LVAFLTDLALIADIDQLDKEEESGGKDAITLMTLHAAKGLEFPVVFLMGLEEGVFPHSRSLM
GEOB   508 RLENLDEFLSVTKHFEN--VSDDKS-------------------------LVAFLTDLALISDLDELNGTDQAAEG-DAVVFMTLHAAKGLEFPVVFLIGMEEGIFPHLRSLE
LIST   446 RLENIDEFLSVTQNFEK--ENDDKT-------------------------LLAFLTDLALVADVDKLEEDNEEQNG--AVTLMTLHSAKGLEFPVVFLVGMEEGIFPHSRAIY
STAPS  503 RLENIDEFMSVPKDYEENTPLEEQS-------------------------LINFLTDLSLVADIDDAQIEDG-------ITLMTMHSAKGLEFPIVFIMGMEESLFPHIRAIK
STAPA  503 RLENIDEFMSVPKDYEENTPLEEQS-------------------------LINFLTDLSLVADIDEADTENG-------VTLMTMHSAKGLEFPIVFIMGMEESLFPHIRAIN
LACT   506 RLENIEELLSVTKQFDEHYEPEEEDS---------------------DLFVDFLAELSLVSDLDSVEEEAS------EVTMMTLHAAKGLEFPVVFLMGMEEGIFPLSRAIL
STREP  503 RVENIEEFLSVTKNFDDTTDVTEEETGL-----------------DKLSRFLNDLALIADTDSGSQETS------EVTLMTLHAAKGLEFPVVFLIGMEENVFPLSRATE
CLOST  504 RIENLEELVSDAVEFEKSNEEDKS-------------------------LSAYLEKVALVQDMDDLEAEDN------YVMLMTVHSSKGLEFPTVFVGMENGIFPNAACFE
CORYN  540 RLDNLNELVSVAREFSSEAANQVAYERMGAENQAPEIPELSEGEAAPGSLMAFLRVSLVADADQIPDNEQG----VVTLMTLHTAKGLEFPVVFLTGWEDGQFPHLRALG
SERR   505 RIENLEELVTATRQFSYQDEYQDLMP----------------------LQAFLSHAALEAGEGHADAYQD------AVQLMTLHSAKGLEFPLVFIVGMEEGMFPSQMSLD

SUBT   582 EEA--EMEEERRLAYVGITRAEQELYLTNAKMRTLFGRTNMNPESRFIAEIPDDLLENLNEKKETRATSARKMQPR------RGPVSRPVSYASKTGGDTLN----------
GEOB   593 NED--EMEEERRLAYVGITRAEEELVLTNAQMRTLFGNIQMNRPSRFLDEIPAHLLETVSRGKPS-------------------VSRSASARLQAGGAIRV----------
LIST   530 EEE--EMEEERRLAYVGITRAEEELFLTSAYSRMLYGRPSSNQESRFIGEIPRDLLELGNENKLTADKPYAKPR----------MPQKATTAYKSSGAETLG----------
STAPS  584 SDDEHEMQEERRICYVAITRAEETLYLTHATSRMLFGRPQSNMPSRFLREIPEDLLENESKGKSKSQAQATSNRFQNATKQPAKRAHSQRATATKQKQSATN----------
STAPA  584 SEDDHEMQEERRICYVAITRAEEVLYITHATSRMLFGRPQSNMPSRFLKEIPESLLENHSSGKRQ--------TIQPKAKPFAKRGFSQRTTSTKKQVSSSD----------
LACT   591 EED--EIEEERRLAYVGITRAESKLYLTNAYSRMLYGRQQSNQASRFVGEIDDQLLDYANGN-------AGSIPERKLPFGKSNSYARATATTYNGPKTTTSHKIADTTGGD
STREP  590 DPD--ELEEERRLAYVGITRAEKILYLTNANSRLLFGRTNYNRPTRFINEISSDLLEYQGLARPANTSFKASYSSGSISFGQGMSLAQALQDRKRGAAPKSIQSSGLPFGQF
CLOST  585 HDN--QMEEARRLCYVGITRAEEKLYMTSAETRMIFGKTVAYGQSDFISEIPANIKDYVNINGKSSSNSFKNTSQTYGETRKTYNPHSIRGGMSVSKANESLNKTLSGKSSG
CORYN  647 EPK--ELAEEERRLAYVGITRARRTLYLSRAMLRSSWGNPVTNPPSRFLSEIPDELMDWRREEPEHSYDNSWGMGYGSSNSRSYGSQNSRSYGSPRTPKLPKHSNKP------
SERR   588 EGG--RLEEERRLAYVGVTRAMQKLTLTYAETRRLYGKEVYHRPSRFIGELPEECVEEVR----------------LRASVSRPVNHRRMGTPISENDTG----------

SUBT   676 -------------WAVGDKAGHKKWGTGTVVSVKGEGEGTELDIAFPSPVGVKRLLAAFAPIEKQ---WAVGDKAGHKKWGTGTVVSVKGEGEGTELDIAFPSPVGVKRLLAA
GEOB   673 -------------WKVGDKANHRKWGIGTVVSVRGEGDDQELDIAFPSPIGIKRLLAKFAPIEKV---WKVGDRANHRKWGIGTVVSVRGEGDDQELDIAFPSPIGIKRLLAK
LIST   620 -------------WTVGDKASHKKWGVGTVVSVKGEGSGMELDIAFPSPTGVKRLLAEFAPIEKV---WTVGDKASHKKWGVGTVVSVKGEGSGMELDIAFPSPTGVKRLLAE
STAPS  686 -------------WNVGDKVMHKSWGEGMVSNVKEKNGSVELDIIFKS-EGPKRLLAQFAPIEKKGE-WNVGDKVMHKSWGEGMVSNVKEKNGSVELDIIFKS-EGPKRLLAQ
STAPA  678 -------------WNVGDKVMHKAWGEGMVSNVKEKNGSIELDIIFKS-QGPKRLLAQFAPIEKV---WNVGDKVMHKAWGEGMVSNVKEKNGSIELDIIFKS-QGPKRLLAQ
LACT   694 KS--------GSWQAGQKVQHKKWGIGTVVKVNGSDQELDIAFKE-QGVKRLLASFAPIEKVE-SWQAGQKVQHKKWGIGTVVKVNGTGSDQELDIAFKE-QGVKRLLAS
STREP  700 TAGAKPASSEANWSIGDIALHKKWGEGTVLEVSGSGARQELKINFPE-VGLKKLLASVAPIEKKI-NWSIGDIALHKKWGEGTVLEVSGSGARQELKINFPE-VGLKKLLAS
CLOST  695 NAG------IDELTMGRKVKHPKFGVGTLITKSKVGDDYKLTIAFDS-QGVKHLMFSFAPLELL--ELTMGRKVKHPKFGVGTLITKSKVGDDYKLTIAFDS-QGVKHLMFS
CORYN  751 ---------AIALNVGDRVNHEKYGLGTVTKVDKMAPADSVTIDFGSSGTVRLMLIGSVPMEKL--ALNVGDRVNHKYGLGTVTKVDKMAPADSVTIDFGSSGTVRLMLIG
SERR   670 ------------YKLGQRVRHPKFGEGTIVNLEGSGEHSRLQIAFPG-EGIKWLVAAYARLETV---YKLGQRVRHPKFGEGTIVNLEGSGEHSRLQIAFPG-EGIKWLVA
```

```
SUBT   773  FAPIEKQ--
GEOB   770  FAPIEKV--
LIST   717  FAPIEKV--
STAPS  783  FAPIEKKGE
STAPA  775  FAPIEKKED
LACT   795  FAPIEKVE-
STREP  809  VAPIEKKI-
CLOST  797  FAPLELL--
CORYN  852  SVPMEKL--
SERR   765  YARLETV--
```

Appendix II: Media and Buffers

Luria Bertani (LB) media

10 g tryptone

5 g yeast extract

5 g NaCl

1 ml NaOH (1M)

1 L with MQW - Autoclaved

Nutrient agar (NA)

23 g Oxoid nutrient agar

1 L MQW - autoclaved

Auto-induction media (AIM)

1 % tryptone

0.5 % yeast extract

25 mM Na_2HPO_4

50 mM KH_2PO_4

5 mM Na_2SO_4

2 mM $MgSO_4$

0.05 % glucose

0.2 % lactose

0.5% glycerol - Autoclaved

Selective Media: Antibiotics

Ampicillin – (*E. coli)*

Stock solution 100 mg/ml

Final concentration 100 μg/ml

Chloramphenicol – *(B. subtilis)*

Stock solution 10 mg/ml

Final concentration 5 μg/ml

Erythmycin – (*B. subtilis)*

Stock solution 10 mg/ml

Final concentration 1 µg/ml

E. coli competent cells

RFI

30 mM potassium acetate

10 mM $CaCl_2.2H_2O$

50 mM $MnCl_2.4H_2O$

15 % glycerol

100 mM RbCl

pH 5.8 with 0.2 M acetic acid

filter sterilised

RFII

10 mM MOPS

75 mM $CaCl_2.H_2O$

15 % glycerol

50 mM RbCl

pH 6.8 with NaOH

filter sterilised

Small Scale Plasmid Prep -Alkaline Lysis Method

SDS-alkaline

0.2 NaOH

1% SDS

Make up fresh

TEG

50 mM Glucose

10 mM EDTA

25 mM Tris-HCl, pH 8.0

High Salt Buffer

3 M Potassium Acetate

2 M Acetic Acid

B. subtilis Transformation

10 x PC

0.6 M K_2HPO_4	(21.4 g, FW = 174.18)
0.4 M KH_2PO_4	(12 g, FW = 136.09)
0.03 M Na_3citrate.$5H_2O$	(2 g, FW = 348.1)
0.03 M Na_3citrate.$2H_2O$	(1.7 g, FW = 294.1) = 0.06 Na_3 citrate

Make up to 200 ml with MQW and autoclave

MD media

2 ml 10 x PC

0.5 ml L-tryptophan (2 mg/ml)

0.1 ml Ferric ammonium citrate (2.2 mg/ml, filter sterilised)

1 ml L-aspartate (50 mg/ml, pH 7.0 with KOH, filter sterilised)

1 ml 40% glucose

60 μl $MgSO_4$ (1M)

15.34 ml MQW

DNA Agarose Gel Electrophoresis

50 x TAE

2 M Tris-HCl

50 mM Glacial Acetic Acid

0.1 M EDTA

DNA Loading Buffer (1ml)

200 µl Loading Dye

800 µl 50% Glycerol

SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE)

4 x Resolving Buffer

90.83 g Tris

2.0 g SDS

pH 8.8 with HCl

Make up to 500 ml with dH$_2$O

4 x Stacking Buffer

0.5 M Tris

0.4 % (w/v) SDS

pH 6.8 with HCl

10 x SDS Running Buffer

30.02 g Tris

188 g Glycine

10 g SDS

Make up to 1 L with MQW

2 x SDS Loading Buffer

0.15 g DTT

1.25 ml 0.5 M Tris-HCl (pH 6.8)

5.0 ml 50 % (v/v) Glycerol

1.0 ml 20 % (w/v) SDS

0.6 ml 0.5 % (w/v) Bromophenol blue

~ 2.55 ml MQW

Resolving Gel (10%) for 10 ml

5.0 ml MQW

2.5 ml 40 % Acrylamide

2.5 ml Resolving buffer

200 μl 10 % (w/v) APS

15 μl TEMED

Stacking Gel (6%)

3.2 ml MQW

0.675 ml 40 % Acrylamide

1.25 ml stacking buffer

50 μl 10 % (w/v) APS

15 μl TEMED

Coomassie Blue Stain

0.25 g Coomassie brilliant blue

20 ml Methanol

10 ml Acetic Acid

70 ml MQW

Destain

10 % (v/v) Acetic Acid

20 % (v/v) Methanol

Western Blotting

Transfer buffer

100 ml 10x SDS running buffer

200 ml methanol

Make up to 1L with MQW

Blocking buffer

1 x PBS

0.1 % (v/v) Tween 20

5 % (w/v) skim milk

Wash buffer

1 x PBS

0.1 % (v/v) Tween 20

Far Western Blotting

TFWB wash buffer

20 mM Tris

100 mM NaCl

0.5 mM EDTA

10 % (v/v) Glycerol

0.1 % (v/v) Tween-20

pH 7.6

TFWB refold/block buffer

TFWB wash buffer containing 5 % (w/v) skim milk powder

TFWB probe buffer

TFWB refold buffer containing the protein probe (~ 40 nM)

Overproduction of proteins

IPTG

Final concentration 0.05 - 0.5 mM

10 x TES

2 M Tris-HCl (pH 7.5)

50 mM EDTA

1 M NaCl

Lysozyme

10 mg/ml in TES or lysis buffer

PcrA Purification

Lysis Buffer

50 mM Tris pH 8.0

0.1 % (v/v) triton

1 mM DTT

Polymin P Wash Buffer

50 mM Tris pH 8.0

1 mM DTT

0.4 M NaCl

Phenyl Sepharose Resuspension Buffer

50 mM Tris pH 8.0

0.2 M NaCl

Phenyl Sepharose Buffer A

50 mM tris pH 8.0

Phenyl Sepharose Buffer B

50 mM Tris pH 8.0

26.2 g NH_4SO_4 / 100ml

Heparin Buffer A (equilibrating)

50 mM Tris pH 8.0

100 mM NaCl

1 mM DTT

Heparin Buffer A (gradient)

50 mM Tris pH 8.0

250 mM NaCl

1 mM DTT

Heparin Buffer B

50 mM Tris pH 8.0

1.25 M NaCl

Dialysis (Storage) Buffer

50 mM Tris

100 mM NaCl

50 % (v/v) glycerol

Nickel affinity purification

Lysis buffer

20 mM KH_2PO_4

500 mM NaCl

20 mM Imidazole

0.5 mg/ml Lysozyme

pH 8.0 with NaOH

Buffer A – binding buffer

20 mM KH_2PO_4

500 mM NaCl

20 mM Imidazole

pH 8.0 with NaOH

Buffer B – elution buffer

20 mM KH_2PO_4

500 mM NaCl

500 mM Imidazole

pH 8.0 with NaOH

Insoluble nickel affinity purification

Resuspension/binding buffer

20 mM KH_2PO_4

500 mM NaCl

20 mM Imidazole

8 M Urea

pH 8.0 with NaOH

Affinity chromatography

Wash Buffer

50 mM NaH_2PO_4

100 mM NaCl

pH 7.5

Tandem affinity purification (TAP-tag)

Buffer A

10 mM Tris-HCl pH 7.5

150 mM NaCl

Buffer B

10 mM Tris-HCl pH 7.5

150 mM NaCl

0.2 mM EDTA

0.1 % (v/v) TritonX-100

TEV buffer

50 mM Tris-HCl pH 7.5

150 mM NaCl

0.2 mM EDTA

0.1 % (v/v) TritonX-100

1 mM DTT

CBB buffer

10 mM Tris-HCl pH 7.5

150 mM NaCl

0.1 % (v/v) TritonX-100

1 mM DTT

2 mM $CaCl_2$

CWB buffer

10 mM Tris-HCl pH 7.5

150 mM NaCl

0.1 % (v/v) TritonX-100

1 mM DTT 0.1 mM $CaCl_2$

CEB buffer

50 mM Tris-HCl pH 7.5

150 mM NaCl

1 mM DTT

3 mM EGTA

Single particle reconstruction

Complex buffer

10 mM Tris-HCl pH 7.8

150 mM NaCl

10 mm $MgCl_2$

Appendix III: Refinement loop for the single particle reconstruction of the PcrA-RNAP complex

Initial model: threed.0a.mrc same as RNAP core model 39 (Yang, et al 2009)

Particles filtered: start.* lp=13 and start.filt.* lp=13 hp=210

Refinement loop

refine 2 mask=45 hard=30 pad=144 ang=6 classkeep=1 classiter=8 usefilt refine goodbad proc=16 median phasecls

refine 8 mask=45 hard=30 pad=144 ang=6 classkeep=1 classiter=5 usefilt refine goodbad proc=16 median phasecls

- model 6 was filtered to 20Å as the initial model

refine 2 mask=45 hard=30 pad=144 ang=6 classkeep=1 classiter=8 usefilt refine goodbad proc=8 median phasecls

- class averages with significant bias was manually removed and the 2nd model was filtered

proc3d threed.2a.mrc threed.2filt.mrc apix=3.9 automask2=6,1.3,3.6

Refinement loop

refine 2 mask=45 hard=30 pad=144 ang=4 classkeep=1 classiter=8 usefilt refine goodbad proc=8 median phasecls

refine 8 mask=45 hard=30 pad=144 ang=4 classkeep=1 classiter=5 usefilt refine goodbad proc=8 median phasecls

refine 14 mask=45 hard=30 pad=144 ang=4 classkeep=1 classiter=3 usefilt refine goodbad proc=8 median phasecls

refine 16 mask=45 hard=30 pad=144 ang=2 classkeep=1 classiter=8 usefilt refine goodbad proc=8 median phasecls

Initial model: threed.0a.mrc same as 16th model from above

Particles filtered: start.* lp=15 and start.filt.* lp=20 hp=210

Refinement loop

refine 2 mask=45 hard=30 pad=144 ang=4 classkeep=1 classiter=8 usefilt refine goodbad proc=8 median phasecls

refine 8 mask=45 hard=30 pad=144 ang=4 classkeep=1 classiter=5 usefilt refine goodbad proc=8 median phasecls

refine 14 mask=45 hard=30 pad=144 ang=4 classkeep=1 classiter=3 usefilt refine goodbad proc=8 median phasecls

Appendix IV: PcrA protein sequence identity and similarity

1 :

QUERY – *B. subtilis* PcrA

SUBJECT – *G. stearothermphilus* PcrA

IDENTITY – 70%   **SIMILARITY – 84%**

2 :

QUERY – *B. subtilis* PcrA

SUBJECT – *E. coli* UvrD

IDENTITY – 41%   **SIMILARITY – 61%**

3 :

QUERY – *G. stearothermphilus* PcrA

SUBJECT – *E. coli* UvrD

IDENTITY – 43%   **SIMILARITY – 63%**

1: QUERY – *G. stearothermophilus* PcrA

SUBJECT – *B. subtilis* PcrA

```
Query    1   MNFLSEQLLAHLNKEQQEAVRTTEGPLLIMAGAGSGKTRVLTHRIAYLMAEKHVAPWNIL   60
             MN++S QLL+ LN  QQEAV+TT+GPLL+MAGAGSGKTRVLTHRIAYLMAEKHVAPWNIL
Sbjct    1   MNYISNQLLSGLNPVQQEAVKTTDGPLLLMAGAGSGKTRVLTHRIAYLMAEKHVAPWNIL   60

Query   61   AITFTNKAAREMRERVQSLLGGAAEDVWISTFHSMCVRILRRDIDRIGINRNFSILDPTD   120
             AITFTNKAAREM+ERV+S+LG  A+D+WISTFHSMCVRILRRDIDRIGINRNFSILD  D
Sbjct   61   AITFTNKAAREMKERVESILGPGADDIWISTFHSMCVRILRRDIDRIGINRNFSILDTAD   120

Query  121   QLSVMKTILKEKNIDPKKFEPRTILGTISAAKNELLPPEQFAKRASTYYEKVVSDVYQEY   180
             QLSV+K ILKE+N+DPKKF PR+ILGTIS+AKNEL  PE+F+K A  YY++VVSDVY +Y
Sbjct  121   QLSVIKGILKERNLDPKKFDPRSILGTISSAKNELTEPEEFSKVAGGYYDQVVSDVYADY   180

Query  181   QQRLLRNHSLDFDDLIMTTIQLFDRVPDVLHYYQYKFQYIHIDEYQDTNRAQYTLVKKLA   240
             Q++LL+N SLDFDDLIMTTI+LFDRVP+VL +YQ KFQYIH+DEYQDTNRAQY LVK+LA
Sbjct  181   QKKLLKNQSLDFDDLIMTTIKLFDRVPEVLEFYQRKFQYIHVDEYQDTNRAQYMLVKQLA   240

Query  241   ERFQNICAVGDADQSIYRWRGADIQNILSFERDYPNAKVILLEQNYRSTKRILQAANEVI   300
             ERFQN+C VGD+DQSIYRWRGADI NILSFE+DYPNA VILLEQNYRSTKRIL+AANEVI
Sbjct  241   ERFQNLCVVGDSDQSIYRWRGADITNILSFEKDYPNASVILLEQNYRSTKRILRAANEVI   300

Query  301   EHNVNRKPKRIWTENPEGKPILYYEAMNEADEAQFVAGRIREAVERGERRYRDFAVLYRT   360
             ++N NRKPK +WTEN EG  I YY   NE E QFVAG+I +   G+R+  D A+LYRT
Sbjct  301   KNNSNRKPKNLWTENDEGIKISYYRGDNEFGEGQFVAGKIHQLHSTGKRKLSDIAILYRT   360

Query  361   NAQSRVMEEMLLKANIPYQIVGGLKFYDRKEIKDILAYLRVIANPDDDLSLLRIINVPKR   420
             NAQSRV+EE LLKA + Y IVGG KFYDRKEIKDILAYLR+++NPDDD+S  RI+NVPKR
Sbjct  361   NAQSRVIEETLLKAGLNYNIVGGTKFYDRKEIKDILAYLRLVSNPDDDISFTRIVNVPKR   420

Query  421   GIGASTIDKLVRYAADHELSLFEALGELEMIGLGAKAAGALAAFRSQLEQWTQLQEYVSV   480
             G+GA++++K+  YAA + LS F+A+ +++ IG+ AKAA AL +FR   +E  T +Q+Y+S+
Sbjct  421   GVGATSLEKIASYAAINGLSFFQAIQQVDFIGVSAKAANALDSFRQMIENLTNMQDYLSI   480

Query  481   TELVEEVLDKSGYREMLKAERTIEAQSRLENLDEFLSVTKHFENVSDDKSLIAFLTDLAL   540
             TEL EE+LDK+ YREMLKAE++IEAQSRLEN+DEFLSVTK+FE  S+DK+L+AFLTDLAL
Sbjct  481   TELTEEILDKTEYREMLKAEKSIEAQSRLENIDEFLSVTKNFEQKSEDKTLVAFLTDLAL   540

Query  541   ISDLDELDGTEQAAEG-DAVMLMTLHAAKGLEFPVVFLIGMEEGIFPHNRSLEDDDEMEE   599
             I+D+D+LD  E+ + G DA+ LMTLHAAKGLEFPVVFL+G+EEG+FPH+RSL ++ EMEE
Sbjct  541   IADIDQLDQKEEESGGKDAITLMTLHAAKGLEFPVVFLMGLEEGVFPHSRSLMEEAEMEE   600

Query  600   ERRLAYVGITRAEEELVLTSAQMRTLFGNIQMDPPSRFLNEIPAHLLE----------TA   649
             ERRLAYVGITRAE+EL LT+A+MRTLFG  M+P SRF+ EIP  LLE          T+
Sbjct  601   ERRLAYVGITRAEQELYLTNAKMRTLFGRTNMNPESRFIAEIPDDLLENLNEKKETRATS   660

Query  650   SR----RQAGASRPAVSRPQASGAVGSWKVGDRANHRKWGIGTVVSVRGGGDDQELDIAF   705
             +R    R+   SRP   + G   +W VGD+A H+KWG GTVVSV+G G+   ELDIAF
Sbjct  661   ARKMQPRRGPVSRPVSYASKTGGDTLNWAVGDKAGHKKWGTGTVVSVKGEGEGTELDIAF   720

Query  706   PSPIGIKRLLAKFAPIEK   723
             PSP+G+KRLLA FAPIEK
Sbjct  721   PSPVGVKRLLAAFAPIEK   738
```

2: QUERY – *B. subtilis* PcrA

SUBJECT – *E. coli* UvrD

```
GENE ID: 948347 uvrD | DNA-dependent ATPase I and helicase II
[Escherichia coli str. K12 substr. MG1655] (Over 100 PubMed links) Score =
555 bits (1429), Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 305/738 (41%), Positives = 452/738 (61%), Gaps = 33/738 (4%)


Query  8    LLSGLNPVQQEAVKTTDGPLLLMAGAGSGKTRVLTHRIAYLMAEKHVAPWNILAITFTNK   67
            LL  LN  Q+EAV       LL++AGAGSGKTRVL HRIA+LM+ ++ +P++I+A+TFTNK
Sbjct  6    LLDSLNDKQREAVAAPRSNLLVLAGAGSGKTRVLVHRIAWLMSVENCSPYSIMAVTFTNK   65


Query  68   AAREMKERVESILGPGADDIWISTFHSMCVRILRRDIDRIGINRNFSILDTADQLSVIKG   127
            AA EM+ R+  ++G      +W+ TFH + R+LR      + ++F ILD+ DQL ++K
Sbjct  66   AAAEMRHRIGQLMGTSQGGMWVGTFHGLAHRLLRAHHMDANLPQDFQILDSEDQLRLLKR   125


Query  128  ILKERNLDPKKFDPRSILGTISSAKNELTEPEEFSKVAGGYYDQVVSDVYADYQKKLLKN   187
            ++K  NLD K++ PR + I+S K+E  P     G  +Q   VY  YQ+   +
Sbjct  126  LIKAMNLDEKQWPPRQAMWYINSQKDEGLRPHHIQSY-GNPVEQTWQKVYQAYQEACDRA   184


Query  188  QSLDFDDLIMTTIKLFDRVPEVLEFYQRKFQYIHVDEYQDTNRAQYMLVKQLAERFQNLC   247
             +DF +L++    +L+   P +L+ Y+ +F  I VDE+QDTN  QY  ++ LA     +
Sbjct  185  GLVDFAELLLRAHELWLNKPHILQHYRERFTNILVDEFQDTNNIQYAWIRLLAGDTGKVM   244


Query  248  VVGDSDQSIYRWRGADITNILSFEKDYPNASVILLEQNYRSTKRILRAANEVIKNNSNRK   307
            +VGD DQSIY WRGA + NI  F  D+P A  I LEQNYRST  IL AAN +I+NN+ R
Sbjct  245  IVGDDDQSIYGWRGAQVENIQRFLNDFPGAETIRLEQNYRSTSNILSAANALIENNNGRL   304


Query  308  PKNLWTENDEGIKISYYRGDNEFGEGQFVAGKIHQLHSTGKRKLSDIAILYRTNAQSRVI   367
             K LWT+  +G IS Y  NE  E +FV  +I    G  L++ AILYR+NAQSRV+
Sbjct  305  GKKLWTDGADGEPISLYCAFNELDEARFVVNRIKTWQDNGG-ALAECAILYRSNAQSRVL   363


Query  368  EETLLKAGLNYNIVGGTKFYDRKEIKDILAYLRLVSNPDDDISFTRIVNVPKRGVGATSL   427
            EE LL+A + Y I GG +F++R+EIKD L+YLRL++N +DD +F R+VN P RG+G  +L
Sbjct  364  EEALLQASMPYRIYGGMRFFERQEIKDALSYLRLIANRNDDAAFERVVNTPTRGIGDRTL   423


Query  428  EKIASYAAINGLSFFQAIQQ-VDFIGVSAKAANALDSFRQMIENLTNMQDYLSITELTEE   486
            + +   +   L+ +QA ++ +    ++ +AA+AL  F ++I+ L     + + +  T+
Sbjct  424  DVVRQTSRDRQLTLWQACRELLQEKALAGRAASALQRFMELIDALAQETADMPLHVQTDR   483


Query  487  ILDKTEYREMLKAEKSIEAQSRLENIDEFLSVTKNFEQKSEDKTLV---AFLTDLALIAD   543
            ++ +   R M + EK  + Q+R+EN++E ++ T+ F    ED+ L+   AFL+  AL A
Sbjct  484  VIKDSGLRTMYEQEKGEKGQTRIENLEELVTATRQFSYNEEDEDLMPLQAFLSHAALEAG   543


Query  544  IDQLDQKEEESGGKDAITLMTLHAAKGLEFPVVFLMGLEEGVFPHSRSLMEEAEMEEERR   603
             Q D      +DA+ LMTLH+AKGLEFP VF++G+EEG+FP    SL E  +EEERR
Sbjct  544  EGQADT------WQDAVQLMTLHSAKGLEFPQVFIVGMEEGMFPSQMSLDEGGRLEEERR   597


Query  604  LAYVGITRAEQELYLTNAKMRTLFGRTNMNPESRFIAEIPDDLLENLNEKKETRATSARK   663
            LAYVG+TRA Q+L LT A+ R L+G+   +  SRFI E+P++ +E +        RAT
Sbjct  598  LAYVGVTRAMQKLTLTYAETRRLYGKEVYHRPSRFIGELPEECVEEVR----LRAT----   649


Query  664  MQPRRGPVSRPVSYASKTGGDTL----NWAVGDKAGHKKWGTGTVVSVKGEGEGTELDIA   719
                   VSRPVS+ + G  +      + +G + H K+G GT+V+++G GE + L +A
Sbjct  650  -------VSRPVSH-QRMGTPMVENDSGYKLGQRVRHAKFGEGTIVNMEGSGEHSRLQVA   701


Query  720  FPSPVGVKRLLAAFAPIE   737
            F     G+K L+AA+A +E
Sbjct  702  FQGQ-GIKWLVAAYARLE   718
```

3: **QUERY** – *G. stearothermophilus* PcrA

SUBJECT – *E. coli* UvrD

```
Query  8    LLAHLNKEQQEAVRTTEGPLLIMAGAGSGKTRVLTHRIAYLMAEKHVAPWNILAITFTNK  67
            LL  LN +Q+EAV     LL++AGAGSGKTRVL HRIA+LM+ ++ +P++I+A+TFTNK
Sbjct  6    LLDSLNDKQREAVAAPRSNLLVLAGAGSGKTRVLVHRIAWLMSVENCSPYSIMAVTFTNK  65

Query  68   AAREMRERVQSLLGGAAEDVWISTFHSMCVRILRRDIDRIGINRNFSILDPTDQLSVMKT  127
            AA EMR R+  L+G +   +W+ TFH +  R+LR      + ++F ILD  DQL ++K
Sbjct  66   AAAEMRHRIGQLMGTSQGGMWVGTFHGLAHRLLRAHHMDANLPQDFQILDSEDQLRLLKR  125

Query  128  ILKEKNIDPKKFEPRTILGTISAAKNELLPPEQFAKRASTYYEKVVSDVYQEYQQRLLRN  187
            ++K  N+D K++ PR  +  I++ K+E L P        +  E+   VYQ YQ+   R
Sbjct  126  LIKAMNLDEKQWPPRQAMWYINSQKDEGLRPHHIQSYGNPV-EQTWQKVYQAYQEACDRA  184

Query  188  HSLDFDDLIMTTIQLFDRVPDVLHYYQYKFQYIHIDEYQDTNRAQYTLVKKLAERFQNIC  247
            +DF +L++    +L+  P +L +Y+ +F  I +DE+QDTN  QY   ++ LA       +
Sbjct  185  GLVDFAELLLRAHELWLNKPHILQHYRERFTNILVDEFQDTNNIQYAWIRLLAGDTGKVM  244

Query  248  AVGDADQSIYRWRGADIQNILSFERDYPNAKVILLEQNYRSTKRILQAANEVIEHNVNRK  307
             VGD DQSIY WRGA ++NI  F  D+P A+ I  LEQNYRST   IL AAN +IE+N R
Sbjct  245  IVGDDDQSIYGWRGAQVENIQRFLNDFPGAETIRLEQNYRSTSNILSAANALIENNNGRL  304

Query  308  PKRIWTENPEGKPILYYEAMNEADEAQFVAGRIREAVERGERRYRDFAVLYRTNAQSRVM  367
            K++WT+  +G+PI  Y A NE DEA+FV  RI+    + G    + A+LYR+NAQSRV+
Sbjct  305  GKKLWTDGADGEPISLYCAFNELDEARFVVNRIKTWQDNGG-ALAECAILYRSNAQSRVL  363

Query  368  EEMLLKANIPYQIVGGLKFYDRKEIKDILAYLRVIANPDDDLSLLRIINVPKRGIGASTI  427
            EE LL+A++PY+I GG++F++R+EIKD L+YLR+IAN +DD +  R++N P RGIG  T+
Sbjct  364  EEALLQASMPYRIYGGMRFFERQEIKDALSYLRLIANRNDDAAFERVVNTPTRGIGDRTL  423

Query  428  DKLVRYAADHELSLFEALGE-LEMIGLGAKAAGALAAFRSQLEQWTQLQEYVSVTELVEE  486
            D + + + D +L+L++ A  E L+   +AA AL  F   ++ Q    + +    +
Sbjct  424  DVVRQTSRDRQLTLWQACRELLQEKALAGRAASALQRFMELIDALAQETADMPLHVQTDR  483

Query  487  VLDKSGYREMLKAERTIEAQSRLENLDEFLSVTKHFENVSDDKSLI---AFLTDLALISD  543
            V+  SG R M + E+  + Q+R+ENL+E ++ T+ F     +D+ L+   AFL+  AL +
Sbjct  484  VIKDSGLRTMYEQEKGEKGQTRIENLEELVTATRQFSYNEEDEDLMPLQAFLSHAALEAG  543

Query  544  LDELDGTEQAAEGDAVMLMTLHAAKGLEFPVVFLIGMEEGIFPHNRSLEDDDEMEEERRL  603
            + + D +      DAV LMTLH+AKGLEFP VF++GMEEG+FP    SL++   +EEERRL
Sbjct  544  EGQADTWQ-----DAVQLMTLHSAKGLEFPQVFIVGMEEGMFPSQMSLDEGGRLEEERRL  598

Query  604  AYVGITRAEEELVLTSAQMRTLFGNIQMDPPSRFLNEIPAHLLETASRRQAGASRPAVSR  663
            AYVG+TRA ++L LT A+ R L+G     PSRF+ E+P  +E  R +A  SRP   +
Sbjct  599  AYVGVTRAMQKLTLTYAETRRLYGKEVYHRPSRFIGELPEECVEEV-RLRATVSRPVSHQ  657

Query  664  PQASGAV---GSWKVGDRANHRKWGIGTVVSVRGGGDDQELDIAFPSPIGIKRLLAKFAP  720
            + V     +K+G R  H K+G GT+V++ G G+    L +AF    GIK L+A +A
Sbjct  658  RMGTPMVENDSGYKLGQRVRHAKFGEGTIVNMEGSGEHSRLQVAFQGQ-GIKWLVAAYAR  716

Query  721  IEKV  724
            +E V
Sbjct  717  LESV  720
```

Appendix V: PcrA Structure

To view the structural changes when PcrA binds to DNA please see the **PcrA structural morph.py** on the supplied CD. This movie is view using the program UCSF Chimera. The .exe file to install UCSF Chimera on Windows XP, Vista and & has also been provided. Alternatively it can be downloaded for free at http://www.cgl.ucsf.edu/chimera/download.html.

Morphing shows the comparison between unbound PcrA (1PJR) and PcrA in complex with DNA (3PJR). Pressing play in the MD movie control panel will morph the 1PJR to match the structure of 3PJR. Importantly the RNAP interacting regions of the two PcrA structures are well aligned.

PcrA (1PJR) – dark green, PcrA in complex with DNA (3PJR) – light green.

PcrA amino acids 85-189 – blue, 189-310 – yellow, and the region immediately adjacent to the unstructured CTD – pink