Author:
Thompson-Carter, Shani L

Title:
Experimental and Computational Studies of PcrA helicase Interactions with Partner Proteins

Author:
Thompson-Carter, Shani L

Title:
Experimental and Computational Studies of PcrA helicase Interactions with Partner Proteins

Experimental and Computational Studies of PcrA helicase Interactions with Partner Proteins

Shani Thompson-Carter

MScR Thesis

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Master of Science by Research in the Faculty of Biomedical Sciences, School of Biochemistry, March 2023.

Word Count: 23,709

Abstract

Playing important roles in rolling circle DNA replication, homologous recombination and DNA repair, PcrA belongs to the well-established UvrD-like family of helicases. PcrA's involvement in many DNA transactions is directed by interactions with several partner proteins, including RNA polymerase (RNAP). Recent structural data uncovered interactions between the core of PcrA and areas surrounding the DNA and RNA exit channels of RNAP. Additional contacts between PcrA's C-terminal domain and RNAP's SI1 domain identified a putative helicase interaction motif (HIM), also found in several other PcrA interactors. Furthermore, biochemical assays implicated PcrA in DNA:RNA hybrid unwinding and R-loop homeostasis, but the mechanism behind this remains unknown.

A high resolution structure of the PcrA:RNAP complex is now required to determine its precise contacts and to assist in establishing the mechanism by which PcrA resolves R-loops. In this thesis, several methods for formation of a stable PcrA:RNAP complex for downstream imaging applications are explored. We find that individually combining native *B.subtilis* RNAP with recombinant *B.subtilis* PcrA in the presence of an artificial DNA:RNA scaffold yields a stable complex but that other approaches, including co-expression of interaction partners and the use of recombinant RNAP, fail. Moreover, EMSAs showed that multiple shifted species were generated upon addition of PcrA to the RNAP:scaffold complex, revealing an unexpected complexity in the species formed that requires further investigation.

As an alternative and complementary approach to understanding the interactions of PcrA with partner proteins, we use AlphaFold-Multimer to model PcrA complexes with partners containing a HIM. High confidence contacts were predicted between PcrA's CTD and both the SI1 domain of RNAP and a region of YxaL containing the HIM, yielding testable models for the complexes formed. A lower confidence interaction was also predicted between PcrA's 2B domain and RNAP's β-subunit, which is consistent with structural data, but requires experimental validation.

Acknowledgements

Firstly, I wish to express my sincere gratitude to my project supervisor, Professor Mark Dillingham, for his expert guidance, feedback, warm encouragement and support, both inside and outside of the lab. His support both for the practical aspect of the project and the writing of my thesis has been immense, and I am extremely grateful for the positive experience that I have had, largely due to his supervision. I am also very appreciative of his willingness and openness to supervise me for my MScR at the outset.

I want to extend this thanks to the rest of the Dillingham group, as well to Dr Alan Cheung and Professor Nigel Savery for their advice and encouragement of my development throughout the project. In particular, I want to thank Shreya for her endless hands-on help and selflessness inside the lab, as well as her friendship outside of the lab. I would have been lost on many occasions without her guidance and knowledge, so I am forever indebted to her.

I also want to thank the D40 group as a whole for providing a supportive and welcoming environment to work in, as well as great company throughout the last year and a half and several great friendships. The weekly Robin Hood trips certainly kept me sane during the challenging periods and they will be greatly missed!

Finally, I would like to thank my parents for their unconditional support and the sacrifices that they have made to ensure that I always had the resources and opportunities available to me to pursue my passions and to succeed. I also owe a very large thank you to my close friends and family (you know who you are) for their care, advice, support and the extreme positive impact they have on my confidence and well-being, not only related to my academic pursuits, but also in my personal life. I appreciate everything that they have done and continue to do for me.

Author’s Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: [redacted] DATE: 17.03.2023

Table of Contents

List of Tables and Figures

List of Abbreviations

ALS4 – amyotrophic lateral sclerosis type 4

AOA2 – ataxia with oculomotor apraxia type 2

ATP – adenosine triphosphate

CASP – Critical Assessment of Structure Prediction

CTD – C-terminal domain

dsDNA – double-stranded DNA

DTT – dithiothreitol

EDTA – ethylenediaminetetraacetic acid

EMSA – electrophoretic mobility shift assay

HDX – hydrogen-deuterium exchange

HIM – helicase interaction motif

IEX – ion-exchange chromatography

IMAC – immobilised metal affinity chromatography

IPTG – isopropyl β-D-1-thiogalactopyranoside

LB – Luria Bertani

LBA – Luria Bertani Agar

MS – mass spectrometry

MWCO – molecular weight cut-off

NER – nucleotide excision repair

NTP – nucleoside triphosphate

NTS – non-template strand

PAGE – polyacrylamide gel electrophoresis

PDB – Protein Data Bank

PISA – Protein interfaces, surfaces and assemblies

RNAP – RNA polymerase

RT – room temperature

SDS – sodium dodecyl sulfate

SEC – size-exclusion chromatography

SF – superfamily

ssDNA – single-stranded DNA

TCR – transcription-coupled repair

TEC – transcription elongation complex

TS – template strand

Chapter 1

Introduction

1.1 PcrA and its interaction with RNA polymerase

Fundamental to a plethora of essential processes, including DNA repair, replication and transcription, helicases are ubiquitous DNA motor proteins that utilise energy from nucleoside triphosphate (NTP) hydrolysis to partition or move along nucleic acid strands (Gorbalenya and Koonin, 1993). Since their discovery in the 1970's, they have been ordered into 6 superfamilies (SF) based upon conserved motifs within their primary structures (Singleton et al., 2007). Belonging to SF1, the UvrD-like helicases are some of the most widely studied motor proteins, and include *Escherichia coli* UvrD and *Bacillus subtilis* PcrA (Dillingham et al., 2001). UvrD and PcrA were identified independently in Gram-negative (*E. coli*) and Gram-positive (*Staphylococcus aureus* and *B.subtilis*) bacteria respectively, whereby UvrD was implicated in nucleotide excision repair (NER) whilst PcrA was shown to promote plasmid rolling circle replication as a plasmid copy number regulatory factor (Caron et al., 1985, Petit et al., 1998*,* Carr et al., 2016). However, based on similarities between their functions, biochemical characteristics and primary sequences, they are considered to be orthologues.

PcrA functions by translocating along ssDNA in the 3′-5′ direction using an inchworm mechanism, hydrolysing one ATP molecule per base movement (Velankar et al., 1999, Dillingham et al., 2000). Its *Bacillus stearothermophilus* orthologue has been structurally characterised using x-ray crystallography, in which four domains (1A, 1B, 2A and 2B) were identified (Figure 1-1) (Subramanya et al., 1996, Velankar et al., 1999). These structures unveiled that upon a series of conformational changes initiated by ssDNA binding, domains 1B and 2B interact with duplex DNA whilst domains 1A and 2A have motor activity, with ATP hydrolysis occurring at a site located at the interface of the 1A and 2A domains (Figure 1-1, black ATP). The same structures, as well as structures of UvrD, identified that PcrA requires a region of 8-9 nucleotides for stable DNA binding (Velankar et al., 1999, Lee and Yang, 2006). In addition to the core domains, PcrA possesses a peripheral C-terminal domain (CTD), which is attached to the main body of the helicase by a long flexible linker, and adopts a Tudor fold (Sanders et al., 2017). This CTD is central to an interaction between PcrA and RNA polymerase (RNAP) and it is also both known and predicted to be important in interactions with other partners (Manelyte et al., 2009, Urrutia-Irazabal et al., 2021). Furthermore, there have been reports of PcrA dimerization, and it has been speculated that this occurs through its CTD, but that oligomerisation is not essential for its activity (Mechanic et al., 1999a, Leah et al., 1999, Yang et al., 2008). Epshtein et al. (2014) also proposed that UvrD is capable of associating with RNAP both in monomeric and multimeric forms. However, several monomeric crystal structures of both UvrD and PcrA have been obtained, and little convincing evidence for dimerization through the CTD exists (Mechanic et al., 1999b, Velankar et al., 1999, Lee

and Yang, 2006). As a result, the idea of oligomerisation in these helicases is disputed and remains as a debate within the field.

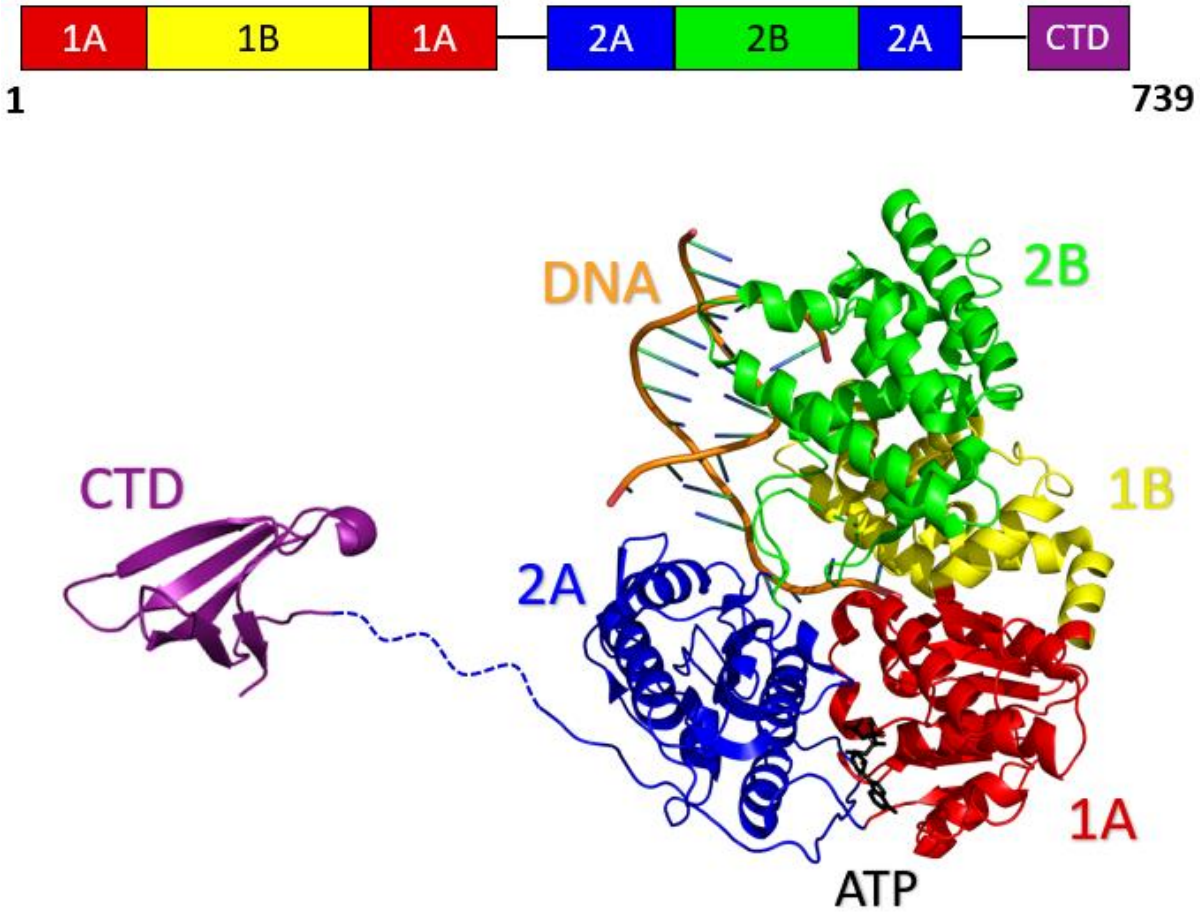

Figure 1-1. Structure of PcrA. Primary and tertiary domain organisation of *B.stearothermophilus* PcrA, homologous to *B.subtilis* PcrA and *E.coli* UvrD. The structure shows location of DNA and ATP binding. The CTD is proposed to join the main body of the helicase by a long flexible linker (blue dashed line), but this is not present in any available structure. (PDB: 3PJR and 5DMA).

PcrA is also essential to many different processes, including homologous recombination, NER and likely, mismatch repair (Lahue et al., 1989, Merrikh et al., 2015, Urrutia-Irazabal et al., 2021). Akin to SF1 helicases, the diversity of PcrA/UvrD's functions can be attributed to its ability to physically interact with several partner proteins, including RecA, UvrB and MutL (Veaute et al., 2005, Manelyte et al., 2009, Machón et al., 2010). In fact, PcrA's helicase activity is low in vitro but its processivity has shown to increase on interaction with partner proteins (Soultanas et al., 1999, Noirot-Gros et al., 2002b, Niedziela-Majka et al., 2007, Zhang et al., 2007). In recent years, there has been particular interest in the interaction between PcrA and RNAP, suggesting the involvement of PcrA in transcription (Noirot-Gros et al., 2002a, Delumeau et al., 2011, Gwynn et al., 2013, Urrutia-Irazabal et al., 2021). However, the function

of the interaction remains unclear, but three interconnected hypotheses exist: PcrA's involvement in transcription-replication conflicts, transcription-coupled repair (TCR) and R-loop dissolution.

The transcription machinery frequently encounters proteins involved in other DNA transactions, such as replication, since they occupy the same substrates. This can either occur co-directionally, whereby transcription and replication occur in the same direction, or head-on, whereby they converge, with the latter having the most damaging effects (Figure 1-2). This shared involvement of DNA also means that these protein complexes are inevitably confronted with DNA lesions. These encounters can be detrimental to the cell and so cells require strict and systematic responses to alleviate conflict and to promote harmonious interactions between the processes. Research from *E.coli* has already identified a role for UvrD in easing transcription-replication conflicts, whereby it acts as an accessory helicase to assist translocation past nucleoprotein barriers, including stalled transcription elongation complexes (Guy et al., 2009, Boubakri et al., 2010, Hawkins et al., 2019). Similarly, *B.subtilis* PcrA has been shown to lessen replisome stalling caused by transcription-related effects (Merrikh et al., 2015). However, it's CTD, which has been shown to directly contact RNAP, is dispensable for this function. In addition, Moreno-del Alamo et al. (2020) concluded that PcrA functions at the interface of transcription, DNA replication, recombination and chromosomal segregation, after genetic analysis.

Investigation of UvrD's direct interaction with RNAP led to the second hypothesis for the PcrA:RNAP interaction, in which an alternative method for TCR was found. In this pathway, RNAP is supposedly forced backwards on the DNA by UvrD in the first step, exposing DNA lesions which are then repaired by NER (Epshtein et al., 2014). However, this claim has been disputed since high-throughput sequencing of oligonucleotides generated during NER identified that the transcribed strand/non-transcribed strand repair ratio actually increased slightly in UvrD mutants whilst it decreased two-fold on the deletion of the classical TC-NER factor, Mfd. Thus, it was concluded that UvrD only acts to unwind damaged ssDNA in NER, and does not have an overall role in TCR (Adebali et al., 2017).

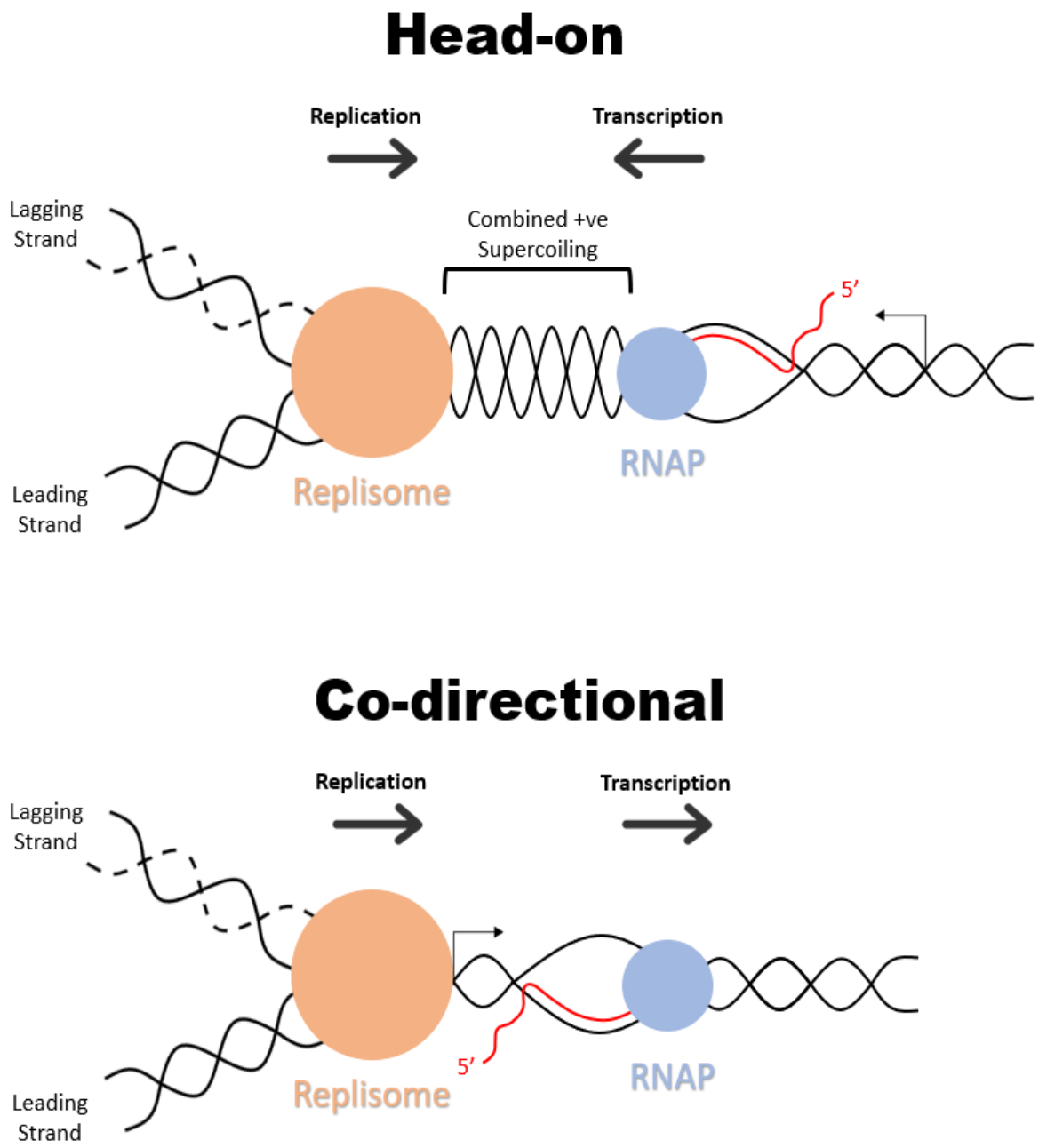

Figure 1-2. Replication-transcription conflicts. Conflicts between the cells replication machinery and RNA polymerase can either result from head-on collisions or by co-directional collisions. Head-on collisions can occur when transcription of a gene present on the lagging strand occurs at the same time and location as DNA replication, resulting in a build-up of positive supercoils ahead of both machineries and stalling of the replication fork. This can lead to genomic instability. Co-directional collisions occur when genes on the leading strand are transcribed at the same time and location as DNA replication, and are often less detrimental to the cell. Newly transcribed RNA is shown in red.

Recent publications studying the structural basis for interaction between UvrD/PcrA and RNAP have started to provide more detailed clues about the function of PcrA in transcription. Crosslinking studies carried out by Nudler's group identified interactions between the 1B and

2B subdomains of UvrD and the β-flap tip and β′-subunit N-terminus of RNAP (Figure 1-3, orange regions) (Epshtein et al., 2014). Later evidence from hydrogen-deuterium exchange coupled to mass spectrometry (HDX-MS) experiments uncovered interactions between the core of PcrA and the regions surrounding the DNA and RNA exit channels of the β and β′ subunits of RNAP, supporting the previous crosslinking data (Figure 1-3, pink, purple and yellow regions) (Urrutia-Irazabal et al., 2021). PcrA's CTD is also necessary for the interaction and contacts the lineage-specific insert (SI1) domain of RNAP's β-subunit (Figure 1-3, green) (Gwynn et al., 2013, Urrutia-Irazabal et al., 2021). In addition to the HDX-MS data, biochemical assays undertaken by Urrutia et al. (2021) revealed that PcrA is capable of unwinding DNA:RNA hybrids and DNA duplexes with 3′-ssDNA overhangs. This work also revealed that expression of dominant-negative PcrA constructs led to the accumulation of DNA:RNA hybrids in vivo. Thus, our primary hypothesis suggests that PcrA is involved in the resolution of the transient three-stranded nucleic acid structures, known as R-loops.

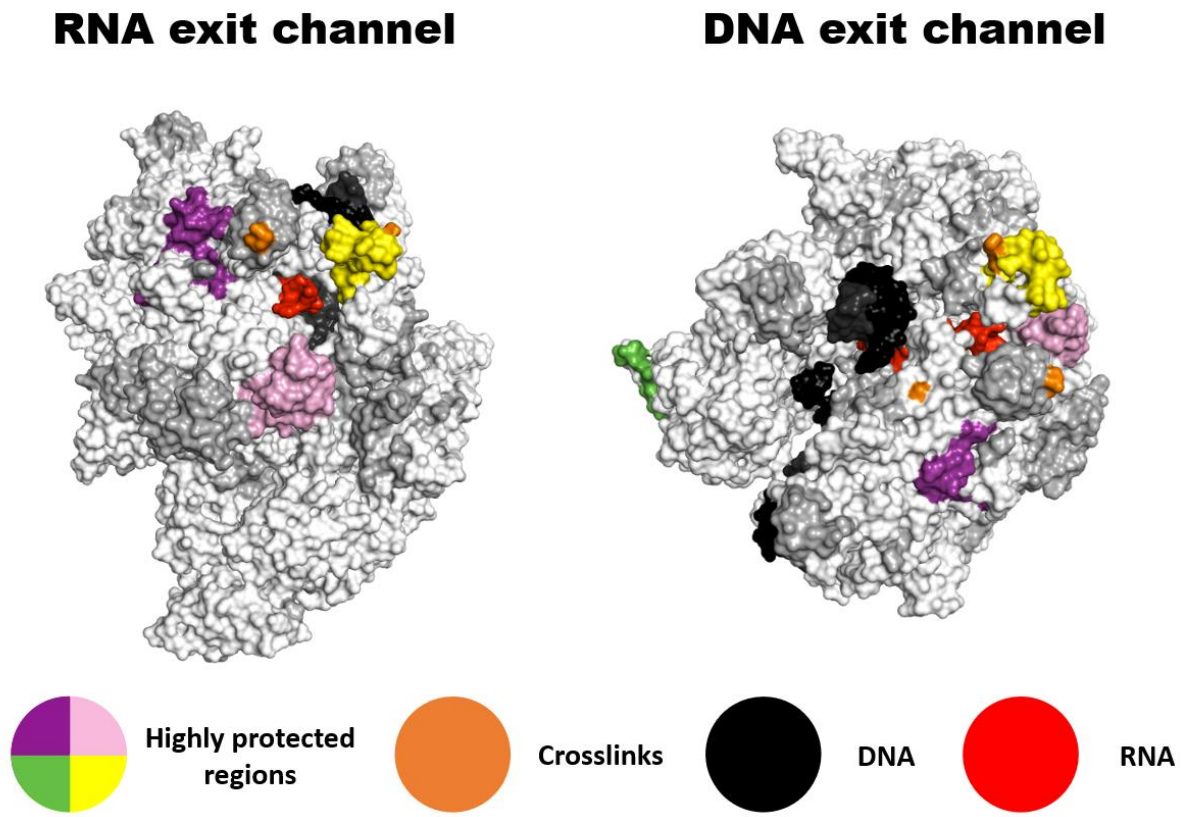

Figure 1-3. PcrA/RNAP interactions. Elongation complex displaying protected regions from interaction with PcrA after HDX-MS (purple, pink, yellow and green) (Urrutia-Irazabal *et al.*, 2021) and crosslinking (orange) (Epshtein *et al.*, 2014) experiments. DNA and RNA are shown in black and red respectively. Protection mainly surrounds the DNA and RNA exit channels of RNA polymerase. The green region contains the helicase interaction motif which interacts with the CTD of PcrA. Data obtained from HDX-MS is broadly consistent with that obtained using crosslinking. Hybrid model created using PBDs: 6WVJ, 6FLQ and DNA/RNA from 6ALF.

Formed as stable intermediates in many DNA transactions, R-loops are DNA:RNA hybrids in which an RNA molecule hybridises to the template strand of a DNA duplex, resulting in local displacement of the opposing DNA strand (Thomas et al., 1976). These structures can be formed both co-transcriptionally and post-transcriptionally, but the most accepted mechanism for their formation involves invasion of the DNA duplex behind the transcription bubble by a nascent RNA transcript (Figure 1-4). In this circumstance, they can act as roadblocks to both replication forks and transcription machineries, as well as precursors of double-strand breaks. It has been reported that they can also result from head-on collisions of transcription and replication, whereby positive supercoiling ahead of both processes meet and combine, leading to stalling of the replication fork and RNAP, and thus subsequent toxicity (García-Muse and Aguilera, 2016, Hamperl et al., 2017, Lang et al., 2017). They are predominantly found at highly-transcribed regions of the genome, and a number of factors promote their formation, including guanine-rich regions, negative supercoiling and DNA lesions (Allison and Wang, 2019).

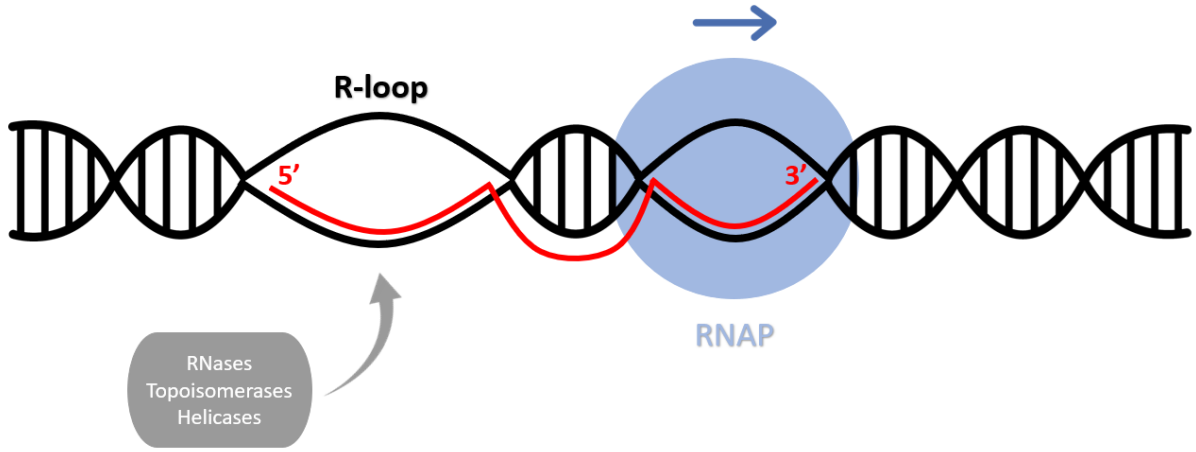

Figure 1-4. Formation and resolution of co-transcriptional R-loops. During transcription, R-loops can be formed when the 5′ end of the nascent RNA molecule (red) invades the DNA duplex (black) behind the transcribing RNA polymerase (blue), annealing to the non-coding strand of DNA and causing dissociation of the coding strand. R-loops can lead to genomic instability if they accumulate and remain unresolved. Cells can resolve R-loops using a number of methods, including RNases, topoisomerases and helicases. The direction of transcription is indicated by a blue arrow.

R-loops have a number of physiological roles, including gene regulation, immunoglobulin class-switch recombination, promotion of telomere repair and mitochondrial DNA replication (Lieber et al., 2003, Pohjoismäki et al., 2010, Arora et al., 2014, Fernandes et al., 2021). However, accumulation of aberrant R-loops poses a threat to genome integrity by instigating mutations, recombination and chromosomal rearrangements (Aguilera and García-Muse,

2012). Therefore, they are of particular interest to the biomedical community since genomic instability is a hallmark of cancer, and high levels of R-loops have been linked to neurological disorders (Halazonetis et al., 2010, Richard and Manley, 2017). Due to their stability, they require active removal. Thus, cells have developed several ways to prevent their formation, but also mechanisms to resolve them if these fail. These include RNases, topoisomerases and helicases. UvrD was first implicated in the unwinding of DNA:RNA hybrids by Matson (1989), and more recent studies have linked the helicase to R-loop homeostasis, as well as the potential resolution of R-loops by PcrA, as mentioned above (Wolak et al., 2020, Urrutia-Irazabal et al., 2021). Furthermore, the human SF1 helicase, Senataxin, has also been linked to R-loop resolution (Skourti-Stathaki et al., 2011, Cohen et al., 2018, Hasanova et al., 2022). Mutations in the *SETX* gene can lead to one of two neurodegenerative conditions: a form of amyotrophic lateral sclerosis (ALS4) or ataxia with oculomotor apraxia type 2 (AOA2). ALS4 is characterised by muscle weakness, atrophy and hyperreflexia whilst AOA2 affects muscle control, eye movement and can even lead to cognitive impairment. This again highlights the relevance of understanding the role helicases play in R-loop homeostasis. However, the mechanism by which helicases act to remove R-loops is poorly understood.

Given the well-established 3′-5′ translocation polarity of PcrA on ssDNA and the location of contacts found between PcrA and RNAP, two possible mechanisms for the removal of R-loops satisfy current biochemical and structural data. The first hypothesis evolved from observations made by Epshtein et al. (2014), in which UvrD was reported to pull RNAP backwards and hence, it was termed the 'backtracking' model. As the name suggests, PcrA indirectly unwinds R-loops by translocating in the opposite direction to RNAP on the non-template strand, leading to the backtracking of the polymerase (Figure 1-5). In this circumstance, the lid domain of the RNAP forces the strands of the R-loop apart. The second, termed the 'co-directional' model, was proposed based on findings from Urrutia-Irazabal et al. (2021), particularly those showing the location of the contacts between PcrA and RNAP. In this model, PcrA moves in the same direction as RNAP behind the transcription bubble. As a result, co-transcriptional R-loops could be unwound almost instantly as they are formed (Figure 1-5). In both cases, PcrA is proposed to initially be recruited to RNAP, and consequently the R-loops, via its CTD. We favour the co-directional model since current HDX-MS data positions the front of the helicase (the 2A domain) next to the RNA/DNA exit channels of RNAP, suggesting that they move in the same direction (Urrutia-Irazabal et al., 2021). However, evidence for the backtracking of RNAP by PcrA from multiple sources also provides support for the first model (Epshtein et al., 2014, Sanders et al., 2017). Clearly, a high-resolution structure of the PcrA:RNAP complex

would help to differentiate between the models and enhance our understanding of this process.

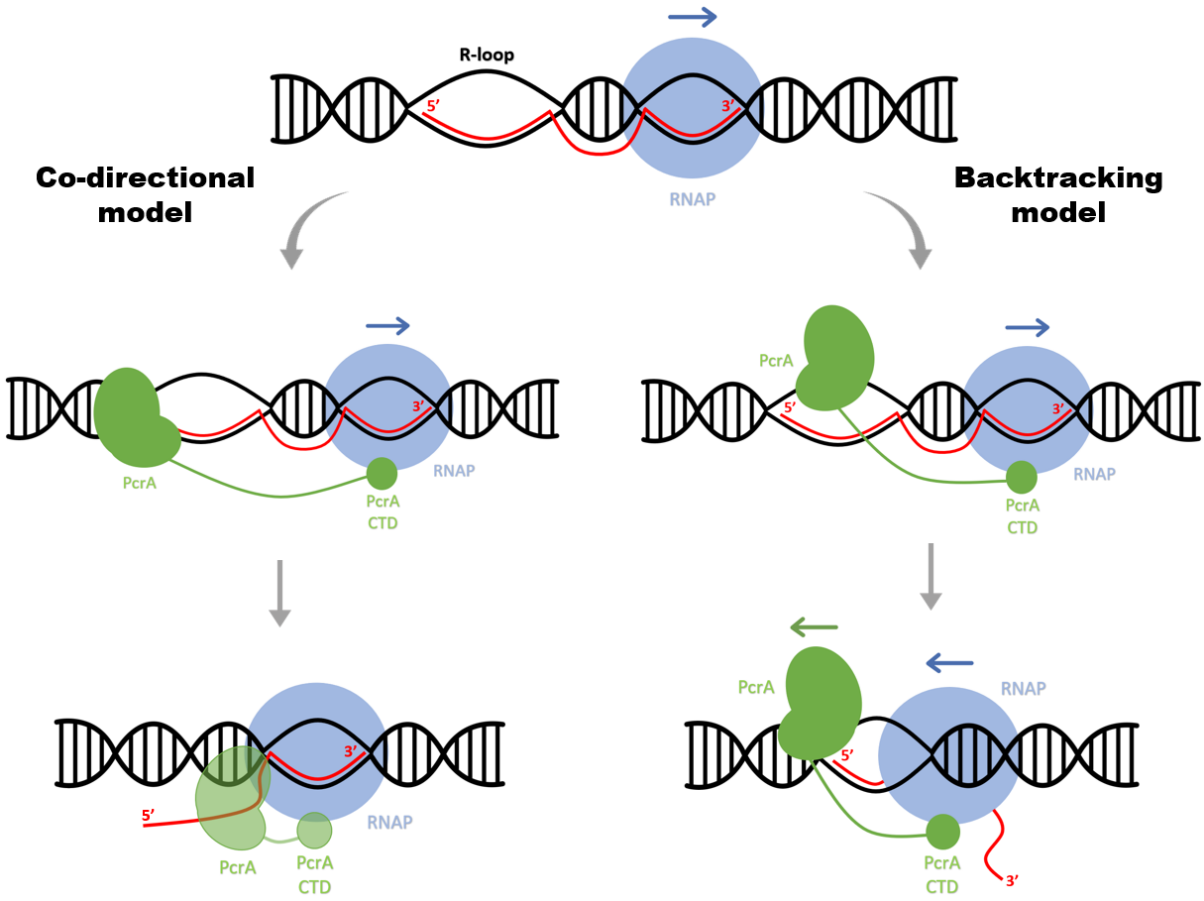

Figure 1-5. Speculative mechanisms of R-loop resolution by PcrA. Two methods for the resolution of co-transcriptional R-loops have been proposed based on biochemical and structural data. The co-directional model (left) suggests that PcrA (green) is recruited to RNA polymerase (blue) via its CTD and translocates in the same direction, unwinding R-loops as they form. The backtracking model (right) suggests that R-loops are indirectly unwound by the backtracking of RNA polymerase, which is initiated by PcrA after recruitment via its CTD.

It is important to highlight, that the three hypotheses generated to explain the function of the PcrA:RNAP interaction are not necessarily mutually exclusive, however, but actually considerably intertwined. It is possible that the effects observed when PcrA function is inhibited are not due to a lack of direct unwinding of R-loops by PcrA, but rather an increase in replication-transcription conflicts resulting from the absence of PcrA and thus, increased R-loop formation. However, it is also possible that the direct resolution of R-loops by PcrA then in turn prevents replication-transcription conflicts. Furthermore, the proposed involvement of PcrA in TCR could also prevent the formation of R-loops, rather than PcrA directly unwinding them, by promoting the repair of DNA lesions that could encourage R-loop formation. PcrA is

predicted to do this by removal of the stalled polymerase from the lesion, which also avoids further conflict and conflict-associated R-loops.

1.2 PcrA CTD interactions with partner proteins containing a helicase interaction motif

Within the SI1 domain of RNAP, which forms interactions with the CTD of PcrA, a nine-residue putative helicase interaction motif (HIM) was identified ("VDPETGEIL" in *B.subtilis*) (Urrutia-Irazabal et al., 2021). The motif adopts a β-hairpin fold in the *E.coli* RNAP structure, with the "TGE" triad occupying the tip of the loop between the two β-strands. The glutamate residue of the triad shows complete conservation across RpoB sequences and mutagenesis experiments revealed its importance in the interaction with PcrA, whereby pulldown of RpoB by PcrA was dramatically reduced in mutants (Urrutia-Irazabal et al., 2021). Remarkably, variations of this conserved HIM were identified in four known interactors of PcrA (RpoB, UvrB, YxaL and YwhK) as well as three new predicted partners (QueA, RplX and YtzB) using the PROSITE database (Urrutia-Irazabal et al., 2021). In the cases where experimental data is available, the regions showing interaction with PcrA are consistent with the location of the HIM in the 3D structure of the proteins (Noirot-Gros et al., 2002b, Manelyte et al., 2009). However, since the identification of the HIM, there has been no further study concerning these interactions.

In recent years, the field of structural biology has been revolutionised with the emergence of exceptionally accurate in silico protein structure prediction programmes, including DeepMind's prodigy, AlphaFold (Jumper et al., 2021). This deep learning technology uses co-evolutionary data from multiple sequence alignments (MSAs) and solved protein structures to predict protein structures from their primary sequences. AlphaFold and AlphaFold2 dominated the Critical Assessment of Structure Prediction (CASP) competition in 2018 and 2020 respectively, and programmes showing the greatest success in the most recent edition have incorporated AlphaFold into their methods (Kryshtafovych et al., 2019, Kryshtafovych et al., 2021). Recently, an adaptation of AlphaFold, in which protein-protein interactions (PPIs) and multi-chain protein complexes could be predicted, has been developed, known as AlphaFold-Multimer. Thus, this provides a tool for us to not only study and predict the structures of the HIMs of PcrA's partner proteins, but to also model interaction between PcrA and these partners.

1.3 Thesis outline

To unveil the precise interactions formed between RNAP and PcrA, and further elucidate the structure to function relationship of the complex, a high-resolution structure is required. The development of AI structural prediction tools also provides us with a means of exploring predicted interactions between PcrA and RNAP, in addition to PcrA and other known and putative partner proteins containing the HIM. In Chapter 3, we assess three broad methods for PcrA:RNAP complex formation in an attempt to establish an effective and reproducible method to generate a complex for downstream imaging applications. These methods include individually combining RNAP and PcrA in the presence of an artificial DNA:RNA scaffold, co-expressing PcrA and RNAP in *E.coli* and mixing cell lysates from separate expressions of RNAP and PcrA. We found that combining native *B.subtilis* RNAP and recombinant *B.subtilis* PcrA individually with a DNA:RNA scaffold was the only method in which the proteins convincingly co-eluted from a size exclusion column, albeit at low concentrations and non-stoichiometrically. Analysis of complex formation using electrophoretic mobility shift assays showed that both native and recombinant *B.subtilis* RNAP interacted with PcrA when combined in the presence of scaffolds formed via two different assembly methods. However, in the case of native RNAP, multiple shifted bands were generated which require identification, whilst in the case of recombinant RNAP, material largely remained in the wells of the gel. In Chapter 4, we explore PcrA's interactions with partner proteins containing a HIM using the protein structure prediction software, AlphaFold-Multimer. We present data showing high confidence interactions between PcrA's CTD and both the HIM of the SI1 domain of RNAP's β-subunit as well as the HIM of YxaL. We also predicted a surface interaction between C-terminal regions of PcrA's 2B subdomain and the β-subunit of RNAP which would be consistent with data obtained by Urrutia-Irazabal et al. (2021), but this prediction had low confidence and would require experimental validation.

Chapter 2

Materials and Methods

2.1 Subcloning

To co-express RNAP and PcrA, the full length PcrA gene was first excised from the pET22b_BSu_PcrA vector and inserted into a pRSFDuet-1 vector. pRSFDuet-1 was chosen here since it possesses a different replication origin and resistance marker to the plasmid expressing core RNAP, pNG1256 (Newing et al., 2020). 10 ng of pRSFDuet-1 and pET22b_BSu_PcrA were digested with NdeI and XhoI (NEB) at 37 °C for 2 hours. Reaction volumes varied, but enzyme concentration was kept below 5 % of the total volume. Digested products were run on a 0.8 % Agarose gel and purified using the QiAquick Gel Extraction Kit (Qiagen, Cat. No. 28706). pRSFDuet-1 underwent dephosphorylation using Antarctic phosphatase (NEB) according to the manufacturer's instructions. The insert and recipient vector were ligated using T4 DNA ligase (NEB) according to the manufacturer's instructions at a 1:1 molar ratio. Reactions were incubated at room temperature (RT) for 2 hours. Cloning success was validated by sequencing.

2.2 Protein expression and purification

2.2.1 PcrA

Rosetta2(DE3) cells were transformed with pET22b_BSu_PcrA and plated on LBA plates containing Ampicillin (50 µg/ml). Single colonies were used to inoculate 5 ml LB containing Ampicillin and incubated overnight at 37 °C and 250 rpm. The following day, 2 ml of the overnight culture was used to inoculate 1 L LB media containing Ampicillin, and cells were grown to an OD_{600} of 0.5-0.6. Expression was induced by the addition of IPTG (1 mM) before cells were incubated at 18 °C, 250 rpm overnight. Cells were harvested by centrifugation at 4500 rpm and 4 °C for 15 minutes, and resulting cell pellets were resuspended in resuspension buffer (50 mM Tris-Cl pH 7.5, 300 mM NaCl, 0.1 mM EDTA, 0.1 mM DTT, 10 % sucrose (w/v)). Resuspended pellets were stored at -80 °C. 1 ml of resuspended cells was assessed for protein expression using enzymatic lysis with lysozyme.

Resuspended cells were supplemented with 100 µM PMSF prior to lysis by sonication. The lysate was centrifuged at 20,000 rpm and 4 °C for 30 minutes. Proteins were precipitated by adding ammonium sulfate to 50 % saturation before centrifugation at 20,000 rpm and 4 °C for 30 minutes. The pellet was resuspended in Buffer A (50 mM Tris-Cl pH 7.5, 2 mM EDTA, 1 mM DTT) until a conductivity equivalent to Buffer A + 200 mM NaCl. The resuspension was passed through a 0.45 µM filter before loading onto a 5 ml Hi-Trap Heparin-Sepharose column (Cytiva) equilibrated with Buffer A + 100 mM NaCl. Protein was eluted over a gradient of Buffer A + 200-1000 mM NaCl. Fractions containing PcrA were dialysed against Buffer A + 200 mM NaCl overnight before passing through a 0.2 µM filter the following day. The filtered solution was loaded onto a MonoQ column (Cytiva) preequilibrated in Buffer A + 200 mM NaCl and

protein was eluted over a gradient of Buffer A + 200-1000 mM NaCl. Fractions containing PcrA were pooled and loaded onto a Superdex 200 16/60 column (Cytiva) equilibrated with Buffer A + 200 mM NaCl. Protein concentration was quantified using a DS-11 series spectrophotometer (DeNovix). Glycerol was added to purified PcrA to a final concentration of 10 % (v/v) before snap freezing and storage at -80 °C.

2.2.2 Recombinant RNAP (from pNG1256)

Rosetta2(DE3) cells were transformed with pNG1256 and plated on LBA plates containing Ampicillin (50 µg/ml). Single colonies were used to inoculate 5 ml/10 ml LB containing Ampicillin and incubated overnight at 37 °C and 250 rpm. The following day, 2 ml of the overnight culture was used to inoculate 1 L LB media containing Ampicillin, and cells were grown to an OD_{600} of 0.5-0.6. Expression was induced by the addition of IPTG (1 mM) and cells were incubated at 27 °C, 250 rpm overnight. Cells were harvested as in 2.2.1, and the resulting pellets were resuspended in HisA buffer (20 mM KH_2PO_4 pH 7.8, 500 mM NaCl, 20 mM Imidazole), before storage at -80 °C. 1 ml of resuspended cells was assessed for protein expression using enzymatic lysis with lysozyme. His SpinTrap TALON (Cytiva, Cat. No. 29000593) columns were used to confirm the presence of its subunits and assess binding of the β′-subunit C-terminal His-Tag (rpoC-His9).

Resuspended cells were supplemented with an EDTA-free protease inhibitor (Roche) and DNaseI (0.1 mg/ml) before lysis by sonication and centrifugation at 20,000 rpm and 4 °C for 30 minutes. The resulting supernatant was passed through a 0.45 µM filter before loading onto a 5 ml HisTrap column (Cytiva) preequilibrated in HisA buffer (20 mM KH_2PO_4 pH 7.8, 500 mM NaCl, 20 mM Imidazole). The column was washed with 4 % HisB (HisA + 500 mM Imidazole) before RNAP was eluted using 50 % HisB. The eluate was dialysed against QA buffer (20 mM Tris-Cl pH 7.8, 150 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT) overnight before loading onto a 1 ml MonoQ equilibrated in QA buffer the following morning. Protein was eluted over a gradient of 0-50 % QA + 1M NaCl and the fractions containing RNAP were pooled. RNAP was concentrated at this point using an Amicon Ultra-15 Centrifugal filter with a 3 kDa MWCO and either snap-frozen and stored at -80 °C or loaded onto a Superose 6 Increase 10/300 Gl column (Cytiva). Eluted protein was snap frozen and stored at -80 °C. Before freezing, protein concentration was quantified using a DS-11 series spectrophotometer (DeNovix).

In some cases, an extra wash step with HisA + 1.5 M NaCl was included during the HisTrap column step of the purification to remove bound nucleic acids. Following this, the NaCl concentration of HisB buffer and QA buffer was increased to 275 mM to avoid protein precipitation.

2.2.3 Native RNAP (from MH5636)

B.subtilis RNAP was also expressed and purified from MH5636, a modified version of the *B.subtilis* 168 strain (Qi and Hulett, 1998). MH5636 cells were plated on LBA plates containing Chloramphenicol (25 µg/ml). Single colonies were used to inoculate 5/10 ml LB containing Chloramphenicol and cells were grown to an OD_{600} of 1.2, before being harvested as in 2.2.1. Cell pellets were resuspended in Buffer R (20 mM Tris-Cl pH 7.5, 300 mM NaCl, 5% glycerol (v/v), 0.1 mM DTT, 10mM Imidazole) and stored at -80°C. 1 ml of resuspended cells was assessed for protein expression using enzymatic lysis with lysozyme. His SpinTrap TALON (Cytiva, Cat. No. 29000593) columns were used to confirm the presence of its subunits and assess binding of the β′-subunit C-terminal His-Tag (rpoC-His10).

Resuspended cells were supplemented with an EDTA-free protease inhibitor (Roche) and lysozyme (0.5 mg/ml) before lysis by sonication and centrifugation at 20,000 rpm and 4 °C for 30 minutes. The resulting supernatant was passed through a 0.45 µM filter before loading onto a 5 ml HisTrap column (Cytiva) preequilibrated in Buffer R. Protein was eluted over a gradient of Buffer R + 10 mM-20 mM Imidazole. Fractions containing protein were dialysed against Buffer Q (10 mM Tris pH 7.5, 100 mM NaCl, 5 % glycerol (v/v), 0.1 mM DTT, 0.1 mM EDTA) overnight, before loading onto a 1 ml MonoQ equilibrated in Buffer Q the following morning. Protein was eluted over a gradient of Buffer Q + 100-600 mM NaCl. Protein concentration was quantified using a DS-11 series spectrophotometer (DeNovix), before being snap frozen and stored at -80 °C.

The yields of protein achieved in this work are presented in Table 2-1.

2.2.4 Co-expressed PcrA and RNAP

Rosetta2(DE3) cells were sequentially transformed, first with pNG1256 followed by pRSFDuet-1 containing the PcrA insert. Cells were plated on LBA plates containing Ampicillin (50 µg/ml) and Kanamycin (50 µg/ml). Single colonies were used to inoculate 5 ml/10 ml LB containing Ampicillin and Kanamycin and incubated overnight at 37 °C and 250 rpm. The following day, 2 ml of the overnight culture was used to inoculate 1 L LB media containing Ampicillin and Kanamycin. Cells were grown to an OD_{600} of 0.5-0.6 before expression was induced by the addition of IPTG (1 mM). Cells were incubated at 37°C, 250 rpm for 3 hours. Cells were harvested as in 2.2.1 and the resulting pellets were resuspended in HisA buffer and stored at -80 °C. 1 ml of resuspended cells was assessed for protein expression using enzymatic lysis with lysozyme and Ni-NTA agarose (Qiagen, Cat. No. 30410) was used to confirm PcrA pulldown with RNAP.

Proteins resulting from co-expression were purified using the same method as that of RNAP from pNG1256 (section 2.2.2), but Ni-NTA agarose was used for the affinity purification step

instead of the HisTrap column. In this case, the resin was centrifuged at 4000 x g for 2 minutes before removal of EtOH and washing with His A buffer. The supernatant was added to the resin and incubated at 4 °C with gentle rocking for 1 hour. The resin mix was then centrifuged at 3000 x g and 4 °C for 1 minute, before removal of the supernatant. The beads were washed again with HisA a minimum of three times before transfer to a gravity flow column and elution with HisB.

Table 2-1. Achieved protein yields.

Protein	Molecular Weight (kDa)	Extinction Coefficient ($M^{-1}cm^{-1}$)	Yield (mg/L)
PcrA	83.44	66,380	1.9643
***B.subtilis* RNAP (from pNG1256)**	344.20	187,230	1.4363
***B.subtilis* RNAP (from MH5636)**	382.09	241,425	0.9803
Washed *B.subtilis* RNAP (from pNG1256)	344.20	187,230	1.9605

2.3 PcrA:RNAP complex formation

2.3.1 Using individual components

Transcription elongation complex (TEC) formation was adapted from previous work (Sidorenkov et al., 1998). RNA was first incubated with scaffold template strand (TS) at 45 °C for 5 minutes in Buffer T (50 mM Tris-Cl pH 7.9, 150 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT) before cooling to 20 °C at a rate of 1 °C/min. RNAP was added and incubated at RT for 10 minutes before the addition of scaffold non-template strand (NTS) and further incubation at 37 °C for 10 minutes. The oligonucleotides used in this work are given in Table 2-2. RNAP was added to the highest concentration possible allowing for the molar ratios 1(RNAP):2(TS):10(NTS):4(RNA) in the smallest injection volume possible, up to 400 µl. PcrA was then added at a molar ratio of 1(RNAP):1.3(PcrA). See Figure 3-8 (RHS flow-diagram) for a schematic representation of complex formation.

PcrA:RNAP ‘complexes’ were run on a Superose 6 Increase 10/300 GI (Cytiva) equilibrated in Buffer T. Fractions containing protein were assessed using SDS-PAGE.

Table 2-2. Oligonucleotides used for scaffold formation

Strand	Sequence (5′-3′)	Modification
RNA	AUCGAGAGG	-
Template strand	TTCGCCGTGTCCCTCTCGATGGCTGTAAGT	5′ Cy5
Non-template strand	ACTTACAGCCATCGAGAGGGACACGGCGAA	-

2.3.2 Combining cell lysates

Resuspended cell pellets from individual expressions of PcrA and recombinant RNAP were combined and simultaneously lysed by sonication. The lysates were incubated at 4 °C on a rocker prior to purification using the same methods as section 2.2.2.

2.4 Electrophoretic mobility shift assay (EMSA)

Scaffold and complex formation were determined using TBE-PAGE native gels (12 % and 6 % respectively). Scaffolds were either formed as above (2.3.1), but omitting the addition of RNAP and PcrA, or by combining all nucleic acid components in equimolar quantities and heating to 95 °C for 10 minutes, before cooling to RT overnight. Protein complexes were either formed by incubation of RNAP with the pre-formed scaffolds for 10 minutes at RT, followed by the addition of PcrA and incubation at RT for 10 minutes, or by following the methods above (2.3.1). See Figure 3-8 for schematic representation. Gels were run at 150 V and 4 °C for 20 minutes for assessment of scaffold formation, or 35 minutes for protein complex formation. Gels were ran using the Mini-PROTEAN Tetra Cell system (Bio-Rad) and imaged using a Typhoon FLA 9500 (GE Healthcare).

2.5 AlphaFold-Multimer protein complex prediction

Interactions between PcrA and partner proteins were predicted using Alphafold2-Multimer via ColabFold (Jumper et al., 2021, Mirdita et al., 2022, Varadi et al., 2022). Unless stated, default parameters were used for all query submissions. The resulting protein structures were visualised using PyMOL (Schrödinger, Version 2.5) or ChimeraX (Goddard et al., 2018, Pettersen et al., 2021). Any interactions at the protein-protein interfaces predicted by AlphaFold2-Multimer were analysed using the ‘Protein interfaces, surfaces and assemblies' service (PISA) at the European Bioinformatics Institute (EBI) (Krissinel and Henrick, 2007) and ChimeraX.

Chapter 3

Exploring methods for formation of the PcrA:RNAP complex

3.1 Introduction

Since the first evidence of an interaction between RNAP and UvrD/PcrA from Runyon et al. (1993), the interaction and its implications for PcrA in transcription and related conflicts has become a point of interest within the biochemical community. As mentioned previously, several groups have studied the biochemistry of the interaction in detail, in which both UvrD and PcrA have been implicated in the resolution of replication-transcription conflicts. In addition, recent studies have started to shed light on the structural basis for this interaction. Crosslinking experiments carried out by Epshtein et al. (2014) identified contacts between the domains forming the core of UvrD and the β-subunits of RNAP. Data obtained several years later using HDX-MS was found to be broadly consistent with that obtained by crosslinking, whereby contacts were identified between PcrA's core and the regions surrounding the DNA and RNA exit channels of RNAP (Urrutia-Irazabal et al., 2021). Furthermore, interactions between the CTD of PcrA and RNAP have been identified by multiple studies (Gwynn et al., 2013, Urrutia-Irazabal et al., 2021). The specifics of these interactions are discussed in greater detail in Chapter 1.

The discovery of PcrA's recruitment close to the DNA and RNA exit channels of RNAP in combination with findings that UvrD is capable of unwinding DNA:RNA hybrids (Matson, 1989) led Urrutia-Irazabal et al. (2021) to test PcrA's ability to unwind R-loops. They found that PcrA was not only able to unwind DNA:RNA hybrids in vitro, but that cellular R-loop levels increased when PcrA activity was inhibited. This gave rise to the current hypothesis that PcrA acts as a surveillance helicase by resolving co-transcriptional R-loops in order to alleviate their effects on other DNA transactions. However, in order to gain the precise details of the PcrA:RNAP interaction and to assist in determining the mechanism behind R-loop resolution by PcrA, a high-resolution structure is required.

To gain a high-resolution structure, a stable PcrA:RNAP complex must first be formed. Urrutia-Irazabal et al. (2021) discovered that this required the presence of a DNA:RNA scaffold, which was formed by annealing DNA and RNA strands in a stepwise manner in combination with native *B.subtilis* RNAP to form an artificial transcription elongation complex (TEC), before then interacting with PcrA (Figure 3-8). Recently, a paper studying interactions between the transcription recycling factor, HelD, and RNAP described a method of synthesising a transcription elongation complex (TEC) comprising only the core subunits (α2, β, β′ and ω) of *B.subtilis* RNAP in *E.coli* (Newing et al., 2020). Urrutia-Irazabal (2021) initially found that PcrA had a higher affinity for RNAP in the presence of σ_A after assessment by microscale thermophoresis. However, this result was later dismissed when no direct interaction was found between PcrA and σ_A or PcrA and RNAP:σ_A. As a result, formation of a PcrA:RNAP complex

comprising only the core subunits of RNAP offers a potential advantage over native RNAP, due to the likely production of more homogenous material, which could alleviate issues in downstream structural characterisations. Furthermore, since the RNAP synthesised by Newing et al. (2020) apparently purifies pre-bound to a DNA:RNA hybrid, and Urrutia-Irazabal et al. (2021) proposed that PcrA favours the conformation adopted by the TEC rather than directly engaging with the nucleic acids, it was possible that a PcrA:RNAP complex could be formed without the need for an additional TEC formation step. Thus, we explore the use of both recombinant and native RNAP here in an attempt to form a stable PcrA:RNAP complex from its individual components.

Since PcrA/UvrD is a known interactor of RNAP, and they have been shown to co-purify, it is also plausible that overexpression of the two may encourage formation of a complex in vivo. Assuming that the recombinant material mentioned above is also synthesised as a TEC, this may provide the preferred substrate for PcrA, allowing for us to obtain a stable complex in a single purification. Since the exact composition and structure of the nucleic acids bound to RNAP during the recruitment of PcrA is unknown, co-expression of the proteins may also identify a favoured substrate. Thus, we test a novel approach here for formation of the PcrA:RNAP complex by co-expression in *E.coli*.

Using similar principles to those for co-expression, we also attempt to form a complex by combining cell lysates of PcrA and the recombinant RNAP. Since the recombinant material reportedly forms a TEC, we propose that PcrA may therefore form interactions with the polymerase by mixing lysates from their respective expressions. Again, this may have the advantage of selecting and identifying the favoured substrate of PcrA.

The strategies for PcrA:RNAP complex formation are summarised in Figure 3-1.

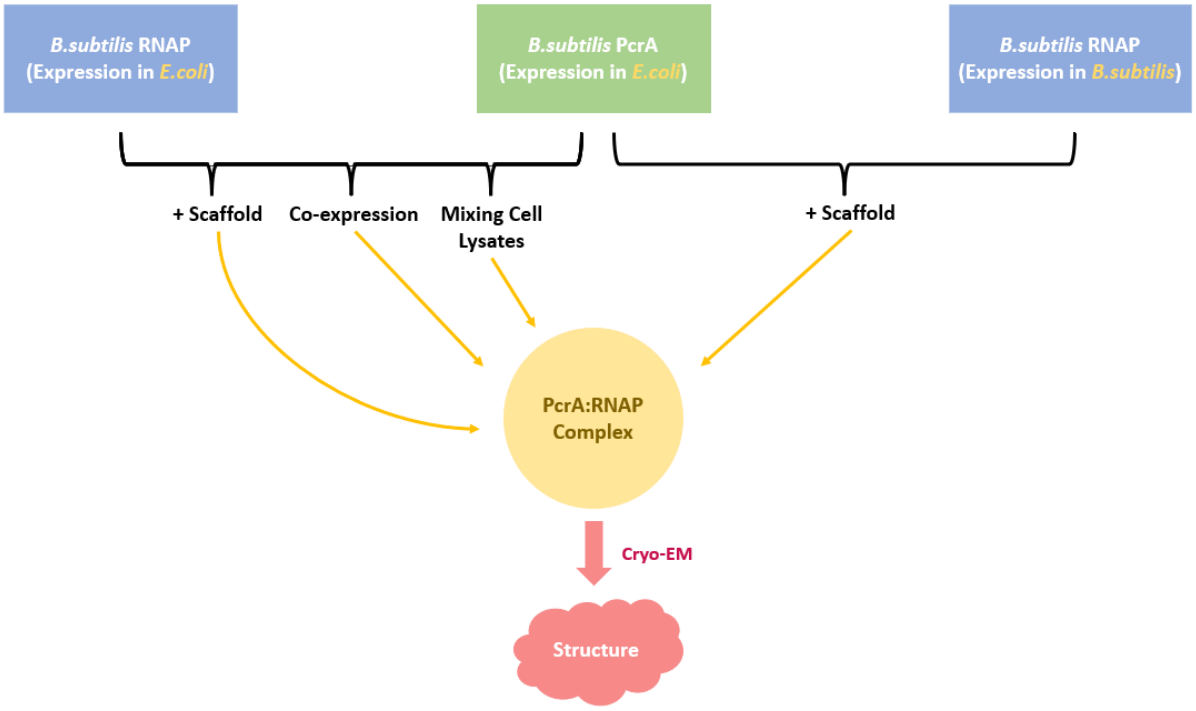

Figure 3-1. Methods for formation of a stable PcrA:RNAP complex for downstream cryo-EM. Flowchart illustrating the methods explored for formation of a stable PcrA:RNAP complex for downstream imaging applications. These methods include combining both recombinant *B.subtilis* RNAP (expressed in *E.coli*) and native *B.subtilis* RNAP (expressed in *B.subtilis)* with recombinant *B.subtilis* PcrA (expressed in *E.coli)* in the presence of a DNA:RNA scaffold, co-expression of *B.subtilis* PcrA and RNAP in *E.coli* and combining cell lysates from separate expressions of recombinant *B.subtilis* RNAP and recombinant *B.subtilis* PcrA.

3.1.1 Aims of the chapter

This chapter explores 3 methods for formation of a stable PcrA:RNAP complex for imaging applications. The aims are to:

1. Optimise methods for production of all proteins involved
2. Identify successful methods for PcrA:RNAP complex formation by:
   a) Combining recombinant *B.subtilis* PcrA and both recombinant and native *B.subtilis* RNA polymerase in the presence of an added DNA:RNA scaffold
   b) Co-expressing recombinant *B.subtilis* PcrA and RNAP
   c) Mixing cell lysates from individual protein expressions of PcrA and recombinant RNAP
3. Validate and assess the PcrA:RNAP interaction using EMSAs

3.2 Results

3.2.1 PcrA:RNAP complex formation from individual components

3.2.1.1 Expression of PcrA and RNA polymerase

To form the PcrA:RNAP complex from its individual components, untagged *B.subtilis* PcrA and *B.subtilis* RNAP with a His-Tag at the C-terminus of its β′-subunit were first expressed individually in Rosetta2(DE3) and BL21(DE3) cells, respectively. In later stages, RNAP was also expressed in a modified *B.subtilis* strain, MH5636 (Qi and Hulett, 1998). The methods used for untagged PcrA expression and expression of native His-tagged RNAP (in MH5636 cells) have previously been published by our lab and thus, required no optimisation (Gwynn et al., 2013). Figure 3-2 shows the resulting expressions of PcrA and native RNAP, whereby intense bands corresponding to PcrA and all subunits of RNAP are present within their respective soluble fractions. A small-scale His-Tag purification of RNAP was also done to assess its formation from all of its subunits at this stage. Figure 3-2B shows the presence of the β, β′, α, δ and ω subunits, which should run at approximately 136 kDa, 134 kDa, 35 kDa, 20 kDa and 8 kDa respectively. However, all subunits ran apparently higher than their respective molecular weights, but this was also observed by Urrutia-Irazabal (2021). Despite this, their presence indicates successful formation of RNAP.

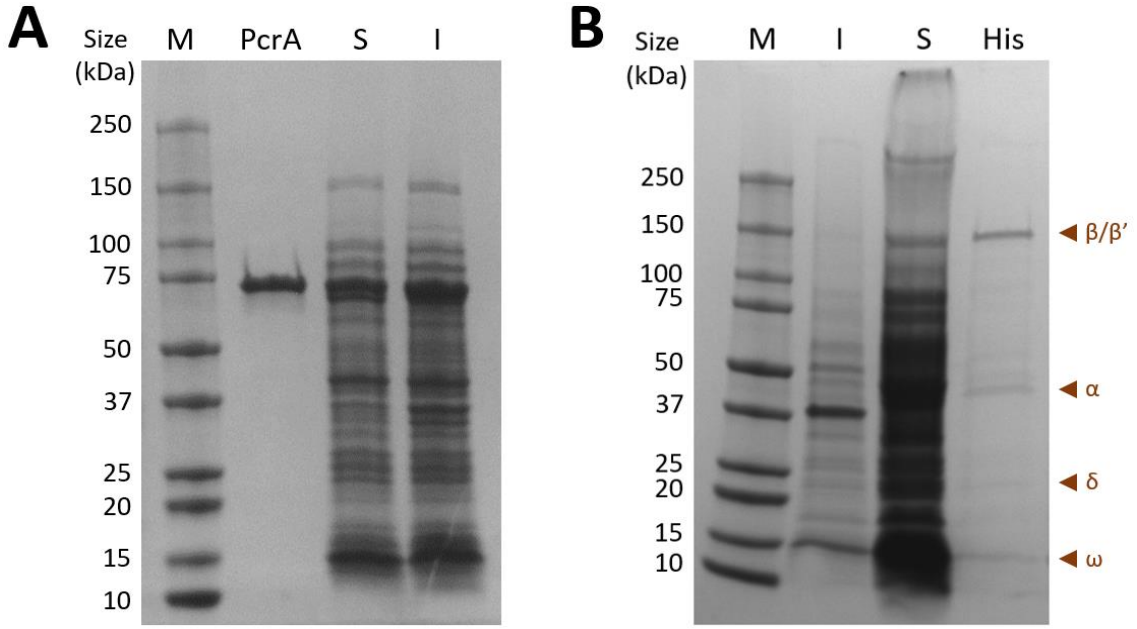

Figure 3-2. Expression of *B.subtilis* PcrA in Rosetta2(DE3) and *B.subtilis* RNAP in MH5636 cells. (A) SDS-PAGE gel showing soluble and insoluble expression of PcrA. A pure sample of PcrA was run to confirm expression of PcrA at the correct molecular weight. (B) SDS-PAGE gel showing soluble and insoluble expression of RNAP from MH5636. Formation of RNAP from its individual subunits was confirmed using His SpinTrap column pulldowns (His). Presence of individual RNAP subunits on the gel are indicated by dark orange arrowheads. In both gels, M represents the 10-250 kDa protein sizing ladder, S represents the soluble fraction and I represents the insoluble fraction.

In contrast to PcrA and the native RNAP, expression of recombinant His-tagged *B.subtilis* RNAP from the plasmid pNG1256 required optimisation since the methods employed directly from Newing et al. (2020) yielded low protein expression in our hands. Four *E.coli* expression strains were selected for expression trials at 37 °C for 3 hours based on their use in a previous dual plasmid expression system for RNAP (Yang and Lewis, 2008). As shown in Figure 3-3, soluble expression of the α-subunits and the ω-subunit was greatest in BL21(DE3) and Rosetta2(DE3) cells, but only soluble expression in Rosetta2(DE3) produced a noticeably intense band corresponding to the β and β′ subunits.

Expression of RNAP was repeated in the BL21(DE3) and Rosetta2(DE3) strains under different conditions: 37 °C for 3 hours, 27 °C for 3 hours and 18 °C overnight. It can be clearly seen in Figure 3-4 that soluble expression was greatest in the Rosetta2(DE3) cells at 27 °C and hence, this condition was used for subsequent RNAP expression. Small-scale His-Tag purifications were also done here to assess the presence of RNAP subunits, with β, β′, α and ω present at the same molecular weights as the native RNAP. The δ subunit is not present in RNAP from pNG1256 since it only encodes the core subunits (β, β′, α_2 and ω).

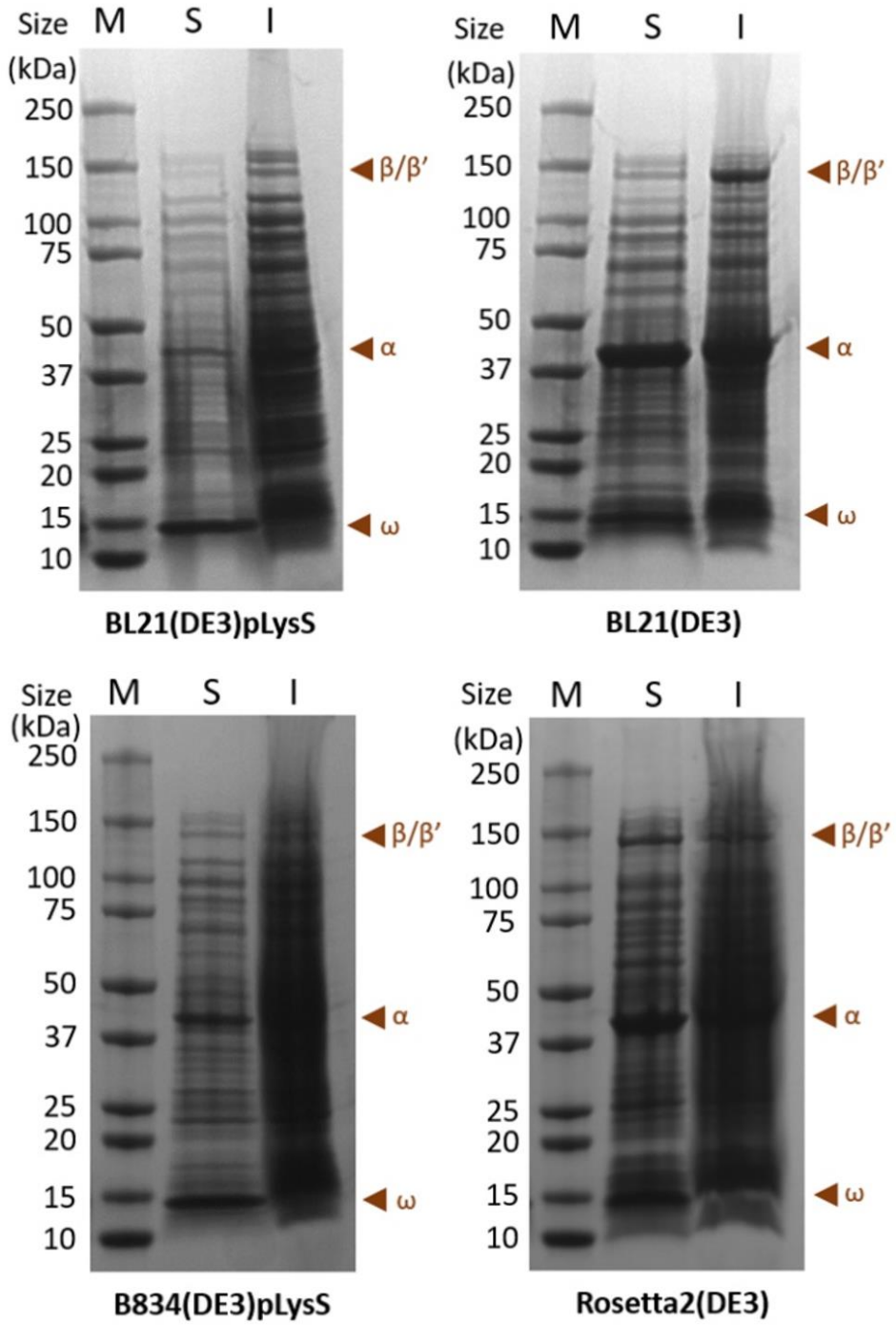

Figure 3-3. *B.subtilis* RNAP expression in different *E.coli* strains. SDS-PAGE page gels displaying soluble and insoluble expression of all core RNAP subunits in BL21(DE3)pLysS, BL21(DE3), B834(DE3)pLysS and Rosetta2(DE3) cells. *E.coli* strains are shown below their respective gels. Expression was induced with IPTG (1mM) at an OD_{600} of 0.5-0.6 and cells incubated at 37 °C for 3 hours. BL21(DE3) and Rosetta2(DE3) yielded the greatest soluble expression of the α and ω subunits and Rosetta2(DE3) yielded the greatest expression of the β and β' subunits (Dark orange arrowheads with symbols). In all gels, M represents the 10-250 kDa protein sizing ladder, S represents the soluble fraction and I represents the insoluble fraction.

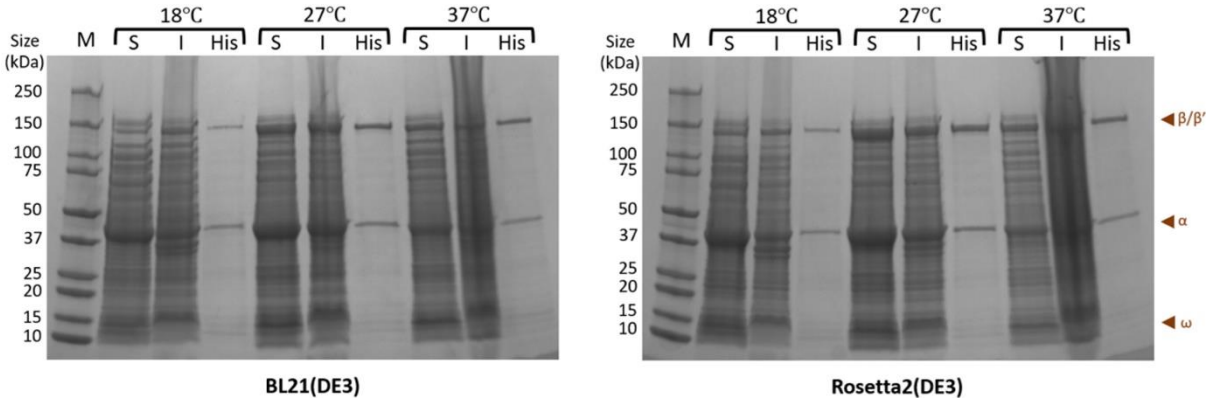

Figure 3-4. *B.subtilis* RNAP expression optimisation in BL21(DE3) and Rosetta2(DE3) *E.coli* strains. SDS-PAGE page gels displaying soluble and insoluble expression of all core RNAP subunits in BL21(DE3) and Rosetta2(DE3) cells under different expression conditions. *E.coli* strains are shown below their respective gel. Expression was induced with IPTG (1mM) at an OD_{600} of 0.5-0.6 and cells incubated at either 37 °C for 3 hours, 27 °C for 3 hours or 18 °C overnight. Formation of RNAP from its core subunits was confirmed using His SpinTrap column pulldowns (His). Rosetta2(DE3) yielded the greatest soluble expression of all core RNAP subunits (Dark Orange arrowheads with symbols). In both gels, M represents the 10-250 kDa protein sizing ladder, S represents the soluble fraction and I represents the insoluble fraction.

3.2.1.2 Purification of PcrA and RNA polymerase

Recombinant RNAP from pNG1256 was first purified to homogeneity as shown in Figure 3-5, whereby bands corresponding to the β, β′ and α subunits can be seen above their approximate masses of 136 kDa, 134 kDa and 35 kDa respectively. A band representing the ω-subunit should be present at approximately 8 kDa, but this could not be seen clearly after size-exclusion chromatography (SEC) (Figure 3-5C). Since pure recombinant RNAP eluted from the size-exclusion column at very low concentrations, it was pooled, dialysed and stored immediately following ion-exchange chromatography (IEX) in subsequent purifications to maximise yield. This was also considered appropriate given that we planned to purify the assembled PcrA:RNAP complex using SEC at later stages in the project in order to assess its formation.

The same rationale was used for the purification of native RNAP from MH5636, whereby a two-column method achieved levels of purity akin to that of the recombinant RNAP after elution from the MonoQ column. Similarly, bands corresponding to the β, β′ and α subunits can be seen above their approximate molecular weights of 136 kDa, 134 kDa and 35 kDa respectively (Figure 3-6). In addition to this, the δ-subunit is present above its mass of 20 kDa and a faint band corresponding to the ω-subunit can be seen above its expected mass of 8 kDa.

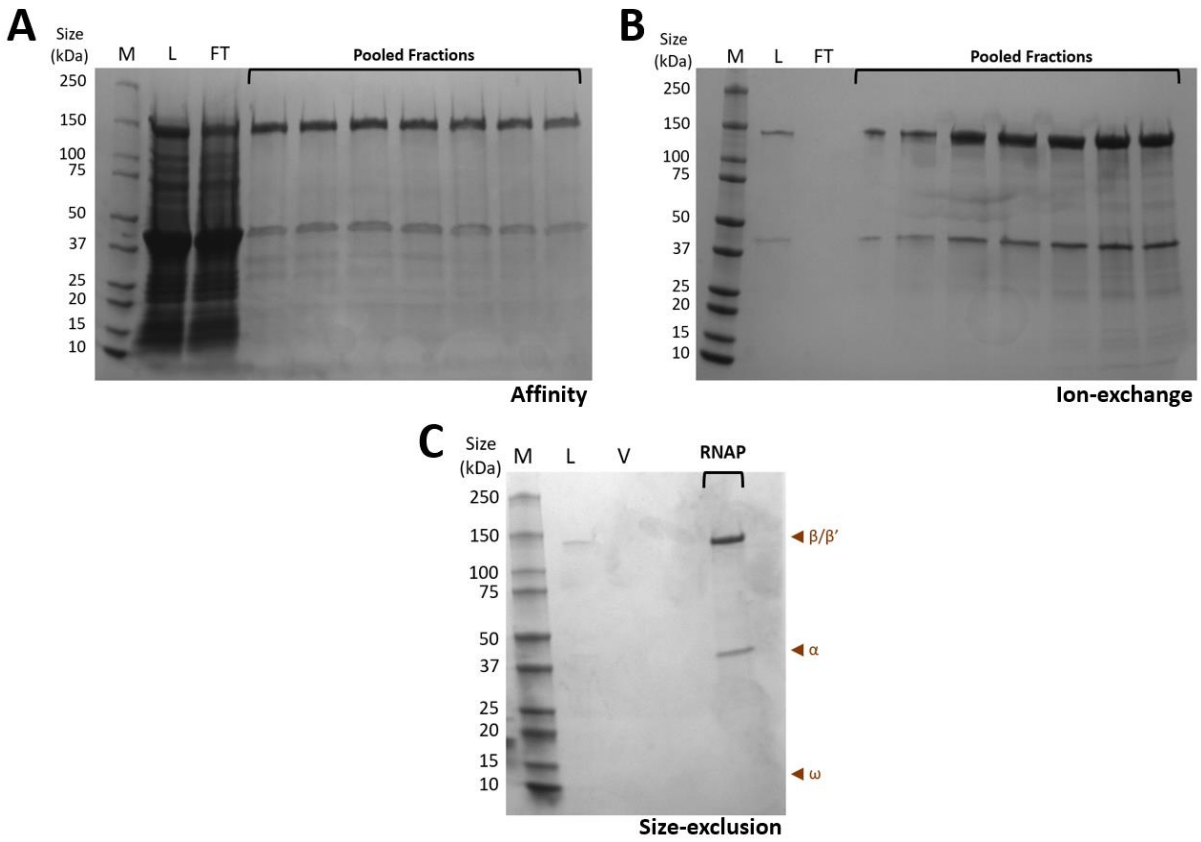

Figure 3-5. Recombinant RNAP purification. The gels are presented in the same order that RNAP was passed through their respective columns. (A) SDS-PAGE page gel of fractions eluted from a 5 ml HisTrap column. (B) SDS-PAGE gel of fractions containing RNAP after MonoQ IEX chromatography. (C) Pure RNAP eluted from a Superose 6 Increase 10/300 Gl size-exclusion column. The β, β' and α subunits can be clearly seen (Dark orange arrowheads). The expected positioning of the ω subunit is also annotated. Where appropriate, M represents the 10-250 kDa protein sizing ladder, L represents load, FT represents the flow through and V represents the void.

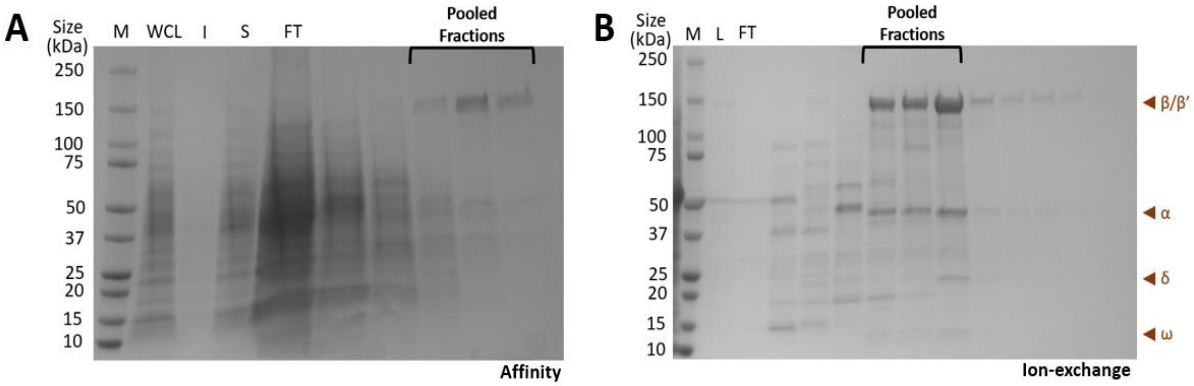

Figure 3-6. Native RNAP purification from MH5636. The gels are presented in the same order that RNAP was passed through their respective columns. (A) SDS-PAGE page gel of fractions eluted from a 5 ml HisTrap column. (B) SDS-PAGE gel of fractions containing RNAP after MonoQ IEX chromatography. The β, β' and α subunits can be clearly seen, with faint bands representing the δ and ω also present (Dark orange arrowheads). Where appropriate, M represents the 10-250 kDa protein sizing ladder, WCL represents whole-cell lysate, I represents the insoluble fraction, S represents the soluble fraction, L represents load and FT represents the flow through.

PcrA was purified using a three-column method, which produced higher concentrations and total yields than recombinant RNAP preparations (Figure 3-7). Purification of PcrA was verified by comparison with pure material previously obtained by our lab (Figure 3-7C). However, a second band was present directly below the expected PcrA band. This band can also be seen in the older material, but it is not as prominent.

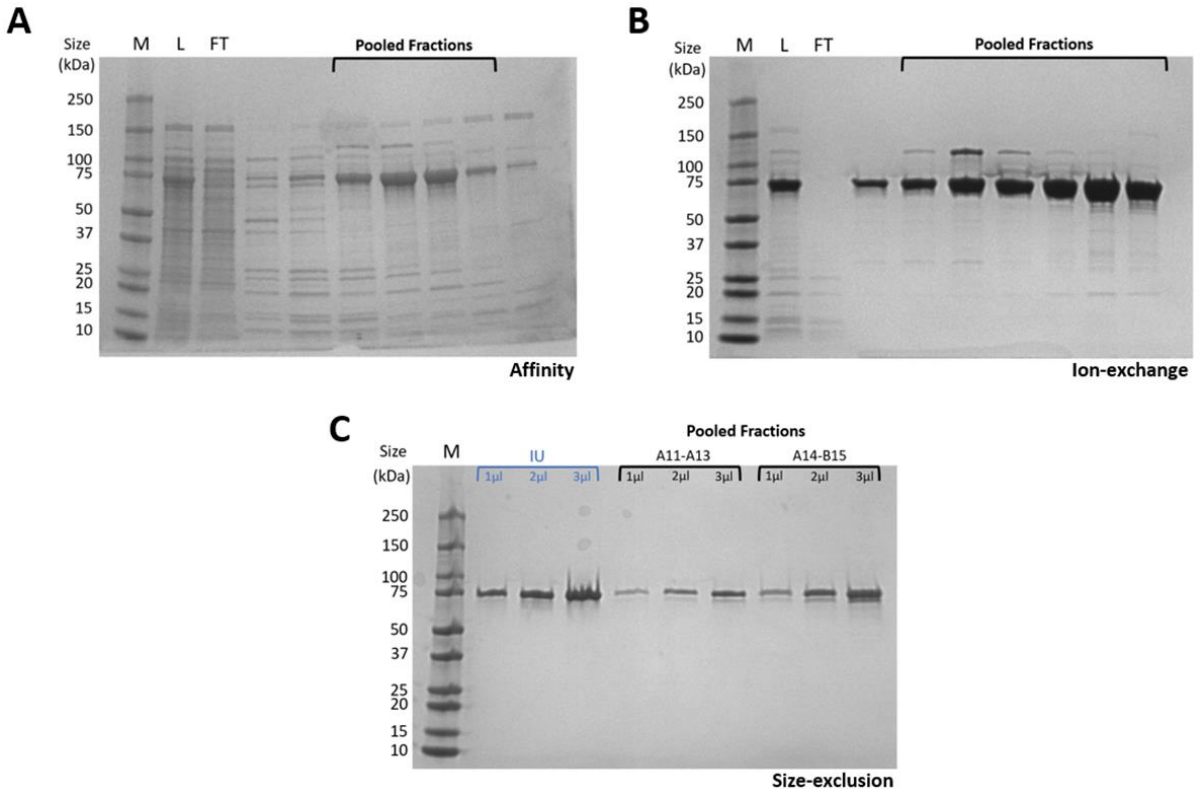

Figure 3-7. PcrA purification. The gels are presented in the same order that PcrA was passed through their respective columns. (A) SDS-PAGE page gel of fractions eluted from a 5 ml HiTrap Heparin HP column. (B) SDS-PAGE gel of fractions containing PcrA after MonoQ IEX chromatography. (C) Comparison of pure PcrA eluted from a Superdex 200 16/60 size-exclusion column with previous material obtained by the lab (IU). Different loading volumes were tested for two separate samples of pooled material (A11-A13 and A14-B15). Where appropriate, M represents the 10-250 kDa protein sizing ladder, L represents load and FT represents the flow through.

3.2.1.3 Assembly of the PcrA:RNAP complex

Despite evidence showing that core RNAP synthesised from pNG1256 can be purified pre-bound to nucleic acids (Newing et al., 2020), we first attempted to form the complex by creating an artificial transcription elongation complex (TEC) using a DNA:RNA scaffold formed and validated previously by our lab (Figure 3-8, pathway 1/RHS) (Urrutia-Irazabal et al., 2021). The reasoning behind this was that PcrA would either favour the substrate already bound to the RNAP, or the additional scaffold would displace nucleic acids already present to form a

more desirable complex for PcrA. However, when RNAP purified from pNG1256 was used, the RNAP and PcrA largely eluted as separate peaks from a SEC column in a reproducible manner (Figure 3-9A).

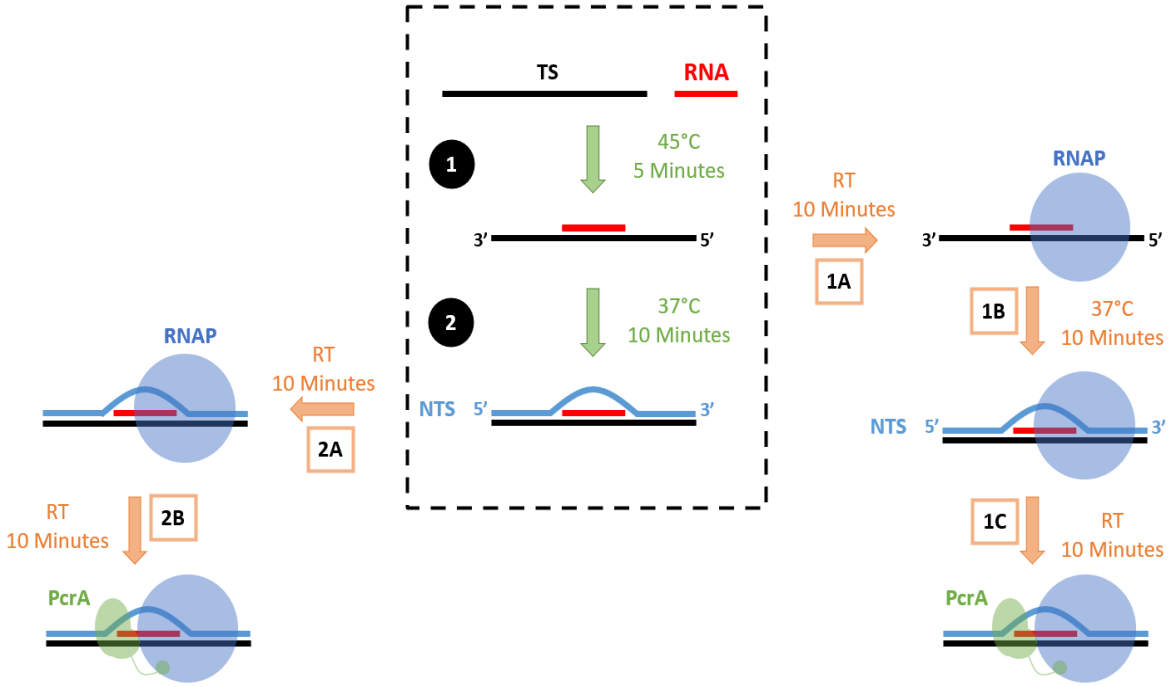

Figure 3-8. Methods for scaffold and PcrA:RNAP complex formation. Pre-formed nucleic acid scaffolds (omitting proteins) were formed by first annealing RNA (red) to the template strand (TS) (black), before incubation with the non-template strand (NTS) (blue line) (black dashed box). Protein complexes were either formed by incubation of a pre-formed scaffold with RNA polymerase (RNAP) (blue circle), followed by incubation with PcrA (green) (Pathway 2, LHS) or by incubation of the RNA:TS hybrid with RNAP, followed by annealing of the NTS and incubation with PcrA (Pathway 1, RHS).

.

To test if this was due to the RNAP material itself or the methods used, native RNAP (from MH5636), which had been purified by a previous lab member, was used to form the complex. When this was done, RNAP and PcrA co-eluted across multiple peaks, albeit not in stoichiometric quantities (Figure 3-9C). As shown in Figures 3-9C and 3-9D, RNAP subunits β, β′, α and ω eluted together with PcrA between ~10-11 ml (peak 2) with the addition of an unidentified band around 50 kDa. PcrA also eluted with RNAP subunits β, β′ and α between ~12-14 ml (peak 3). Excess PcrA appeared to elute as a separate peak later, between ~14-15 ml (peak 4). No protein was present in material from peaks 5 and 6 (Figure 3-9C) and they returned 260/280 ratios of >1.5, indicating that these peaks corresponded to excess nucleic acids.

The first peak in which the native RNAP and PcrA co-eluted (peak 2) is almost completely absent in the chromatograph from SEC with the recombinant RNAP and PcrA (Figure 3-9B). Faint bands corresponding to the β, β′ and α subunits of RNAP were present in material from peak 3 (Figure 3-9A) suggesting some complex may have formed, but at concentrations too low for imaging applications. Furthermore, peak 1 is slightly larger and occurs directly after the column void volume (8ml), where aggregated proteins are eluted. This data suggests that the recombinant RNAP formed aggregates, potentially limiting formation of the PcrA:RNAP complex.

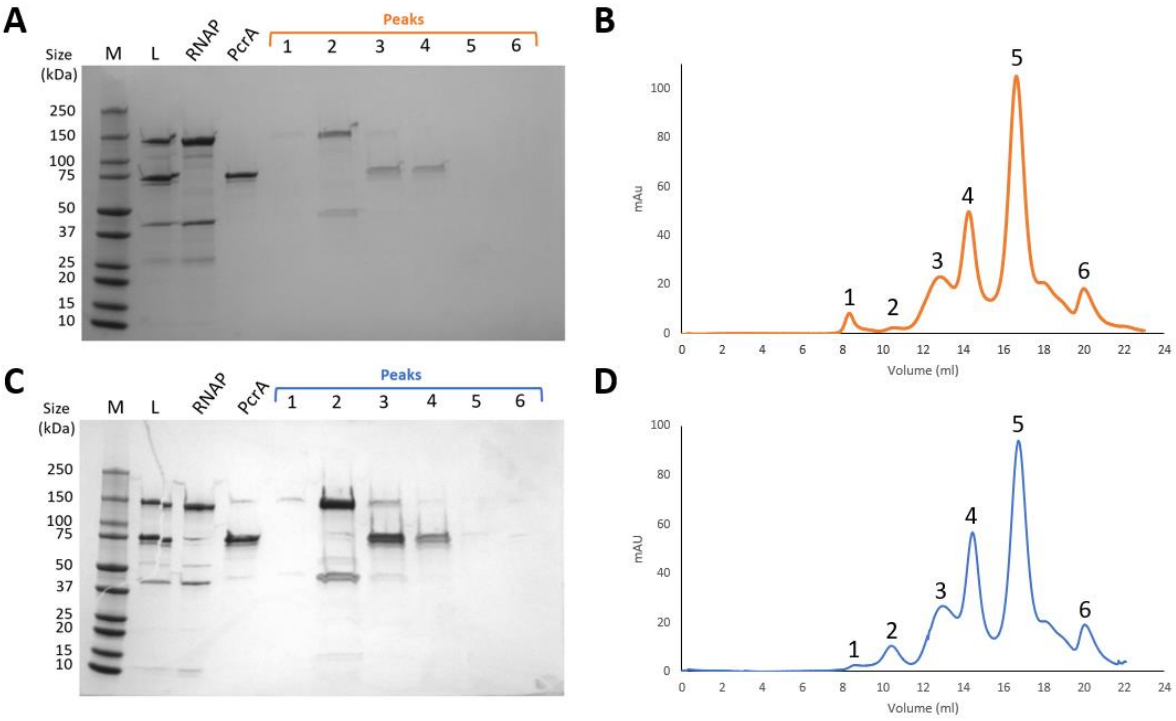

Figure 3-9. In vitro PcrA:RNAP complex formation. (A) SDS-PAGE page gel of eluted proteins from size-exclusion chromatography (SEC) after attempted PcrA:RNAP complex formation using core RNAP purified from pNG1256. (B) Chromatograph from SEC after attempted PcrA:RNAP complex formation using core RNAP purified from pNG1256. Peaks have been annotated with numbers which correspond to the numbered lanes of the SDS-PAGE gel in panel A. (C) SDS-PAGE page gel of eluted proteins from SEC after attempted PcrA:RNAP complex formation using RNAP purified from MH5636. (D) Chromatograph from SEC after attempted PcrA:RNAP complex formation using RNAP purified from MH5636. Peaks have been annotated with numbers which correspond to the numbered lanes of the SDS-PAGE gel in panel C. In both cases, SEC was carried out using a Superose 6 Increase 10/300 Gl column. M represents the 10-250 kDa protein sizing ladder and L represents load.

A second explanation for why the complex may not have formed using the recombinant RNAP is that the nucleic acids already bound to the polymerase were not displaced by the additional scaffold. In this scenario, co-purified nucleic acids are either an unsuitable substrate for PcrA recruitment or promote an inappropriate RNAP conformation. However, we continued to

investigate formation of the PcrA:RNAP complex with the recombinant RNAP due to its more homogenous nature. Hence, the RNAP was purified again, but with a high-salt (1.5 M NaCl) wash step included during the His-column purification in an attempt to remove bound nucleic acids. Initially, the His eluate was dialysed against the standard QA buffer, but the low NaCl concentration caused the protein to precipitate within the dialysis tubing. This is a common property of DNA bound proteins that are devoid of their nucleic acid binding partner, and suggested to us that the high salt-wash step was having the desired effect of removing co-purified DNA/RNA. To avoid precipitation, the NaCl concentration of the QA buffer was increased to 275 mM and the eluted protein was loaded directly onto the MonoQ column, avoiding overnight dialysis. Higher concentrations of recombinant RNAP were achieved this way, but as with the original method, only the β, β′ and α subunits can be seen clearly (Figure 3-10B).

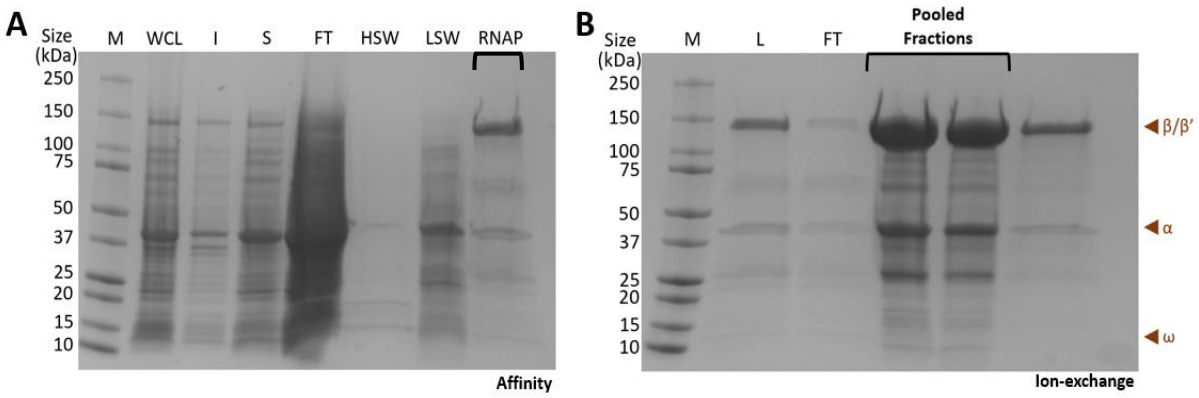

Figure 3-10. Recombinant RNAP purification with high salt wash step. The gels are presented in the same order that RNAP was passed through their respective columns. (A) SDS-PAGE page gel of fractions eluted from a 5ml HisTrap column after a column wash with 1.5 M NaCl. (B) SDS-PAGE gel of fractions containing RNAP after MonoQ IEX chromatography. Presence of the β, β', α and ω subunits is indicated by dark orange arrowheads. Where appropriate, M represents the 10-250 kDa protein sizing ladder, WCL represents the whole cell lysate, I represents the insoluble fraction, S represents the soluble fraction, HSW represents a high salt wash (1.5 M NaCl), LSW represents a low salt wash (500 mM NaCl), L represents load, FT represents the flow through and RNAP represents the eluted RNA polymerase.

The “washed” recombinant RNAP was then used to form the PcrA:RNAP complex with a synthetic scaffold using the same methods described above. However, PcrA and RNAP continued to elute separately (Figure 3-11A). Aggregation of RNAP may have contributed to this due to the tendency of RNAP to aggregate noted during the purification and the presence of a larger aggregate peak (peak 1, Figure 3-11B) in comparison to when the original

recombinant RNAP material was used (peak 1, Figure 3-9B). As a result, little non-aggregated RNAP may be available for PcrA (or the added scaffold) to interact with.

Due to earlier evidence for co-elution of native RNAP and PcrA (Figure 3-9C), the RNAP expressed and purified from MH5636 as part of this work was also used in an attempt to form a stable PcrA:RNAP complex. Two bands representing both the β/β′ and α subunits were present in the second peak eluted from the SEC column, and a very faint band can be seen at the same level as that of PcrA (Figure 3-11C). This band was more visible directly from the gel. However, peak 2 eluted between ~14-15 ml, which is later than the co-elution of RNAP and PcrA when using the older material, and any complex formed was again, nonstoichiometric. SEC of the complex formed using the older material was carried out on a different AKTA to the “washed” recombinant and newer native material. Although this should not alter the elution profile in theory, the differences seen here could be an artefact of the change in system. Nevertheless, whilst some complex may be present, it is likely that it is at concentrations too low for imaging applications.

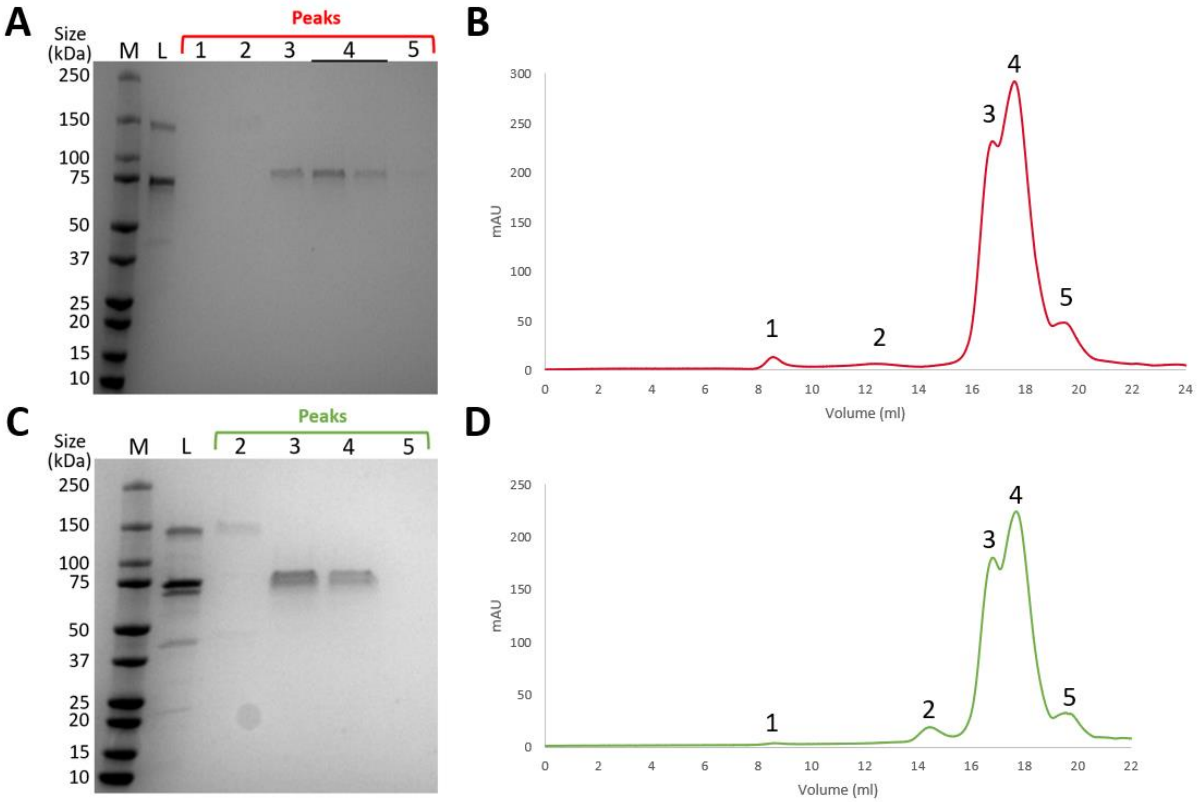

Figure 3-11. In vitro PcrA:RNAP complex formation using washed recombinant RNAP and *B.subtilis* expressed RNAP. (A) SDS-PAGE page gel of eluted proteins from size-exclusion chromatography (SEC) after attempted PcrA:RNAP complex formation using core RNAP purified from pNG1256 that had undergone a 1.5 M NaCl wash. (B) Chromatograph from SEC after attempted PcrA:RNAP complex formation using core RNAP purified from pNG1256 that had undergone a 1.5 M NaCl wash. Peaks have been annotated with numbers which correspond to the numbered lanes of the SDS-PAGE gel in panel A. (C) SDS-PAGE page gel of eluted proteins from SEC after attempted PcrA:RNAP complex formation using RNAP purified from MH5636. (D) Chromatograph from SEC after attempted PcrA:RNAP complex formation using RNAP purified from MH5636. Peaks have been annotated with numbers which correspond to the numbered lanes of the SDS-PAGE gel in panel C. In both cases, SEC was carried out using a Superose 6 Increase 10/300 Gl column. M represents the 10-250 kDa protein sizing ladder and L represents load.

3.2.2 PcrA:RNAP complex formation in vivo

Since PcrA is known to interact with RNAP (Gwynn et al., 2013), our second approach to obtaining a stable complex involved co-expressing the proteins to form the complex in *E. coli*. This was done using the recombinant RNAP expressed from pNG1256 and PcrA from pET-22b_PcrA. Co-expression was first assessed in small-scale (100 ml) trials under three conditions: 37 °C for 3 hours, 27 °C for 3 hours and 18 °C overnight. Whilst the individual proteins had their greatest expression at lower temperatures, co-expression yielded the greatest expression of both RNAP and PcrA at 37 °C (Figure 3-12). Hence, these conditions were used for subsequent expressions. We also assessed if untagged PcrA would be pulled

down with His-tagged RNAP using Ni-NTA resin. As shown in Figure 3-12, it appeared as if both proteins eluted together stoichiometrically, providing evidence that interactions had successfully formed in vivo. There were, however, two bands present at the expected position of the PcrA band.

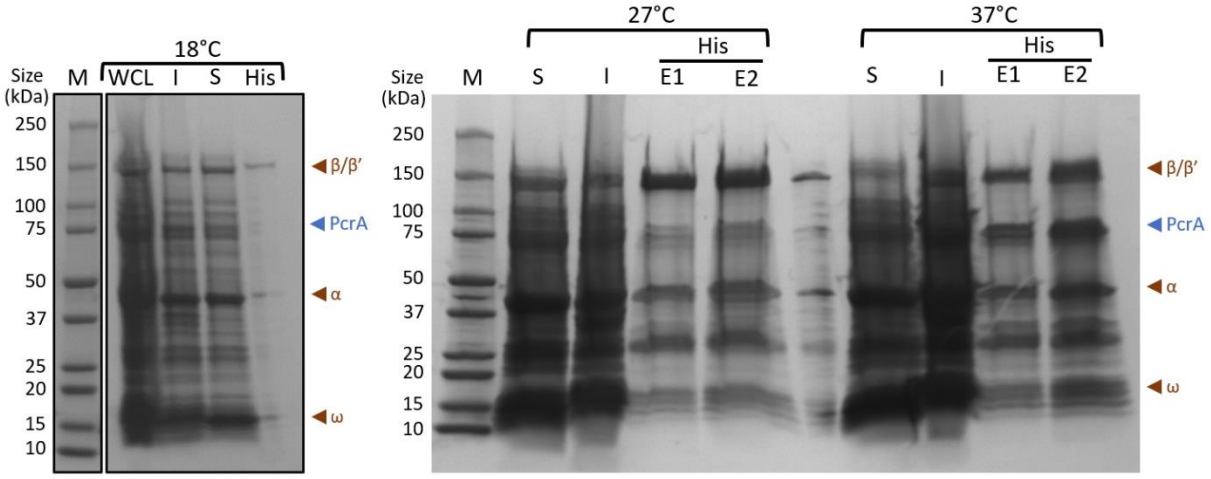

Figure 3-12. PcrA/RNAP co-expression optimisation. SDS-PAGE gels displaying soluble and insoluble expression of RNAP and PcrA under different conditions. (*Left*) Expression at 18 °C overnight (*Right*) Expression at 27 °C for 3 hours and expression at 37 °C for 3 hours. Ni-NTA Agarose and His SpinTrap columns were used to assess PcrA pull-down with RNAP. Where appropriate, M represents the 10-250 kDa protein sizing ladder, WCL represents whole cell lysate, S represents the soluble fraction, I represents the insoluble fraction, PD represents pull-down. Expected size of the PcrA band is indicated by blue arrowheads.

When material from large-scale expressions was purified using Ni-NTA resin, most of the protein was lost in the flow-through and wash (Figure 3-13A). Despite this, two bands present at the expected mass of PcrA still co-eluted with RNAP from the Ni-NTA resin as well as from the MonoQ (Figure 3-13). However, not all RNAP subunits were visible in all fractions containing the suspected PcrA after IEX. Fractions which appeared to contain all the subunits of RNAP and PcrA were pooled and loaded onto a Superose 6 Increase 10/300 SEC column. It is not clear from the gel image that all RNAP subunits were present in the pooled fractions (Figure 3-13B), but bands were visible on the physical gel. Protein eluted from the SEC column was at concentrations too low for visualisation on an SDS-PAGE gel, even after it had been concentrated.

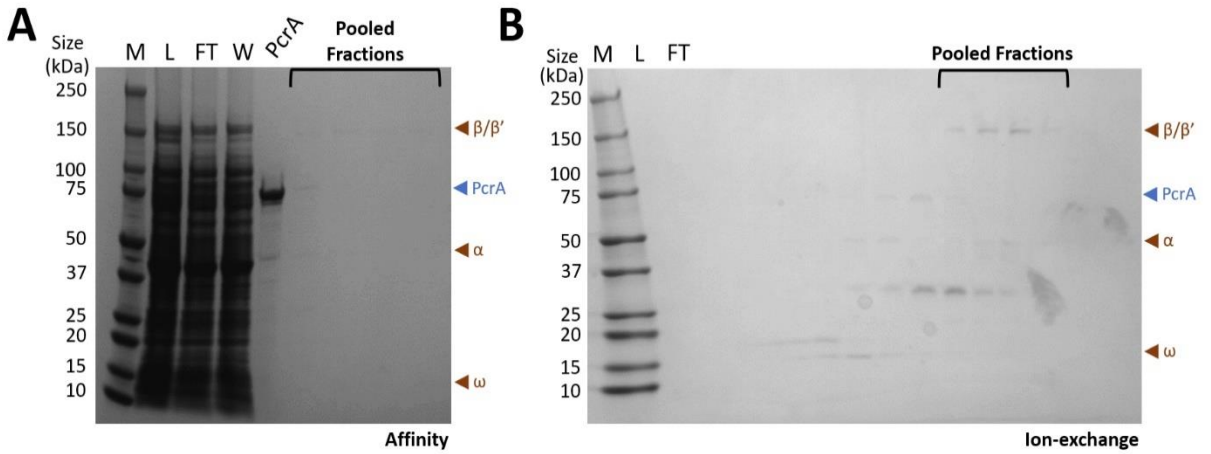

Figure 3-13. In vivo complex formation. The gels are presented in the same order that the co-expressed material was passed through their respective columns. (A) SDS-PAGE gel of fractions eluted from a 5 ml HisTrap column (B) SDS-PAGE gel of fractions containing RNAP and a band of the same size as PcrA after MonoQ IEX chromatography. Where appropriate, M represents the 10-250 kDa protein sizing ladder, L represents load, FT represents the flow through, W represents the wash. Expected band sizes for PcrA and RNAP subunits β, β', α and ω are indicated by blue and dark orange arrowheads respectively.

To determine if and which band present at the expected mass of PcrA was in fact the helicase, both bands were excised from an SDS-PAGE gel after Ni-NTA purification and analysed by mass spectrometry (MS). Measurements and analysis were completed by the University of Bristol Proteomics Facility. Results returned from MS identified that the higher band of the two had a sequence coverage (% of protein sequences covered by the identified peptides) of 8.66 % for PcrA but had a sequence coverage of 60.45 % for the *E.coli* protein, ArnA. ArnA is just over 9 kDa smaller than PcrA and hence, will run close to PcrA on an SDS-PAGE gel. There were also further *E.coli* contaminants with higher sequence coverage scores than PcrA (Table A-1, Appendix). Additionally, the lower band of the two had no sequence coverage for PcrA. Therefore, we conclude the band we thought represented PcrA that had co-purified with RNAP was in fact predominantly *E.coli* ArnA. Whilst some PcrA was present, this would be at concentrations not feasible for imaging applications of the complex, particularly as more would undoubtedly be lost in subsequent purification steps.

3.2.3 PcrA:RNAP complex formation by mixing of cell lysates

The third method used in an attempt to form the PcrA:RNAP complex involved combining resuspended cell pellets from separate PcrA and recombinant RNAP expressions, and simultaneously lysing the cells. When the resulting soluble material was loaded onto a HisTrap column, the large majority of the PcrA passed through in the flowthrough, and most of the

remaining PcrA was removed during the wash (Figure 3-14A). This suggests that either the complex was not able to form under these conditions, or that the interactions formed were too weak to remain intact on the HisTrap column during the wash steps. Despite this, the eluate was pooled and loaded onto a MonoQ column. A band of the same molecular weight as PcrA was eluted from the column first, and in later fractions, RNAP subunits were present (Figure 3-14B). Interestingly, the RNAP appeared to dissociate on the MonoQ column, with fractions containing the α and ω eluting first without the β/β′ subunits.

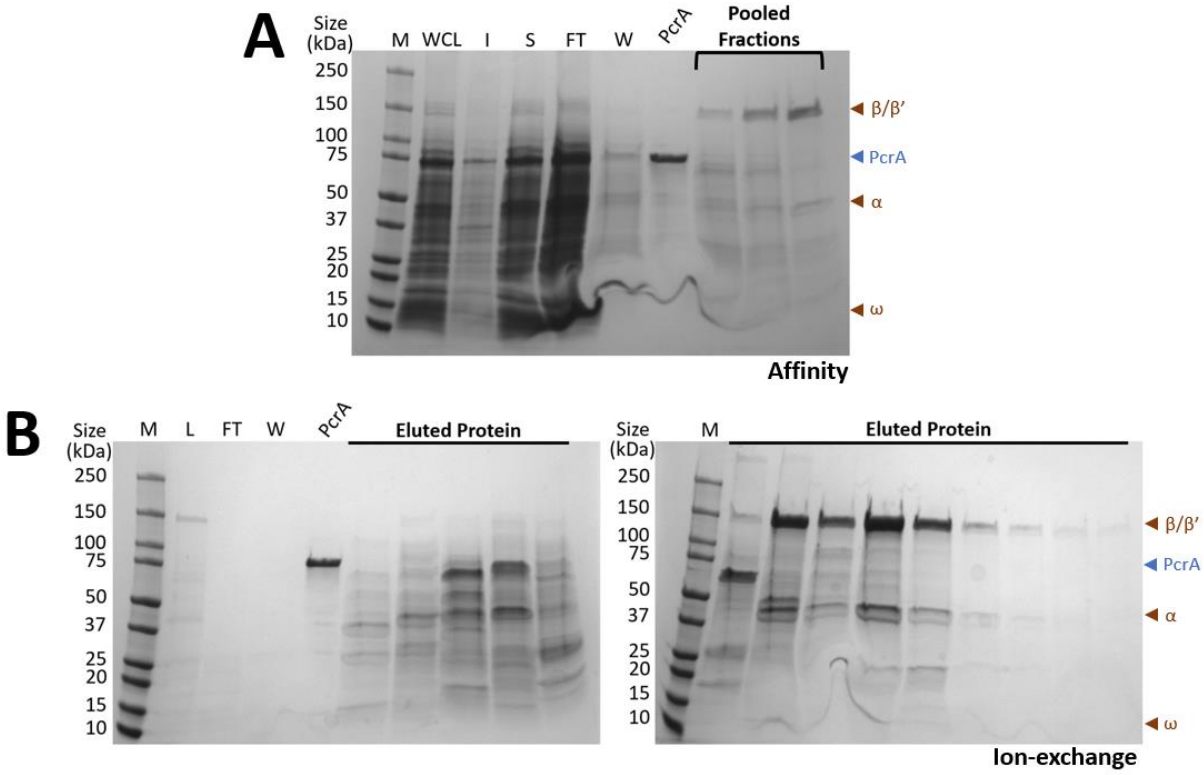

Figure 3-14. Attempted complex formation by mixing cell lysates. The gels are presented in the same order that the mixed material was passed through their respective columns. (A) SDS-PAGE gel of fractions eluted from a 5ml HisTrap column (B) SDS-PAGE gel of fractions eluted from a MonoQ IEX column. Where appropriate, M represents the 10-250 kDa protein sizing ladder, WCL represents whole cell lysate, S represents the soluble fraction, I represents the insoluble fraction, L represents load, FT represents the flow through, W represents the wash, PcrA represents a pure sample of PcrA obtained by a previous member of the lab. Expected band sizes for PcrA and RNAP subunits β, β', α and ω are indicated by blue and dark orange arrowheads respectively.

3.2.4 Testing for interaction between purified RNAP, PcrA and synthetic scaffolds using EMSAs

In order to determine the composition of the complexes that would be analysed by SEC, EMSAs were set up to further study the interactions between RNAP, PcrA and a synthetic

scaffold in vitro. This method has advantages over SEC alone for several reasons: it is not possible to fully ascertain complete and correct complex formation via SEC, EMSAs provide crude insights into binding affinities and testing interaction under different conditions is more practical with EMSAs. Formation of the scaffold alone was first assessed, in which the template strand (TS) was 5′ Cy5 labelled. The scaffold was either formed using similar methods to those used for the complex formation, whereby the TS is first annealed to the RNA, before omitting addition of the RNAP and instead annealing the non-template strand (NTS), or by combining all scaffold components in equimolar amounts at once and heating to 95 °C. As shown in Figure 3-15, the results were identical for both methods. In each case, the band corresponding to the TS was shifted on the addition of the NTS as well as both the NTS and RNA together. There were however two species present, with one band remaining at the same level as the TS alone. No observable shift occurred on addition of the RNA, implying one of two possibilities: the RNA did not anneal, possibly due to its lability, or the EMSA is not sensitive enough to detect a shift on the addition of such short RNA strands (9 nucleotides).

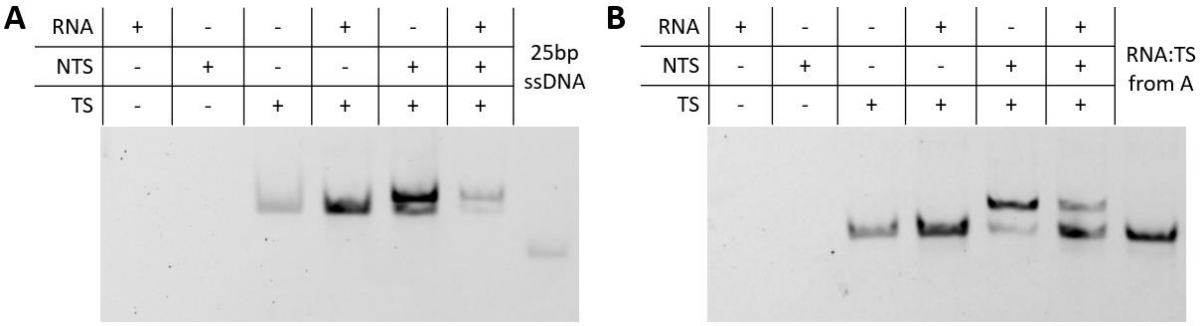

Figure 3-15. Scaffold formation. EMSA's were used to assess formation of the nucleic acid scaffold alone. 12 % TBE-PAGE gels were run at 150 V for 30 minutes using the Mini-PROTEAN Tetra Cell system (Bio-Rad). In both cases, the Template strand (TS) is 5' Cy5 labelled. (A) Scaffold formation by combining the TS, Non-template strand (NTS) and RNA at once and heating to 95 °C, before cooling to room temperature overnight. The control is a Cy5 labelled 25 bp ssDNA oligonucleotide. (B) Scaffold formation by first annealing the RNA to the TS at 45 °C, before then annealing the NTS at 37 °C. The control in this case is the RNA:TS hybrid from the 95 °C scaffold formation in panel A.

Next, interactions between RNAP (both purified from MH5636 and the "washed" recombinant material) with the scaffold (formed using both methods in Figure 3-8) were assessed. As shown in Figure 3-16, both RNAP's were able to shift the scaffold formed by both methods, indicated by the decreasing intensity of the scaffold bands and an increasing intensity of the RNAP bands as RNAP concentration was increased. This implies formation of a TEC in each case. However, RNAP from MH5636 was able to shift the scaffold at somewhat lower concentrations (between 0.24-0.48 µM RNAP) compared with the recombinant RNAP

(between 0.48-0.95 µM RNAP) (Figure 3-16). The shift is less well observed when using the recombinant RNAP since it remained concentrated within the wells of the gel (Figures 3-16C and 3-16D). The EMSAs also showed that PcrA was able to shift the scaffold much more effectively than either RNAP. This is unsurprising as, in the absence of RNAP, the scaffold contains a sufficient region of unpaired ssDNA which is a good binding substrate for PcrA. When interactions were tested using the method for TEC formation that involves annealing strands separately, laddering both above and below the position of the scaffold alone occurred. Interestingly, the laddering above the scaffold has a somewhat regular pattern and is more intense with the recombinant RNAP. In addition, multiple species were formed at the top of the gel when the recombinant RNAP was used, making it difficult to determine which band represents RNAP:Scaffold complex and the identity of the other bands.

Following the successful shift of each scaffold with RNAP alone, PcrA was next titrated against a fixed concentration of the pre-formed TEC (RNAP:Scaffold complex) in an attempt to form a PcrA:TEC complex. We first used the single-step scaffold with native RNAP at two different concentrations of RNAP (0.475 µM and 0.95 µM; Figure 3-17A) selected on the basis of the previous EMSAs. PcrA was titrated against these fixed concentrations of the TEC, starting with a ratio of 1:1 (RNAP:PcrA) and increasing to a ratio of 1:2 (at 0.95 µM RNAP) or 1:4 (at 0.475 µM RNAP). At both RNAP concentrations, the TEC band was super-shifted upon addition of PcrA. As PcrA concentrations increased, the shift was greater, but also generated two distinct bands with some material remaining in the wells of the gel.

Since the TEC was shifted at both RNAP concentrations, the lower concentration of 0.475 µM was used in subsequent EMSAs since it allowed higher ratios of RNAP:PcrA to be achieved. When PcrA was titrated against the TEC formed using the same scaffold, but with the recombinant RNAP, the TEC band was super-shifted again. However, in the presence of PcrA, essentially all of the labelled scaffold was retained in the wells, making it difficult to assess the affinity of PcrA for the TEC or the homogeneity of the resulting complex (Figure 3-17C). As has been mentioned before, there were issues with aggregation during the purification of the recombinant RNAP, and this behaviour may reflect this same issue.

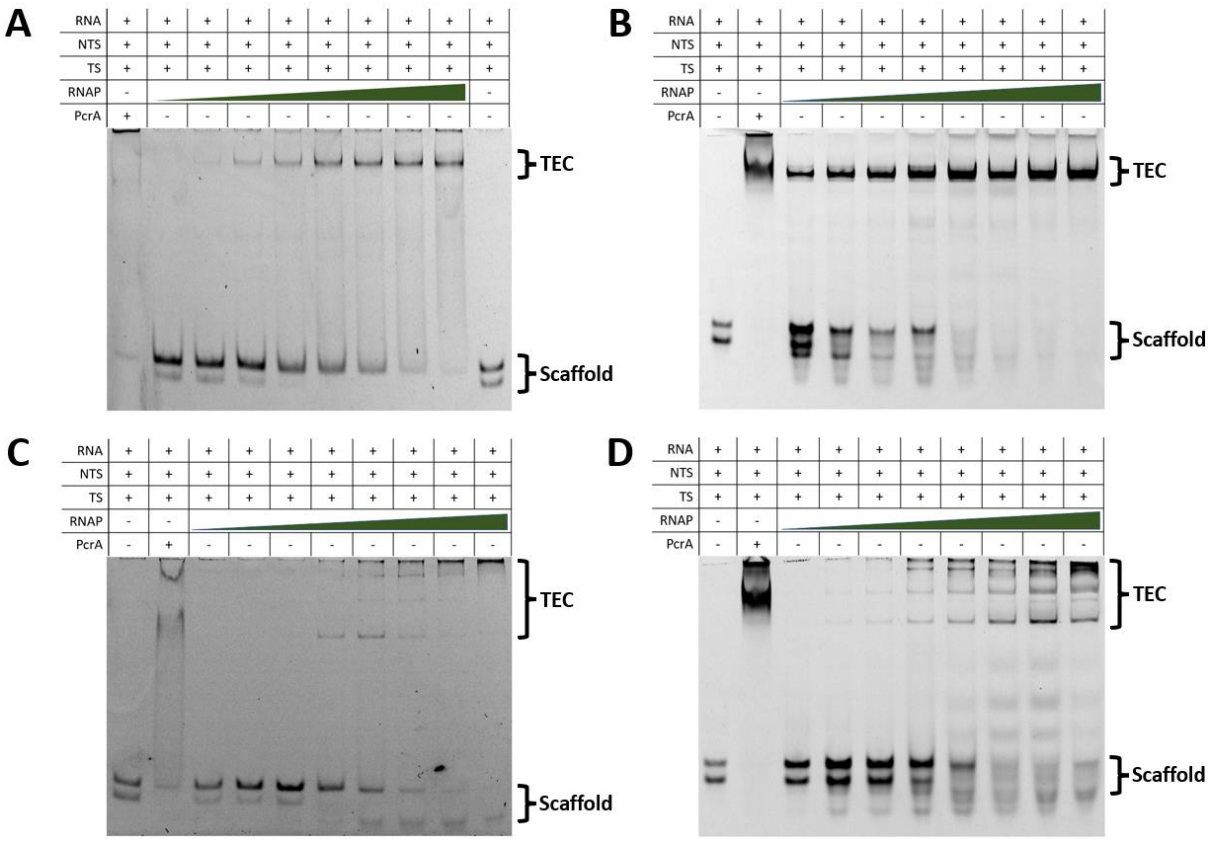

Figure 3-16. Interaction between RNAP and nucleic acid scaffolds. EMSA's were used to assess interaction between the RNAP formed from both methods and the nucleic acid scaffolds formed from both methods. RNAP concentrations increase from 0.09 µM up to 7.57 µM. 6 % TBE-PAGE gels were run at 150 V for 35 minutes using the Mini-PROTEAN Tetra Cell system (Bio-Rad). In all gels, TS represents the template strand and NTS represents the non-template strand. (A) Interaction between native RNAP expressed in MH5636 and the scaffold formed in a single step at 95°C. (B) Interactions between native RNAP expressed in MH5636 and the scaffold formed in the stepwise manner used previously for PcrA:RNAP complex formation. (C) Interactions between the recombinant RNAP expressed from pNG1256 and the scaffold formed in a single step at 95 °C. (D) Interactions between the recombinant RNAP expressed from pNG1256 and the scaffold formed in the stepwise manner used previously for PcrA:RNAP complex formation.

Interactions with PcrA were also tested with the TECs formed using the 3-step approach. When native RNAP was used, the addition of PcrA resulted in the formation of two distinctive shifted bands (Figure 3-17B). A third band was also present below the level of both the TEC and PcrA:Scaffold bands. Furthermore, at the two lowest concentrations of PcrA, a faint fourth band ran at the same level of the TEC alone, indicating the presence of some residual TEC. As the concentration of PcrA increased, this faint band disappeared, and the lowest band increased in intensity, whilst the two shifted bands decreased in intensity.

When recombinant RNAP was used, the addition of PcrA again resulted in a shift of the TEC (Figure 3-17D). However, as with the complex formed using the 1-step scaffold, material remained in the wells of the gel. A band present at the same distance as that of PcrA bound to the scaffold was also visible, which decreased in intensity at greater PcrA concentrations.

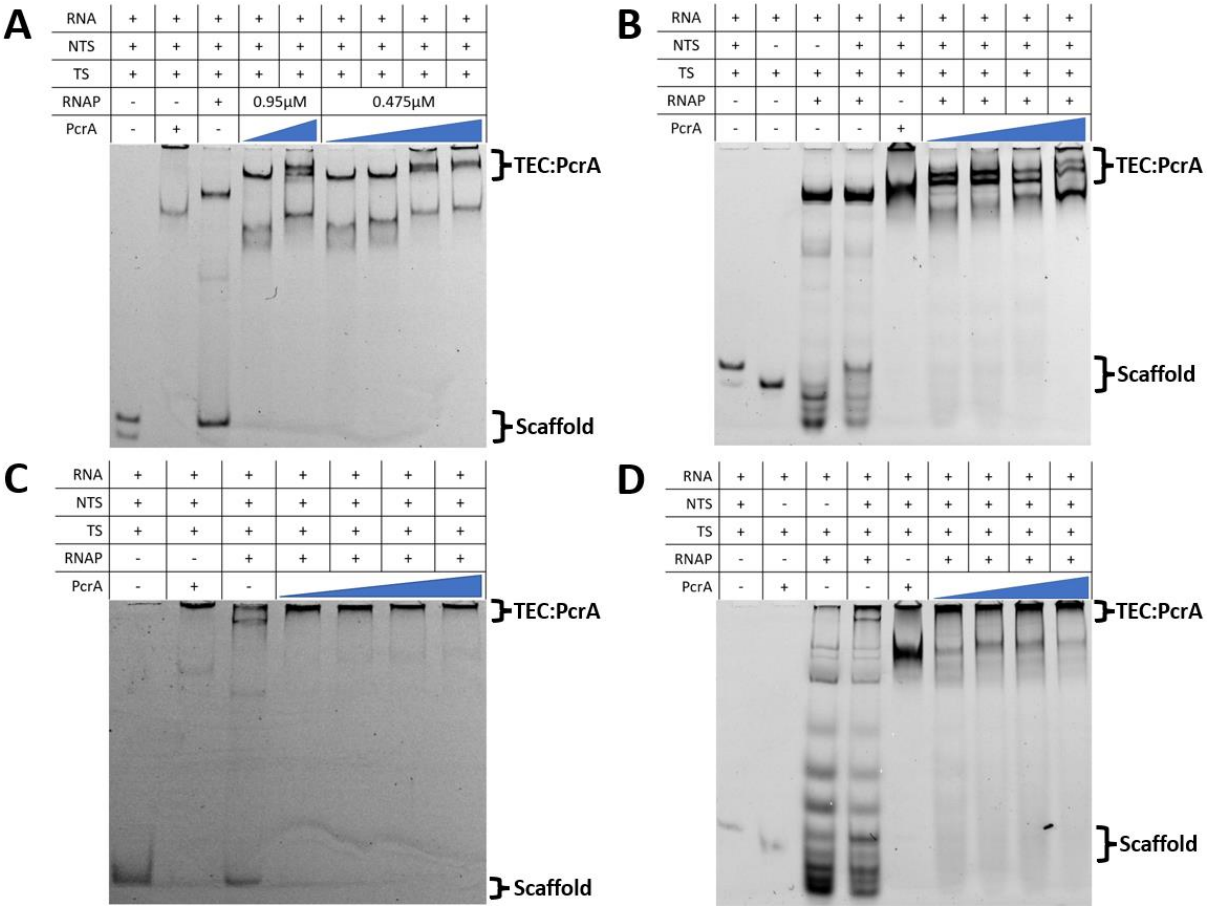

Figure 3-17. Interaction between TECs and PcrA. EMSA's were used to assess interaction between the TEC's formed using different combinations of RNAP and nucleic acid scaffold. RNAP concentrations were fixed at 0.475 µM unless otherwise stated. PcrA was titrated against the TEC starting at a 1:1 ratio and increasing to 1:2 when 0.95 µM RNAP was used, or 1:4 when 0.475 µM RNAP was used. 6 % TBE-PAGE gels were run at 150 V for 35 minutes using the Mini-PROTEAN Tetra Cell system (Bio-Rad). In all gels, TS represents the template strand and NTS represents the non-template strand. (A) Interaction between RNAP expressed from MH5636 bound to the scaffold formed in a single step at 95 °C and PcrA. (B) Interactions between RNAP expressed from MH5636 bound to the scaffold formed by the 3-step method and PcrA. (C) Interactions between the recombinant RNAP expressed from pNG1256 bound to the scaffold formed in a single step at 95 °C and PcrA. (D) Interactions between the recombinant RNAP expressed from pNG1256 bound to the scaffold formed by the 3-step method and PcrA.

As with EMSAs using RNAP alone, laddering occurred above and below the position of the scaffold on the addition of the RNAP to both the complete scaffold, and the RNA annealed to

the template strand. Interestingly, the extra bands either reduced in intensity, or disappeared on the addition of PcrA.

In order to validate the PcrA-TEC interactions we had observed, we next investigated the interaction between native RNAP and PcrA with a C-terminal deletion (PcrA ΔCTD). It has been shown previously that PcrA's CTD forms a crucial part of the interface with RNAP, thus PcrA ΔCTD should not shift the TEC (Urrutia-Irazabal et al., 2021). Interactions were tested both with TECs formed using the 1-step method and the 3-step method. RNAP from pNG1256 was not used due to the aggregation issues seen previously. As shown in Figure 3-18, the TEC was not shifted on addition of the PcrA mutant in both cases. However, the intensity of the TEC band decreased on addition of the mutant PcrA, whilst the lower band representing PcrA ΔCTD remained the same intensity as that of PcrA ΔCTD and the scaffold alone. The PcrA ΔCTD band shifted slightly as its concentration increased, but did not run higher than the PcrAΔCTD:Scaffold complex at any point.

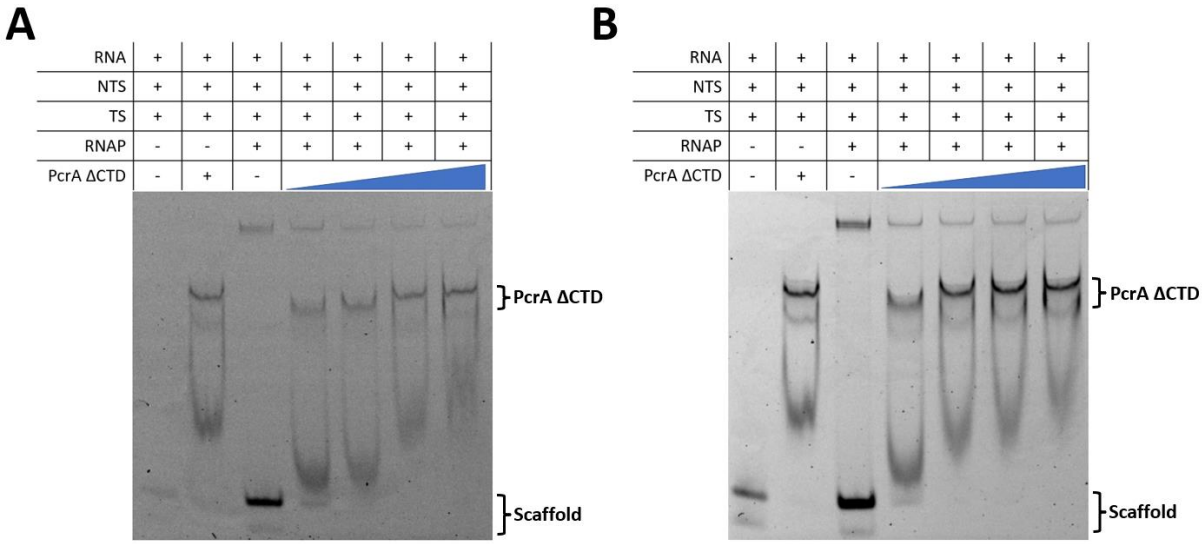

Figure 3-18. ΔCTD PcrA mutant does not interact with RNAP. EMSA's were used to assess interaction between the TEC formed using *B.subtilis*-expressed RNAP and a mutant of PcrA in which the CTD has been removed. RNAP concentrations were fixed at 0.475 µM. PcrA was titrated against the TEC starting at a 1:1 ratio and increasing to 1:4. . 6 % TBE-PAGE gels were run at 150 V for 35 minutes using the Mini-PROTEAN Tetra Cell system (Bio-Rad). In both gels, TS represents the template strand and NTS represents the non-template strand. (A) TEC's formed by combining all the scaffold components and heating to 95 °C before then adding the RNAP. (B) TEC's formed using the 3-step method.

3.3 Discussion and future directions

Here, different methods for formation of a stable PcrA:RNAP complex for imaging applications were explored. Whilst some complex formed by combining native RNAP and PcrA individually in the presence of an added DNA:RNA scaffold, it was present at concentrations too low for imaging applications, and was not stoichiometric. Attempted complex formation by both co-expression and mixing cell lysates failed to yield any convincing evidence of a stable PcrA:RNAP complex and in vivo complex formation was hampered by contamination from *E.coli* ArnA. When interactions between PcrA and an artificial TEC were assessed by EMSAs, we found that PcrA was able to shift the TEC, both for elongation complexes formed using native RNAP and complexes formed using recombinant RNAP, via two different methods. However, it either resulted in the formation of multiple species, or material became trapped in the wells of the gel, making detailed analysis difficult. Thus, further optimisation of the PcrA:RNAP complex pipeline and identification of its products are required.

We first attempted to form the PcrA:RNAP complex by individually combining both native and recombinant RNAP with PcrA in the presence of an artificial TEC scaffold. However, only low concentrations of a nonstoichiometric complex were obtained when native RNAP was used, and no complex was formed when recombinant RNAP was used. Based on the presence of a peak directly following the void volume, which is greater than the peak containing RNAP (Figure 3-9B, peak 2), it is probable that the recombinant RNAP formed aggregates. This would limit the RNAP available for binding the artificial scaffold and interaction with PcrA. However, the non-aggregated RNAP that remained still failed to generate a stable complex, potentially failing to displace the pre-bound scaffold and lacking the added one. This also suggests that either PcrA requires a specific scaffold, or a certain conformation adopted by RNAP when bound to the specific scaffold in order to interact. Alternatively, the RNAP may not have purified with a pre-bound scaffold as seen by Newing et al. (2020) and for some reason, was unable to bind the added scaffold. The latter scenario is unlikely since the later “washed” recombinant material seemingly had a higher propensity to aggregate, which is suggestive of the loss of pre-bound nucleic acids.

Interestingly, it was only the native RNAP that was purified by a previous member of the lab which led to any convincing complex formation with PcrA. Unlike the native RNAP purified as part of this project, this material contained σ_A. Whilst its presence may not be essential to the recruitment of PcrA and formation of a PcrA:RNAP complex as found by Urrutia-Irazabal (2021), it may assist in the stability of the resulting complex. In fact, it was found that complexes containing σ_A had a K_d which was ten-fold less than those lacking it during analysis

by microscale thermophoresis. Hence, this may explain why our material was shifted by PcrA during EMSAs, but mostly failed to remain intact during SEC.

When in vivo complex formation was attempted, issues arose with *E.coli* contamination, particularly by the protein, ArnA, since it runs close to PcrA on SDS-PAGE gels. In fact, ArnA is a frequent contaminant of His-Tag affinity purifications (Andersen et al., 2013), but interestingly, ArnA was not present in the original His-Tag purifications of RNAP alone. Thus, the simple comparison between these earlier gels and those from later His-Tag purifications of the co-expressed material at first indicates that PcrA is actually present, and in stoichiometric quantities. However, ArnA contamination is particularly prevalent during co-expressions as proteins are often poorly expressed under these conditions, and the presence of several surface-exposed His residues allow ArnA to outcompete proteins of interest for Ni-binding. To limit this contamination in the future, the *E.coli* expression strain “LOBSTR”, which is a derivative of BL21(DE3), could alternatively be used for co-expression of RNAP and PcrA. This strain possesses a genetically modified version of the *arnA* gene, which reduces affinity of the protein for Ni resins (Andersen et al., 2013). A second frequent contaminant of IMAC, GlmS, was also present in both bands analysed by MS, but to a lesser extent (Bolanos-Garcia and Davies, 2006). An expression strain, NiCo21(DE3), has been engineered in which GlmS contains a mutation reducing its affinity for IMAC resins and ArnA is tagged for removal by chitin affinity chromatography. Thus, this provides a second route for reducing contamination during co-expression. However, the issue here more likely arises from weak interactions between PcrA and the recombinant RNAP. Therefore, expression in LOBSTR or NiCO21(DE3) would only aid binding of RNAP and at best, result in a minimal increase in complex concentration since the presence of PcrA is dependent on this interaction. Consequently, reduction in ArnA and GlmS would likely only highlight the absence of PcrA.

As mentioned, combining cell lysates of PcrA and RNAP failed to form any complex. In this circumstance, no artificial scaffold was added at any point, and so the only nucleic acids present, if any, were those found to be prebound to the recombinant RNAP (Newing et al., 2020). It is not possible to attribute the lack of interaction solely to a potential lack of scaffold or the presence of a non-specific one. However, based on the observations made during the high salt wash of the same material, the latter argument is more likely. Therefore, these results support the idea that not only does PcrA require an RNAP:Scaffold complex in order to form interactions, but that it requires a specific scaffold of a particular conformation, or a certain conformation of RNAP which it adopts on binding the specific substrate (Urrutia-Irazabal et al., 2021).

During the project's early stages, we failed to obtain high RNAP yields due to low expression and protein loss during His-Tag purifications. It is possible that expression of the His-tagged β′-subunit was lower than expected since it possesses a "TTG" start codon, which is relatively common in *B.subtilis* but rare in *E.coli* (Blattner et al., 1997, Rocha et al., 1999). This in turn limits the concentration of complex obtained as interaction with the Ni-resin is dependent on the presence of the His-tagged β′-subunit. Mutagenizing the start codon to "ATG" may increase protein yields since this start codon is more common in both organisms. Furthermore, loss of protein during His-Tag purification can also be explained by aggregation, whereby the His-Tag becomes sequestered inside aggregates, failing to interact with the column.

To assess the composition of the scaffold and determine if a lack of interaction between the complex components was responsible for the difficulty in obtaining a stable complex, interactions between the scaffold, RNAP and PcrA were investigated using EMSAs. Although it was not initially clear if the scaffold had formed correctly directly from the EMSAs, the protocol for complex formation was taken directly from Urrutia-Irazabal et al. (2021), in which they obtained a stable complex and validated formation of the complex using transcription assays. Thus, the scaffold formed here should be identical and complete. In addition, both the native and recombinant RNAP were able to shift the scaffold generated to establish artificial TECs. These TECs were then shifted on the addition of PcrA, which we suggest requires a specific scaffold to do so.

PcrA was able to shift the TEC formed by the 3-step method using native RNAP, but multiple unidentified species were generated. Interestingly, the two shifted bands decreased in intensity as the PcrA concentration increased, whilst a lower band present below the level of both the TEC and PcrA:Scaffold complex increased in intensity, and transitioned from a smear into a distinctive band (Figure 3-17). These results suggest that PcrA is causing dissociation of the shifted complex, in which either the RNAP is displaced from the scaffold and a PcrA:Scaffold complex is formed, or certain components of the scaffold are being removed. A combination of these possibilities provides an explanation for the variety of species generated. However, since the lower band is present below the level of both the PcrA:Scaffold and the TEC alone, it suggests that PcrA is rebinding only part of the scaffold. For example, the RNA component may be removed from the R-loop, allowing the complementary DNA strands to reanneal and bind to PcrA. Or, since PcrA has been shown to preferentially bind ssDNA and RNA over dsDNA, it may be instead binding to a TS:RNA hybrid or the TS strand alone (Moreno-Del Álamo et al., 2021). This would also involve removal of the RNAP. Since dsDNA, ssDNA and DNA:RNA hybrids have a different conformations and sizes to the R-loop, PcrA bound to any of these would likely run lower on the gel. This would be consistent with the findings of Urrutia-Irazabal et al. (2021), in which they proposed that PcrA is involved in the

resolution of R-loops. Further studies have also implicated PcrA in R-loop removal (Moreno-Del Álamo et al., 2021).

It is also possible that the two distinct shifted bands represent complexes with either a monomer or dimer of PcrA bound. This would also explain why two bands were generated at higher PcrA concentrations when the single-step scaffold was used with the native RNAP. However, if the proposition that PcrA dimerises through its CTD is true, PcrA would be unable to dimerise and interact with RNAP simultaneously, since the CTD is essential for this interaction (Urrutia-Irazabal et al., 2021). Thus, we favour the former explanation for the presence of multiple species.

The shift of the TEC formed using “washed” recombinant RNAP by PcrA was less clear due to the fact that material concentrated in the wells of the gel. This was true for TECs generated using both complex formation approaches. Since the aggregate peak from SEC of the PcrA:RNAP complex with the “washed” recombinant RNAP was fairly large, and only a very faint band corresponding to the β-subunits of RNAP was present in the subsequent peak, it is likely that there was aggregation of the RNAP, particularly as there was visible precipitation during purification. Thus, this would impede entry of any complex formed into the gel during the EMSA. When the PcrA:RNAP complex was formed using the 3-step approach, a band was also present at the same height as that of the PcrA:Scaffold complex. This could also be attributed to aggregation as scaffold that is unable to bind to the aggregated RNAP would be free to bind to excess PcrA. Furthermore, the “washed” recombinant RNAP was less effective at binding the DNA:RNA scaffold alone compared with native RNAP, and again remained in the wells of the gel. Since aggregation could inhibit complex formation, further optimisation of the RNAP purification is essential in future work. The percentage of the TBE-PAGE gels could also be decreased to aid entry of the complexes into the gel where this was an issue.

When complexes were formed using the 3-step approach, band laddering was present above and below the scaffold. Interestingly, the bands showed a somewhat regular pattern and reduced in intensity on the addition of PcrA. The bands below the level of the scaffold alone are likely due to nuclease contamination since they must contain species which are smaller than the scaffold in order to run lower, but the cause of the laddering above the scaffold is less obvious. Since the starting material was impure, digestion of the EMSA samples with Proteinase K followed by comparison with undigested samples could test for contamination by other proteins. Considering that the laddering was only obviously present after the 3-step method was used, and lower bands disappeared on the addition of PcrA, it is unlikely that the issue arose due to impure samples, but rather the method used. However, it is plausible that PcrA could be outcompeting any protein contaminants for the scaffold, as well as RNAP. The

regular arrangement of the bands also implies that the size of the species are increasing in a consistent manner. Addition of the NTS appeared to have little effect on the laddering since the bands of the "ladder" were present at the same position in both its presence and absence. Hence, we suggest that the TS and RNA components of the scaffold are either forming regular structures or the TS alone is multimerising or folding to form secondary structures.

Consistent with the findings of Gwynn et al. (2013) and Urrutia-Irazabal et al. (2021), we again showed that the CTD of PcrA is necessary for its interactions with RNAP: no shift of a TEC by the PcrA mutant lacking its CTD was observed here using EMSAs. This result also validates the interactions observed in the EMSAs conducted here, whereby full-length PcrA was able to shift a number of different TECs.

To confirm the identity of the shifted bands in Figure 3-17, several approaches can be explored. Immediately, the lower band may be identified by forming complexes between PcrA and different combinations of scaffold components (PcrA:ssDNA, PcrA:dsDNA, PcrA:DNA/RNA) and running them alongside the native TEC:PcrA samples. To confirm that this lower band contains PcrA, and not partially degraded RNAP with retained binding activity, the sample could be combined with an anti-PcrA antibody and observed for further supershifts. However, the presence of multiple species could make interpretation difficult. The protein content of the bands could also be determined using 2D-gels, MS or by western blotting. Fluorescent labelling of the other scaffold components (RNA and NTS) would confirm their presence, location on the gel and thus, what interactions are formed. Particularly, labelled RNA would confirm correct scaffold formation from the beginning. Correct TEC scaffold formation could be confirmed by blotting using a S9.6 antibody, which detects DNA:RNA hybrids (Boguslawski et al., 1986). Furthermore, the samples both eluted from SEC and those seen in the EMSAs could be analysed by negative stain. In principle, this would allow us to determine if TEC:PcrA complexes had formed and if so, the stoichiometry of the species present and if all the encoded subunits of RNAP associated correctly to assemble the polymerase.

Although aggregation of RNAP likely played at least a partial role in limiting interaction between PcrA and recombinant RNAP, whereby the proteins mostly eluted separately from a SEC column, PcrA was able to shift the RNAP during EMSAs. PcrA was also able to shift the native material but very little complex was obtained after SEC. These results indicate that PcrA interacted with RNAP, but the interactions formed were not stable enough to remain intact on a size-exclusion column. Thus, a future direction of the work could involve cross-linking the proteins before imaging by cryo-EM, particularly as protein complexes can also face issues with dissociation and degradation during grid preparation. A number of robust methods for

cross-linking have already been developed, including the GraFix method (Stark, 2010) and more recently, the AgarFIx method (Adamus et al., 2019). Both provide a sensible advancement for this work, particularly the latter since it is more accessible and gentler, and thus should be explored in addition to further optimising the methods used here.

Other future directions involving different expression systems could also be beneficial. Here, PcrA and RNAP from the *B.subtilis* system were used in an attempt to form a PcrA:RNAP complex. However, based on pulldown experiments carried out by Gwynn et al. (2013), greater concentrations of RNAP were pulled down from both *B.subtilis* and *Geobacillus stearothermophilus* cell lysates by a *G.stearothermophilus* PcrA bait. In addition, they also obtained higher yields of PcrA from *G.stearothermophilus* compared with *B.subtilis*. Thus, it is possible that formation of a stable complex may be more probable if the thermophilic system is used instead. In addition, co-expression of PcrA and RNAP was carried out in *E.coli*. However, co-expression of PcrA and RNAP in their native environment of *B.subtilis* could also make complex formation more favourable. In particular, stimulating R-loop formation whilst simultaneously overexpressing PcrA may encourage formation of a PcrA:RNAP complex in vivo with a possible preferred substrate.

Once a stable PcrA:RNAP complex has been obtained, either by the original methods developed by Urrutia-Irazabal et al. (2021) and explored here, or by supplementation with cross-linking, we can work towards obtaining a high-resolution structure using cryo-EM. Doing so will greatly increase the knowledge surrounding the interaction between PcrA and RNAP, both structurally and functionally. Thus, we hope it will assist in our understanding of PcrA's role in R-loop biology and transcription conflicts.

3.4 Chapter 3 summary

• Combining native *B.subtilis* RNAP with an artificial DNA:RNA scaffold and recombinant *B.subtilis* PcrA in a stepwise manner led to the elution of low concentrations of a non-stoichiometric complex from a size-exclusion column.

• Combining cell lysates from individual expressions of native *B.subtilis* PcrA and recombinant *B.subtilis* RNAP failed to form a stable PcrA:RNAP complex.

• Co-expression of *B.subtilis* RNAP and *B.subtilis* PcrA in *E.coli* did not yield convincing evidence of a stable PcrA:RNAP complex and was susceptible to *E.coli* ArnA contamination.

• EMSAs showed that both native and recombinant *B.subtilis* RNAP interacted with the artificial scaffold formed both by a 1-step and 3-step method. Each RNAP:scaffold complex was then shifted on the addition of PcrA.

• Addition of PcrA to the native *B.subtilis* RNAP:scaffold complexes resulted in the formation of multiple shifted species, which currently remain unknown and require identification.

Chapter 4

Predicting interactions between PcrA and partner proteins using AlphaFold-Multimer

4.1 Introduction

As mentioned, PcrA possesses multifunctionality through its ability to interact with a collection of partner proteins. When its interactions with RNAP were analysed using HDX-MS, a 9-residue HIM, VDPETGEIL, present within the lineage-specific insertion domain (SI1) of the β-subunit was identified. This motif, which contacts the CTD of PcrA, is present in most bacterial RpoB sequences and contains a highly conserved glutamate residue present within its "TGE" triad. On the basis of site-directed mutagenesis experiments, this glutamate residue has been predicted to form a critical electrostatic interaction with K727 of PcrA (Sanders et al., 2017). Interactions between the core of PcrA (2A and 2B subdomains) and the regions surrounding the DNA and RNA exit channels of the polymerase (β and β′ subunits) were also identified in the same study (Urrutia-Irazabal et al., 2021). Furthermore, crosslinking studies with RNAP and PcrA's orthologue, UvrD, found interactions between UvrD's 2B domain and the β-subunit of RNAP (Epshtein et al., 2014).

A search of the *B.subtilis* proteome identified a further 6 proteins possessing the HIM that was discovered in RNAP. In particular, the "TGE" triad present at the tip of a loop between two β-strands was highly conserved. Only three of these proteins (UvrB, YwhK and YxaL) have had interactions with PcrA experimentally validated, whilst the other three (QueA, RplX and YtzB) remain as putative PcrA interactors. Current information on the interactions between these proteins and PcrA is summarised below.

4.1.1 UvrB

UvrB has been implicated in the recruitment of UvrD during nucleotide-excision repair (NER) and was shown to interact with UvrD's 1A domain and CTD using Y2H assays and SPR (Manelyte et al., 2009). This interaction was validated further using pulldown-coupled to MS, in which UvrB was enriched in the presence of PcrA's CTD, but showed lowered enrichment (comparable to the control) in the absence of the CTD (Sanders et al., 2017). More recently, UvrB has been shown to interact with PcrA, specifically through domain 2 (Urrutia-Irazabal et al., 2021) and contains the same conserved "TGE" triad found within the RNAP β-subunit as part of a putative HIM. In fact, a docking model predicted a stabilising interaction between K727 of PcrA and E233 of UvrB, which is the conserved glutamate residue present within the triad.

4.1.2 YwhK and YxaL

Both YwhK and YxaL are predicted to adopt β-propeller structures and have been shown to interact with PcrA through yeast two-hybrid assays (Y2H) (Noirot-Gros et al., 2002). The same study also found that YxaL stimulated helicase activity and moderately increased the

processivity of PcrA in vitro. Both also contain the conserved “TGE” triad and in fact, the putative HIM was present within the fragments of YxaL and YwhK that interacted with PcrA in the Y2H assay. However, little information regarding their functions with respect to PcrA exists. Interaction cluster identification using Y2H identified a number of partners of YwhK and YxaL, only a few of which have roles related to nucleic acid metabolism (Table 4-1) (Marchadier et al., 2011).

Table 4-1. Interaction partners of YxaL and YwhK as determined by Y2H assays (Marchadier *et al.*, 2011).

	Gene	Protein
YxaL	*pcrA*	ATP-dependant DNA helicase
	swrC	Swarming motility protein
	yabA	Initiation-control protein
	yacL	Uncharacterised PIN and TRAM-domain containing protein
	yclM	Aspartokinase 3
	yhcQ	Spore coat protein
	yomI	Uncharacterised transglycosylase
	pdp	Pyrimidine-nucleoside phosphorylase
	polA	DNA Polymerase I
	queE	7-carboxy-7-deazaguanine synthase
YwhK	*pcrA*	ATP-dependant DNA helicase
	dnaE	DNA Primase
	xseA	Exodeoxyribonuclease

4.1.3 QueA, RplX and YtzB

RplX is a 50S ribosomal protein predicted to nucleate assembly of the 50S subunit (Herold and Nierhaus, 1987). Evidence already exists for interaction between PcrA and the ribosome, whereby the ribosomal protein L3 was shown to co-purify with the helicase and stimulate its activity (Soultanas et al., 1998). In addition, the ribosomal protein L14 has been shown to enhance the processivity of *E.coli* Rep, which is a homologue of PcrA (Yancey and Matson, 1991). Thus, contacts between PcrA and RplX would provide further support for the helicases interaction with the ribosome. QueA is a transferase involved in synthesis of the modified tRNA nucleoside, queuosine (Grimm et al., 2006), and no current links have been made between PcrA, its homologues and QueA. YtzB is a currently uncharacterised protein and there is currently no published data on its function.

To gain further insights into the interfaces between PcrA and its partner proteins, we used an adaptation of the AI-driven protein structure prediction program, AlphaFold2 (referred to AlphaFold in the remainder of the text), known as AlphaFold-Multimer. AlphaFold has

outcompeted its rivals on reliability, including RoseTTAFold, on a number of occasions (Kryshtafovych et al., 2021, Bryant et al., 2022, Lee et al., 2022). It has also replaced a number of traditional methods for predicting protein interactions, such as docking (Dominguez et al., 2003, Chen et al., 2003, Vakser, 2014, Bryant et al., 2022) and template-based methods (Chen and Skolnick, 2008, Keskin et al., 2008, Mukherjee and Zhang, 2011), which are limited by force-field accuracy and the availability of experimentally determined structures respectively. Thus, it remains one of the best accessible options for computational protein structure and complex prediction. Furthermore, the development of notebook-based software, such as ColabFold, makes AlphaFold more accessible and user-friendly to many in the scientific community whose knowledge on structural biology may be limited (Mirdita et al., 2022).

AlphaFold ranks each output structure based on the confidence of its prediction, which involves consideration of several metrics. The first of these is its pLDDT (predicted Local Distance Difference Test): a per-residue estimate of its confidence, whereby scores >90 indicate high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and scores <50 indicate very low confidence. Further metrics assist in determining the confidence of the given structure by evaluating the position of residues within and between protein domains. The “Predicted Aligned Error” (PAE) predicts the alignment errors between all pairs of residues and thus, provides confidence of global protein structure, including the folding of domains and their relative positions to one another. This in turn can be used to evaluate protein-protein contacts. The PAE is presented in the form of a heatmap. Here, dark green indicates high confidence, light green indicates some confidence and white indicates low confidence. For each protein:protein interaction pair we analyse below, only the highest ranked output is presented and discussed.

Table 4-2 summarises which of PcrA’s putative and know protein interactors have already been structurally determined, and which rely solely on AlphaFold modelling in this work.

Table 4-2. Structural determination of PcrA's partner proteins containing the helicase interaction motif.

Protein	Method(s) of Structural Determination	PDB
PcrA	X-ray Crystallography	3PJR
	AlphaFold Modelling	-
RNA Polymerase	Cryo-EM	6WVJ
	AlphaFold Modelling	-
UvrB	X-ray Crystallography	3V4R
	AlphaFold Modelling	-
YxaL	X-ray Crystallography	7DXN
	AlphaFold Modelling	-
YwhK	AlphaFold Modelling	-
YtzB	AlphaFold Modelling	-
RplX	AlphaFold Modelling	-
QueA	X-ray Crystallography	1YY3
	AlphaFold Modelling	-

4.1.4 Aim of the chapter

1. Assess AlphaFold-Multimer predictions of protein:protein interactions between PcrA and its partner proteins containing a HIM

4.2 Results

First, we studied interaction between PcrA and the four proteins that contain HIMs and have experimental evidence supporting their interaction with PcrA.

4.2.1 AlphaFold-Multimer analysis of PcrA:RNAP interactions

A limitation of AlphaFold is the submission limit of ~1400 residues. Thus, it is not yet possible to predict interactions within larger protein complexes, such as PcrA:RNAP, using complete sequences. Therefore, different domains within PcrA and RNAP that have shown evidence for interaction via experimental approaches were tested in separate submissions.

4.2.1.1 SI1 domain interaction with the PcrA CTD

As shown in Figure 4-1A, PcrA's CTD was predicted with higher overall confidence, according to the pLDDT (Figure 4-1B), compared with the SI1 domain, which is only predicted with moderate confidence towards the N-terminal region. Within this region is the loop containing the "TGE" triad (Figure 4-1A, black circle). The β-strand of PcrA containing K727 is also predicted with high confidence. However, the loops towards the centre and C-terminal region of the SI1 domain (Figure 4-1A, green asterisks), are predicted with low confidence. There is also very high error in their positions relative to the rest of the SI1 domain (Figure 4-1C, white colouring in SI1 square). This is likely due to their flexibility, and thus their relative positioning and structure should be interpreted with caution. Fortunately, this is mostly irrelevant to the interaction of interest.

The position of the loop containing the "TGE" triad relative to the β-strand containing K727 is predicted with low error, indicated by the dark green colouring within the red boxes of Figure 4-1C. When contacts (≤4 Å) between the loop and β-strand were analysed, all those present were found to be predicted with high confidence (Figure 4-1D). As a result, PDBePISA (Krissinel and Henrick, 2007) was used to assess the chemical nature of these interactions, and it identified a salt bridge between K727 of PcrA and E301 of the β-subunit (Figure 4-1E). Contacts were also predicted between the CTD and other residues in the HIM (Table 4-2). Residues H696 and K727 show high conservation whilst residues R728 and L730 show moderate conservation (Figure 4-2).

Whilst the individual structures were not modelled with complete confidence, the low error prediction of the of the HIM positioning relative to the interacting β-strand of PcrA, along with the high local structural confidence according to the pLDDT, provides confidence the model accurately mimics the true interaction. This is further supported by the experimental evidence from site-directed mutagenesis experiments undertaken by Sanders et al. (2017).

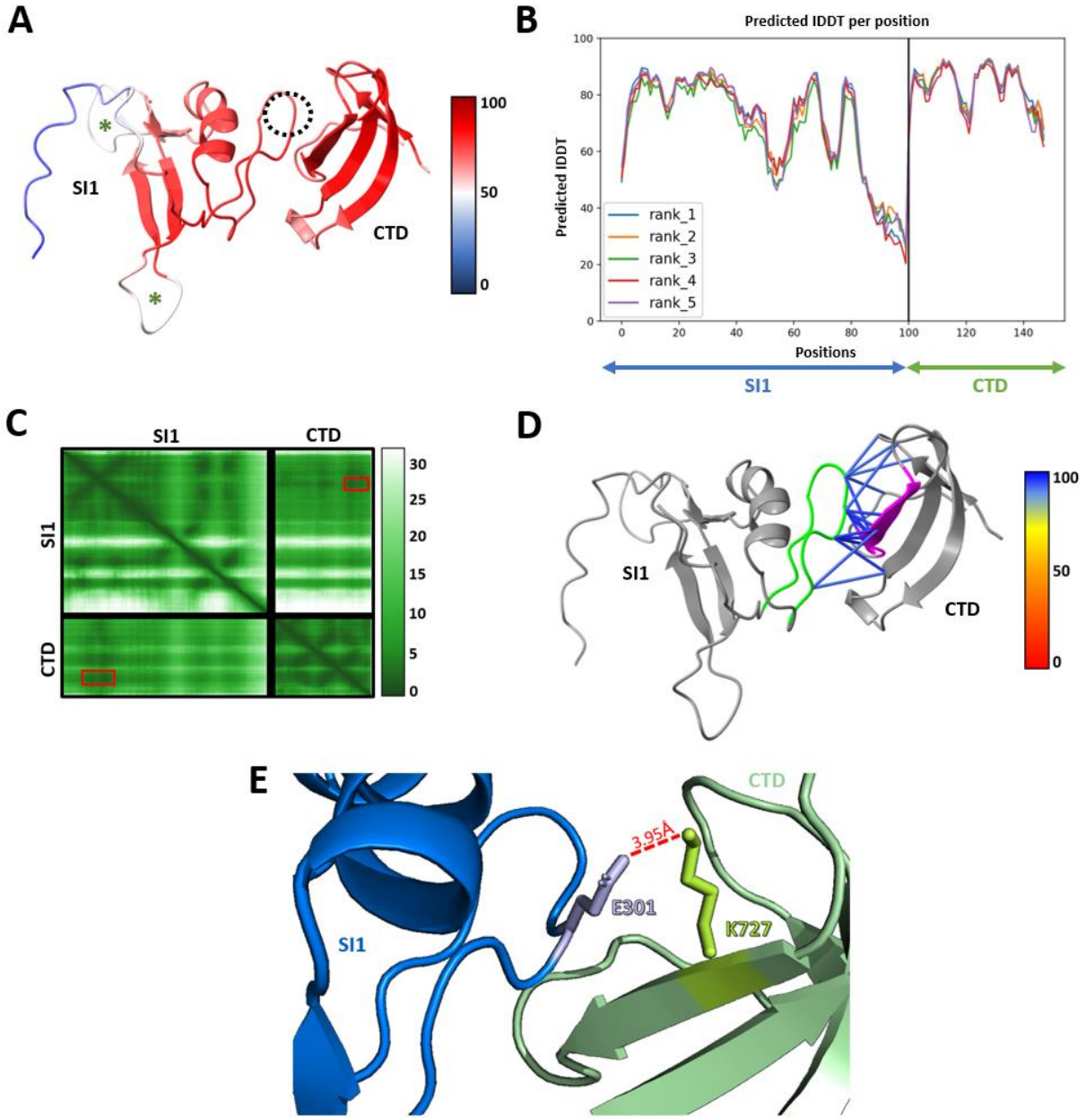

Figure 4-1. RNAP SI1 and PcrA CTD interaction prediction by AlphaFold-Multimer. (A) Confidence of prediction according to the pLDDT. pLDDT>90 indicates high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and pLDDT<50 indicates very low confidence. Region within the black dashed circle represents the expected location of interaction within the SI1 domain. Green asterisks represent regions of low confidence. (B) Graph showing the predicted IDDT per position for all 5 models produced by AlphaFold-Multimer. Rank 1 is represented in blue. (C) PAE plot for the given structure. Regions corresponding to the expected interaction are highlighted by red boxes. Error values are given in Angstroms, where 0 indicates low error and 30 indicates high error. (D) Predicted contacts within 4 Å according to the PAE. Identified contacts are coloured by confidence using the same values as the pLDDT. The loop region containing the putative helicase interaction motif is coloured in green and the β-strand containing K727 of PcrA is coloured in magenta. (E) Salt bridge (red dashed line) of 3.95 Å predicted between E301 of the RNAP β-subunit and K727 of PcrA.

Table 4-3. Interactions between the SI1 domain of RNAP and the CTD of PcrA as predicted by PDBePISA.

Bond Type	β-Subunit Residue	PcrA Residue	Distance (Å)
Salt Bridge	E301	K727	3.95
Salt Bridge	E305	R728	2.54
Hydrogen Bond	T299	H696	2.52
Hydrogen Bond	G300	L730	3.25
Hydrogen Bond	I302	R728	2.69

Figure 4-2. PcrA CTD Weblogo. Weblogo representation of PcrA's CTD multiple sequence alignment conducted using 250 PcrA homologues. Residues predicted to be involved in the PcrA:RNAP interaction are indicated by their residue number.

4.2.1.2 β/β′ subunit interaction with the PcrA 2A domain

(i) PcrA 2A domain & β-subunit interaction

As shown in Figure 4-3A, the structures of both the β-subunit and the 2A domain are mostly predicted with moderate to high confidence according to their respective pLDDT scores, aside from the C-terminal region and two loops of the β-subunit (Figure 4-3A, green asterisks). These low-confidence regions also correspond to the white colouring within the largest square of the PAE plot (Figure 4-3C), meaning that there is also high error in their positioning within the structure. On the other hand, the global structure of the 2A domain is mostly predicted with high confidence (Figure 4-3C).

It was immediately obvious from the model that there were no predicted interactions between the 2A domain and the β-subunit. When contacts of up to 8 Å were displayed on the model, only those present within the individual structures were visible, mostly with moderate to high confidence (Figure 4-3D). Since the confidence of the contacts is interpreted from the PAE, it was unsurprising that there was low confidence in those formed between the loop regions of the β-subunit identified in Figure 4-3A, and the rest of its structure. In fact, when the predicted structure is aligned to an experimentally determined structure of the β-subunit (PDB: 6WVJ), the structure is almost identical, aside from the low-confidence regions produced by the AlphaFold-Multimer prediction (Figure 4-3E). The areas of the β-subunit predicted to interact with PcrA experimentally are directly adjacent to the regions of low confidence (Figure 4-3A, black circles), and only have limited confidence themselves. Therefore, this could explain why

the model does not show any interaction between the β-subunit and the 2A domain. Moreover, since the regions of interaction within the β-subunit are present close to where it meets the β′-subunit within RNAP, they are likely to adopt a different conformation in the complete structure. It is possible that PcrA may only interact in silico when the β/β′ interface is correctly modelled. The lack of any predicted interaction is also highlighted by the high-error prediction of the relative position of 2A to the β-subunit (white boxes in Figure 4-3C).

(ii) PcrA 2A domain & β′-subunit interaction

As shown in Figure 4-4A, the β′-subunit is predicted with mostly moderate to high confidence according to the pLDDT (Figure 4-4B), including the regions which showed protection in HDX-MS (black circles). According to the PAE, however, the global structure is not predicted with high confidence (Figure 4-4C). As expected, the 2A domain was predicted with identical confidence, both at the local and global levels, to that of the 2A/β-subunit interaction prediction (Figures 4-4B and 4-4C).

As with the 2A/β-subunit interaction prediction, no contacts were predicted between the 2A domain and the β′-subunit (Figures 4-4C and 4-4D). To determine if this could be due to an incorrect structural prediction of the putative interaction regions within the β′-subunit, the model was aligned to the same cryo-EM RNAP structure used in Figure 4-4E. Unlike the β-subunit, the predicted structure of the β′-subunit was identical to that of the cryo-EM obtained structure (Figure 4-4E). Thus, lack of interaction is again reflected by the high-error prediction of the position of 2A relative to the β′-subunit.

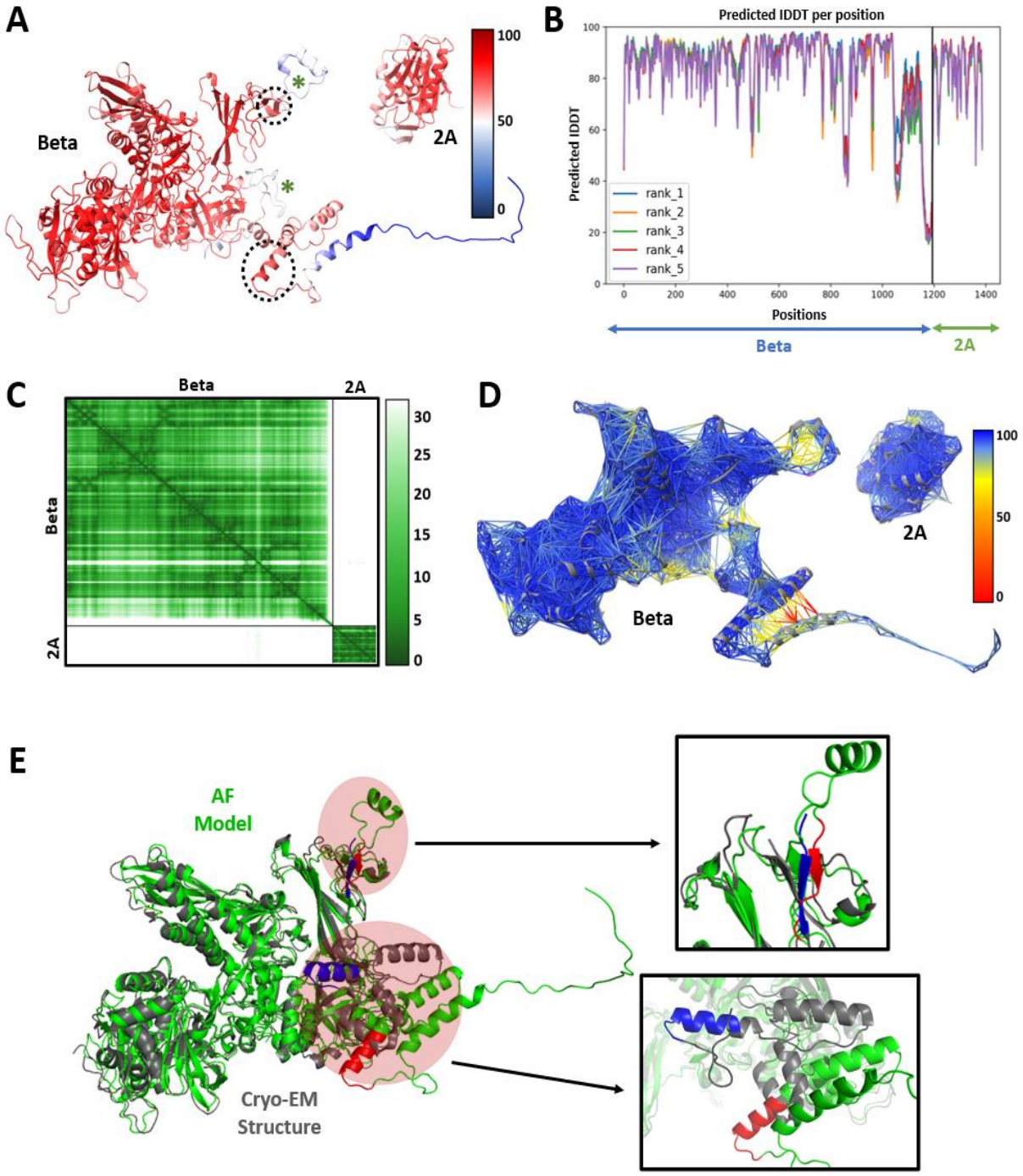

Figure 4-3. RNAP β-subunit and PcrA 2A domain interaction prediction by AlphaFold-Multimer. (A) Confidence of prediction according to the pLDDT. pLDDT>90 indicates high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and pLDDT<50 indicates very low confidence. Regions within the black dashed circles represent the expected locations of interaction within the β-subunit. Green asterisks represent regions of low confidence. (B) Graph showing the predicted IDDT per position for all 5 models produced by AlphaFold-Multimer. Rank 1 is represented in blue. (C) PAE plot for the given structure. Error values are given in Angstroms, where 0 indicates low error and 30 indicates high error. (D) Predicted contacts within 8 Å according to the PAE. Identified contacts are coloured by confidence using the same values as the pLDDT. (E) Alignment of the AlphaFold-Multimer predicted β-subunit structure (green) to the structure of the β-subunit obtained by cryo-EM (grey) (PDB: 6WVJ).

Figure 4-3 continued. Red circles indicate regions of contact with PcrA. Images within black squares are enlarged views of the contact regions. The predicted regions of contact (red) do not align to the cryo-EM structure (blue).

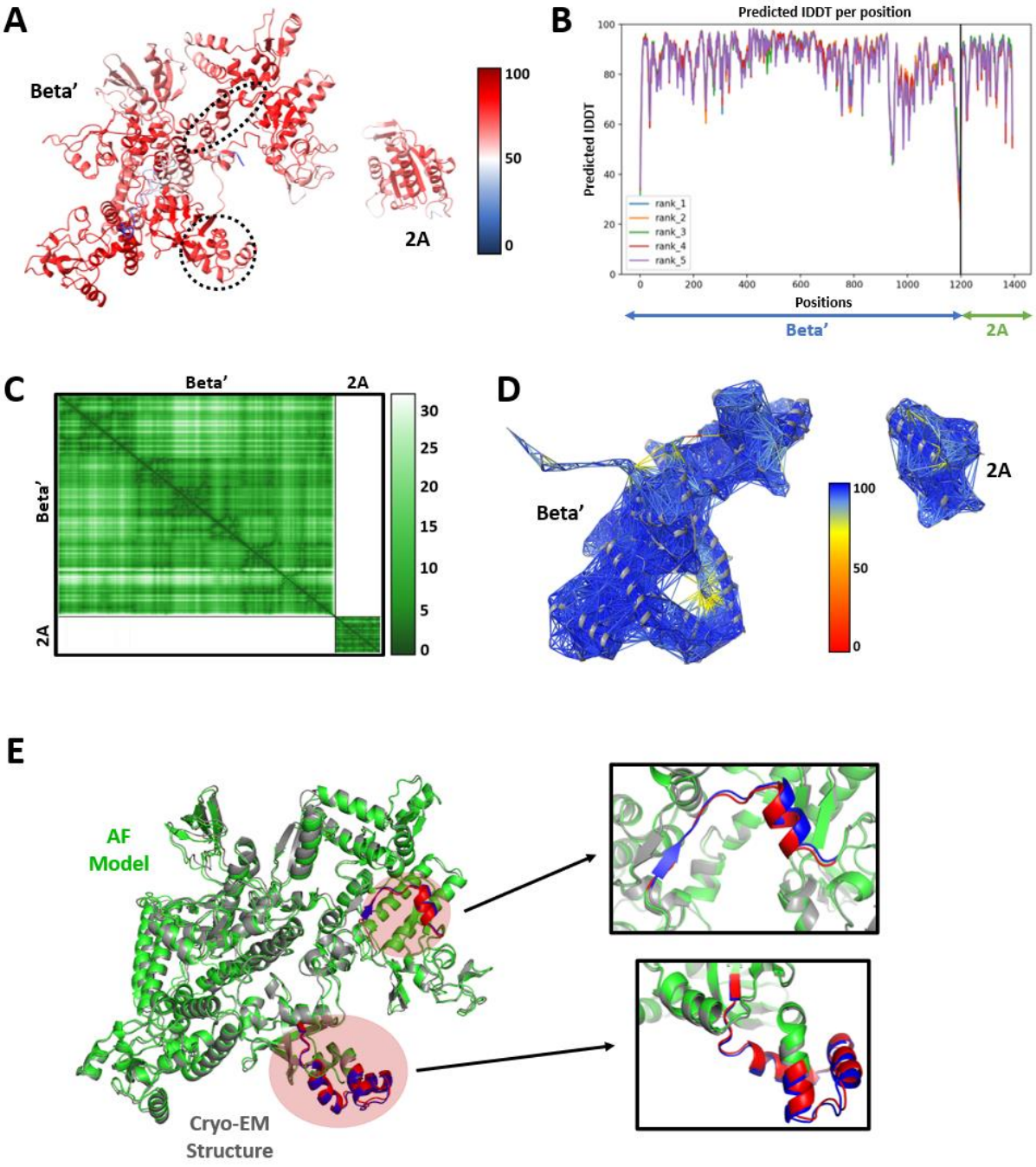

Figure 4-4. RNAP β'-subunit and PcrA 2A domain interaction prediction by AlphaFold-Multimer. (A) Confidence of prediction according to the pLDDT. pLDDT>90 indicates high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and pLDDT<50 indicates very low confidence. Regions within the black dashed circles represent the expected locations of interaction within the β'-subunit. (B) Graph showing the predicted IDDT per position for all 5 models produced by AlphaFold-Multimer. Rank 1 is represented in blue. (C) PAE plot for the given structure. Error values are given in Angstroms, where 0 indicates low error and 30 indicates high error. (D) Predicted contacts within 8 Å according to the PAE. Identified contacts

Figure 4-4 continued. are coloured by confidence using the same values as the pLDDT. Red circles indicate regions of contact with PcrA. Images within black squares are enlarged views of the contact regions. The predicted regions of contact (red) do not align to the cryo-EM structure (blue). (E) Alignment of the AlphaFold-predicted β'-subunit structure (green) to the structure of the β'-subunit obtained by cryo-EM (grey)(PDB: 6WVJ). Red circles indicate regions of contact with PcrA. Images within black squares are enlarged views of the contact regions. The predicted regions of contact (red) directly align to the cryo-EM structure (blue).

4.2.1.3 β/β′ subunit interaction with the PcrA 2B domain

(i) PcrA 2B domain & β-subunit interaction

As shown in Figure 4-5A, the structures of both the β-subunit and the 2B domain are mostly predicted with moderate to high confidence according to their respective pLDDT scores, aside from the C-terminal region and two loops of the β-subunit (Figure 4-5A, green asterisks). Interestingly, the β-subunit C-terminal region was predicted with higher error than when its interaction with the 2A domain was tested (compare Figure 4-5B). These low-confidence regions also correspond to the white colouring within the largest square of the PAE plot (Figure 4-5C). Conversely, the global structure of the 2B domain is mostly predicted with confidence (Figure 4-5C).

AlphaFold-Multimer predicted interactions of ≤4 Å between the C-terminal region of the 2B domain and the region between V1098-P1120 of the β-subunit (Figure 4-5D). These contacts were predicted with very low confidence and did not involve the regions of the β-subunit which showed protection in HDX-MS experiements. However, the residues involved are directly adjacent to these regions and are present on the surface of RNAP in the full structure. As with the 2A/β-subunit interaction model, the β-subunit here was aligned to the same cryo-EM structure of RNAP. Similarly, the structures are mostly identical, aside from the low confidence regions shown in Figure 4-5A, which include the putative interacting regions.

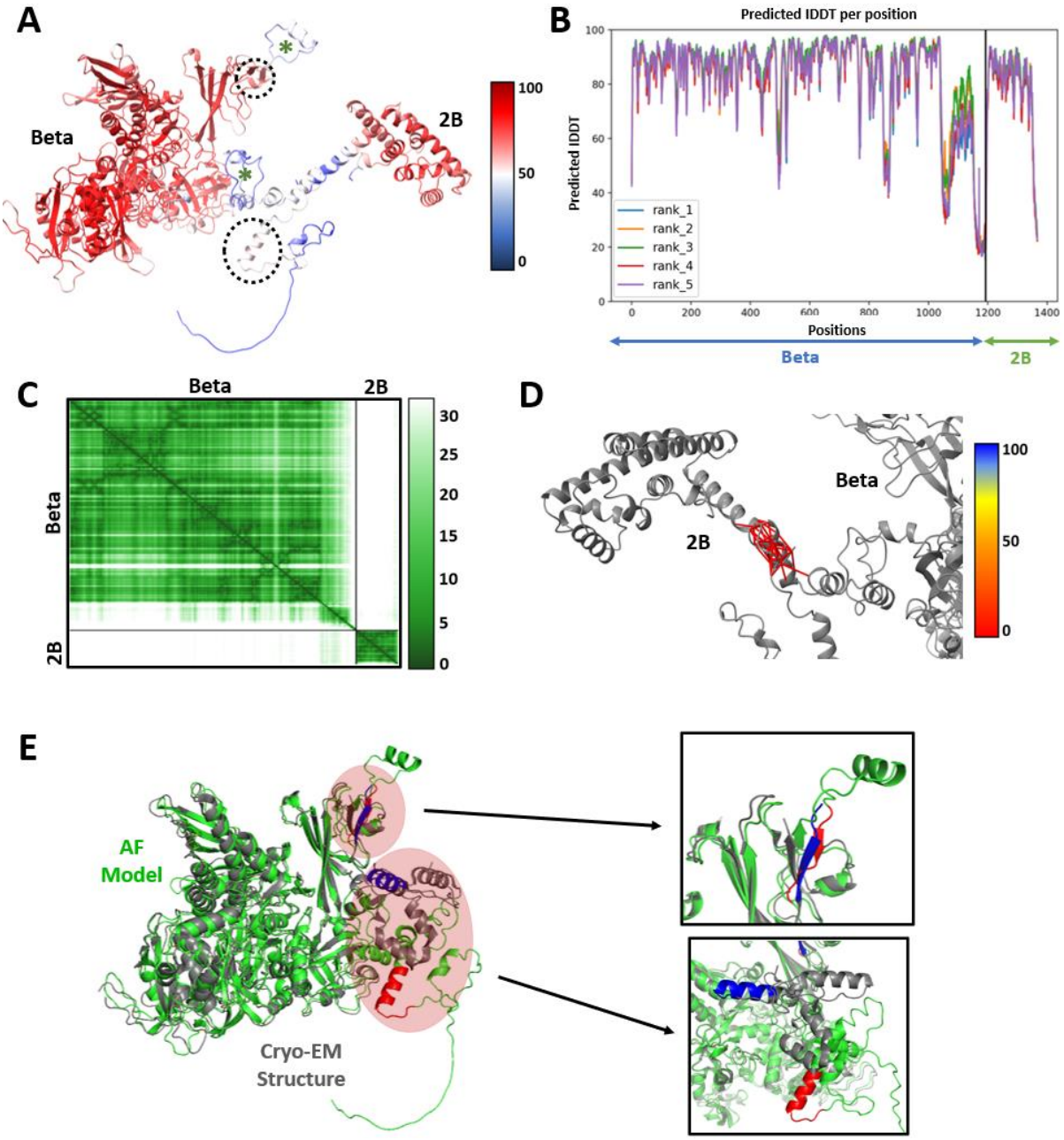

Figure 4-5. RNAP β-subunit and PcrA 2B domain interaction prediction by AlphaFold-Multimer. (A) Confidence of prediction according to the pLDDT. pLDDT>90 indicates high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and pLDDT<50 indicates very low confidence. Regions within the black dashed circles represent the expected locations of interaction within the β-subunit. (B) Graph showing the predicted IDDT per position for all 5 models produced by AlphaFold-Multimer. Rank 1 is represented in blue. (C) PAE plot for the given structure. Error values are given in Angstroms, where 0 indicates low error and 30 indicates high error. (D) Predicted contacts within 4 Å according to the PAE. Identified contacts are coloured by confidence using the same values as the pLDDT. (E) Alignment of the AlphaFold-predicted β-subunit structure (green) to the structure of the β-subunit obtained by cryo-EM (grey)(PDB: 6WVJ). Red circles indicate regions of contact with PcrA. Images within black squares are enlarged views of the contact regions. The predicted regions of contact (red) do not align to the cryo-EM structure (blue).

(ii) PcrA 2B domain & β′-subunit interaction

As shown in Figure 4-6A, the β′-subunit is predicted with mostly moderate to high confidence according to the pLDDT (Figure 4-6B), particularly the regions which showed protection in HDX-MS (black circles). According to the PAE, however, the global structure is not predicted with high confidence (Figure 4-6C). Unsurprisingly, the 2B domain was predicted with identical confidence, both at the local and global levels, to that of the 2B/β-subunit interaction prediction (Figures 4-6B and 4-6C).

Unlike the 2B/β-subunit interaction prediction, no contacts were formed between the 2B domain and the β′-subunit (Figure 4-6D). The high-error prediction reflects the uncertainty in the positioning of the 2B domain relative to the β-subunit (Figure 4-6C, white boxes). To determine if this could be due to an incorrect structural prediction of the putative interaction regions within the β′-subunit, the model was aligned to the same cryo-EM RNAP structure used previously. Unlike the β-subunit, the predicted structure of the β′-subunit was again identical to that of the cryo-EM obtained structure (Figure 4-6E). Due to the high error of the positioning of the 2B domain relative to the β′-subunit, the AlphaFold-Multimer model cannot be trusted.

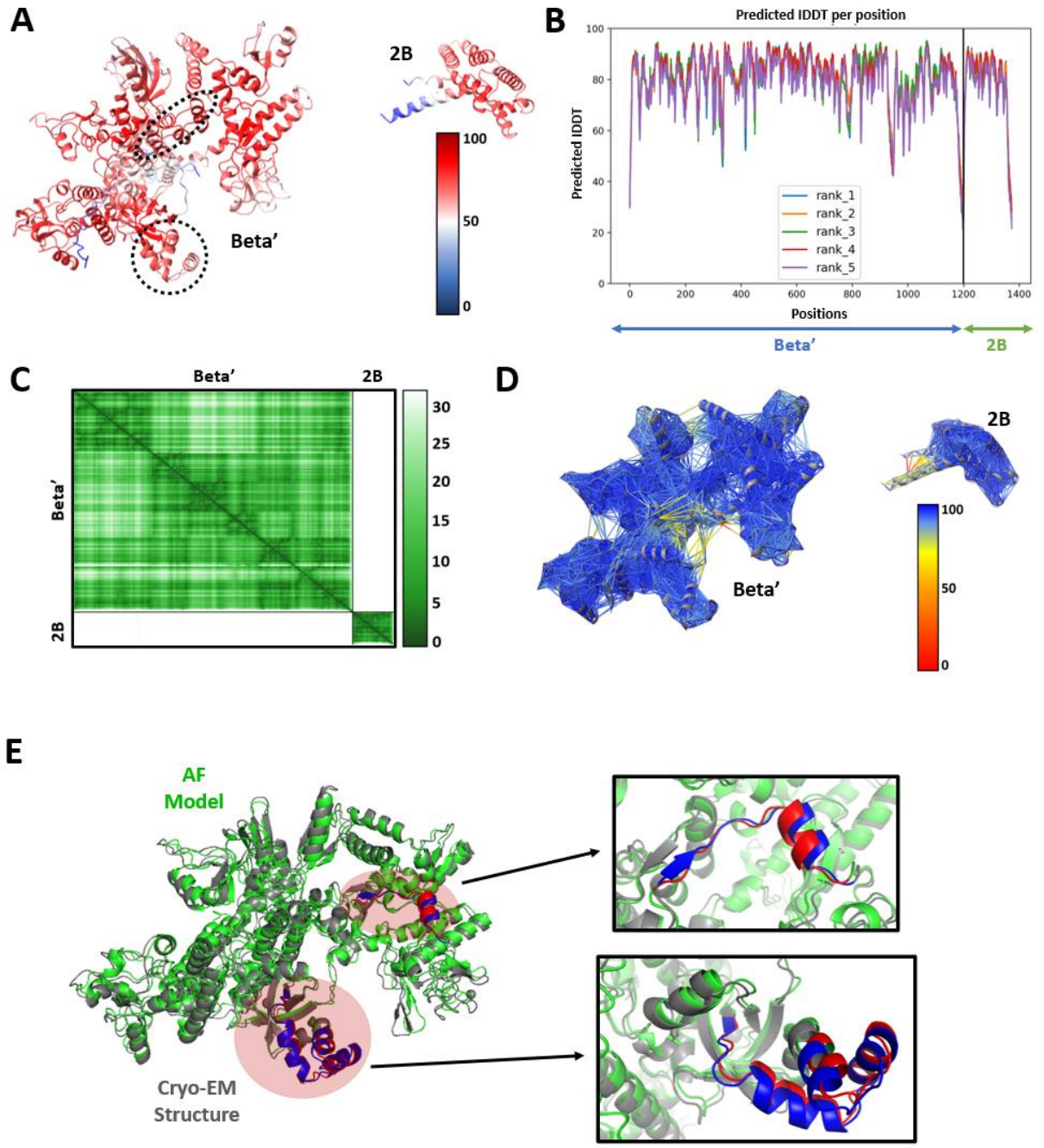

Figure 4-6. RNAP β'-subunit and PcrA 2B domain interaction prediction by AlphaFold-Multimer. (A) Confidence of prediction according to the pLDDT. pLDDT>90 indicates high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and pLDDT<50 indicates very low confidence. Regions within the black dashed circles represent the expected locations of interaction within the β'-subunit. (B) Graph showing the predicted IDDT per position for all 5 models produced by AlphaFold-Multimer. Rank 1 is represented in blue. (C) PAE plot for the given structure. Error values are given in Angstroms, where 0 indicates low error and 30 indicates high error. (D) Predicted contacts within 8 Å according to the PAE. Identified contacts are coloured by confidence using the same values as the pLDDT. (E) Alignment of the AlphaFold-predicted β'-subunit structure (green) to the structure of the β'-subunit obtained by cryo-EM (grey)(PDB: 6wvj). Red circles indicate regions of contact with PcrA. Images within black squares are enlarged views of the contact regions. The predicted regions of contact (red) directly align to the cryo-EM structure (blue).

4.2.1.4 Interaction between the RNAP β/β′ interface and PcrA 2A/2B/CTD

Since the 2B subdomain of PcrA is inserted inside the 2A domain, and they both show protection in HDX-MS experiments, its possible neither were predicted to have confident interactions with either β-subunit of RNAP because the presence of both is required for the correct conformation and interaction. It has also been proposed that the interaction between PcrA's CTD and RNAP recruits the helicase to the polymerase (Urrutia-Irazabal et al., 2021), and so interaction may be very weak in the absence of the CTD. Furthermore, the regions of the β and β′ subunits which showed protection in HDX-MS experiments are close to the interface of the subunits. As seen in Figures 4-3E and 4-5E, the β-subunit did not completely align to the cryo-EM structure of RNAP. Thus, the β-subunit may not adopt the correct fold in the absence of the β′-subunit. To address this, the first 422 residues of the β′-subunit and the last 406 residues of the β-subunit were tested for interaction with the 2A, 2B and CTD of PcrA.

As shown in Figure 4-7A, the β-subunit was overall predicted with moderate to low confidence, whilst the β′-subunit was predicted with overall moderate to high confidence, aside from at its C-terminus (Figure 4-7B). The 2A and 2B subdomains of PcrA were predicted with moderate to high confidence, whilst the CTD was predicted with low confidence. As shown in Figure 4-7C, the individual subunits and domains were predicted with some confidence according to PAE values, but their positionings relative to one another were predicted with lower confidence, indicated by the white squares. This was particularly evident for the relative positioning of the CTD to the 2A/2B domains, which is unsurprising given that the CTD is joined to the main body of PcrA via a flexible linker. As a result, this could cause issues during prediction of the interaction surfaces.

Unfortunately, the model did not predict any interaction between the β subunits of RNAP and the 2A/2B/CTD region of PcrA (Figure4-7D). When contacts within 8 Å were displayed on the model, the high-error predictions for global structure between the individual domains and subunits became more obvious (Figure 4-7D, red and yellow lines). To test these low confidence predictions, the β subunits were aligned to the cryo-EM structure of RNAP, and the 2A/2B/CTD section of PcrA was aligned to a homology model of PcrA (model from PDB: 3PJR). The predicted structure of the combined β subunits was almost identical to that of the cryo-EM structure (Figure 4-7E). However, the predicted structure of the 2A/2B/CTD region was only similar in the 2B domain of the homology model. Thus, it is possible that this contributed to the lack of interaction. The uncertainty and the low-confidence prediction in terms of PAE also provides reasoning for the absence of interaction between the β subunits and the 2A/2B/CTD structure.

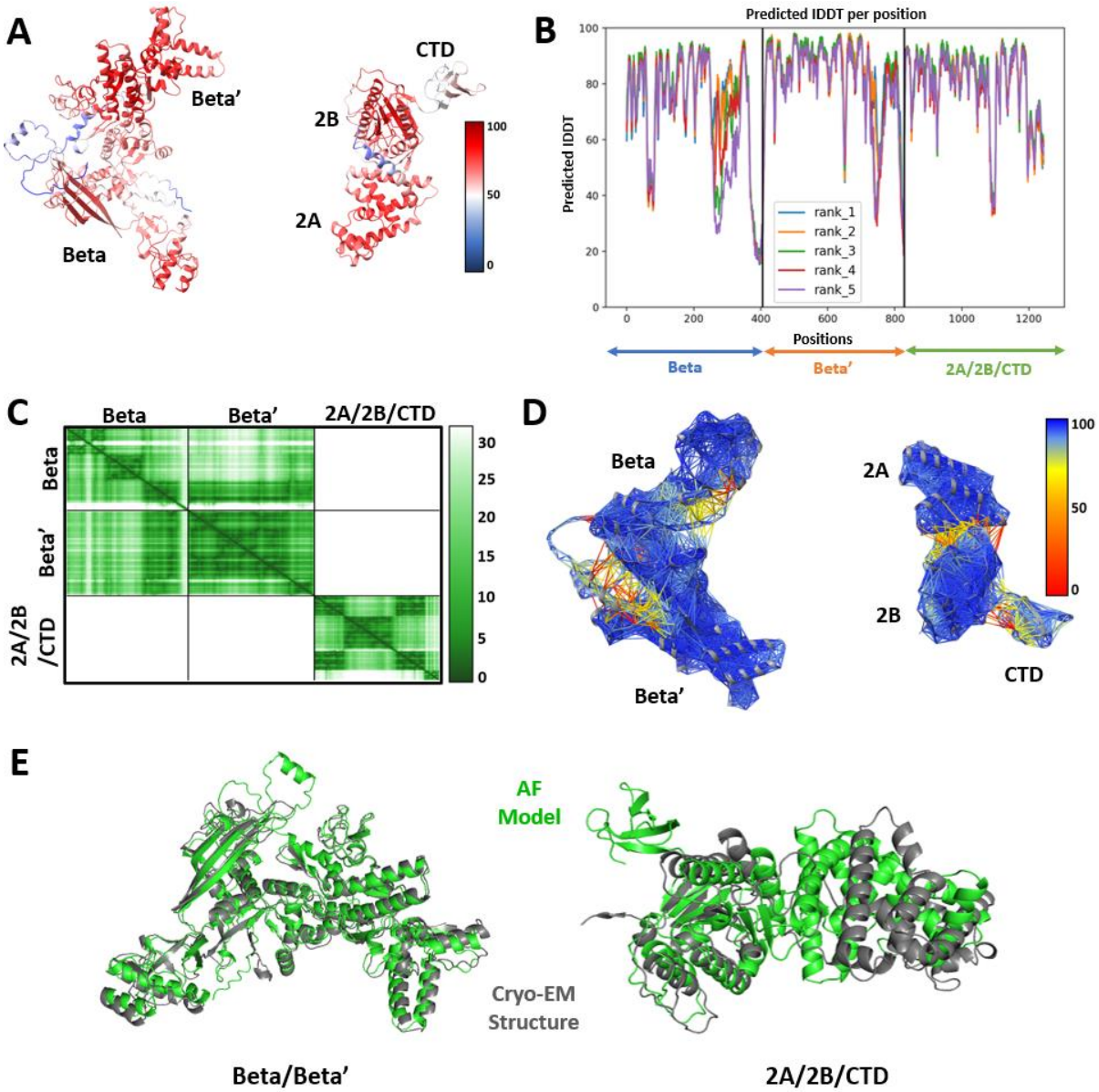

Figure 4-7. RNAP β/β' and PcrA 2A/2B/CTD interaction prediction by AlphaFold-Multimer. (A) Confidence of prediction according to the pLDDT. pLDDT>90 indicates high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and pLDDT<50 indicates very low confidence. (B) Graph showing the predicted IDDT per position for all 5 models produced by AlphaFold. Rank 1 is represented in blue. (C) PAE plot for the given structure. Error values are given in Angstroms, where 0 indicates low error and 30 indicates high error. (D) Predicted contacts within 8 Å according to the PAE. Identified contacts are coloured by confidence using the same values as the pLDDT. (E) Alignment of the AlphaFold-predicted β/β′ structure (green) to the structure of the β subunits obtained by cryo-EM (grey)(PDB: 6WVJ) and alignment of the AlphaFold-predicted 2A/2B/CTD PcrA domain structure to the structure of the 2A/2B domains of PcrA obtained by X-ray Crystallography (grey)(PDB: 3PJR).

4.2.2 AlphaFold-Multimer analysis of PcrA:UvrB interactions

4.2.2.1 UvrB interaction with the PcrA CTD

As shown in Figure 4-8A, the structure of UvrB is mostly predicted with high confidence, according to pLDDT (Figure 4-8B). However, PcrA's CTD is predicted with low confidence, despite adopting the same fold as the high confidence prediction seen in its interaction with the SI1 domain of RNAP (Figure 4-8B). Despite this, both UvrB and PcrA's CTD global structures were predicted with relatively low error (Figure 4-8C). However, there was low confidence in their positionings relative to one another. Therefore, it was not surprising that only two low confidence contacts were found when interactions up to 4 Å were displayed on the model (Figure 4-8D). These interactions did not involve the putative HIM, which contains the E233 residue that is involved with the PcrA/UvrB interaction. In fact, the HIM is present in a domain of UvrB which is not close to the predicted interface (Figure 4-8E). When the predicted UvrB structure was aligned to a structure obtained by x-ray crystallography, they were found to adopt an almost identical fold (Figure 4-8F) (Webster et al., 2012). However, there were some differences in domain 2, which contains the putative HIM, but these were only marginal.

4.2.2.2 UvrB domain 2 interaction with the PcrA CTD

Overall, domain 2 was predicted with high confidence, aside from at its termini and a loop region between residues 52-56, according to the pLDDT scores (Figure 4-9, A and B). The CTD of PcrA was predicted with less confidence, particularly at its termini and between residues 31-36 (125-130 in Figure 4-9B). However, its structure was identical to the predicted CTD's from other tested interactions, which showed higher confidence. Both domains were also predicted with mostly low error in terms of their global structures (Figure 4-9C).

Although the model shows interaction between domain 2 and the CTD, the contacts within 4 Å are predicted with mostly high error and so there is little confidence in the relative position and orientation of the domains (Figure 4-9, C and D). In addition, the interaction between K727 of PcrA and E233 of UvrB was not predicted. In fact, the HIM is present on the side of domain 2 that is not in contact with the CTD (Figure 4-9E). Whilst some structural relevance can be applied to the prediction of the individual domains, the significance of the predicted interactions is unclear.

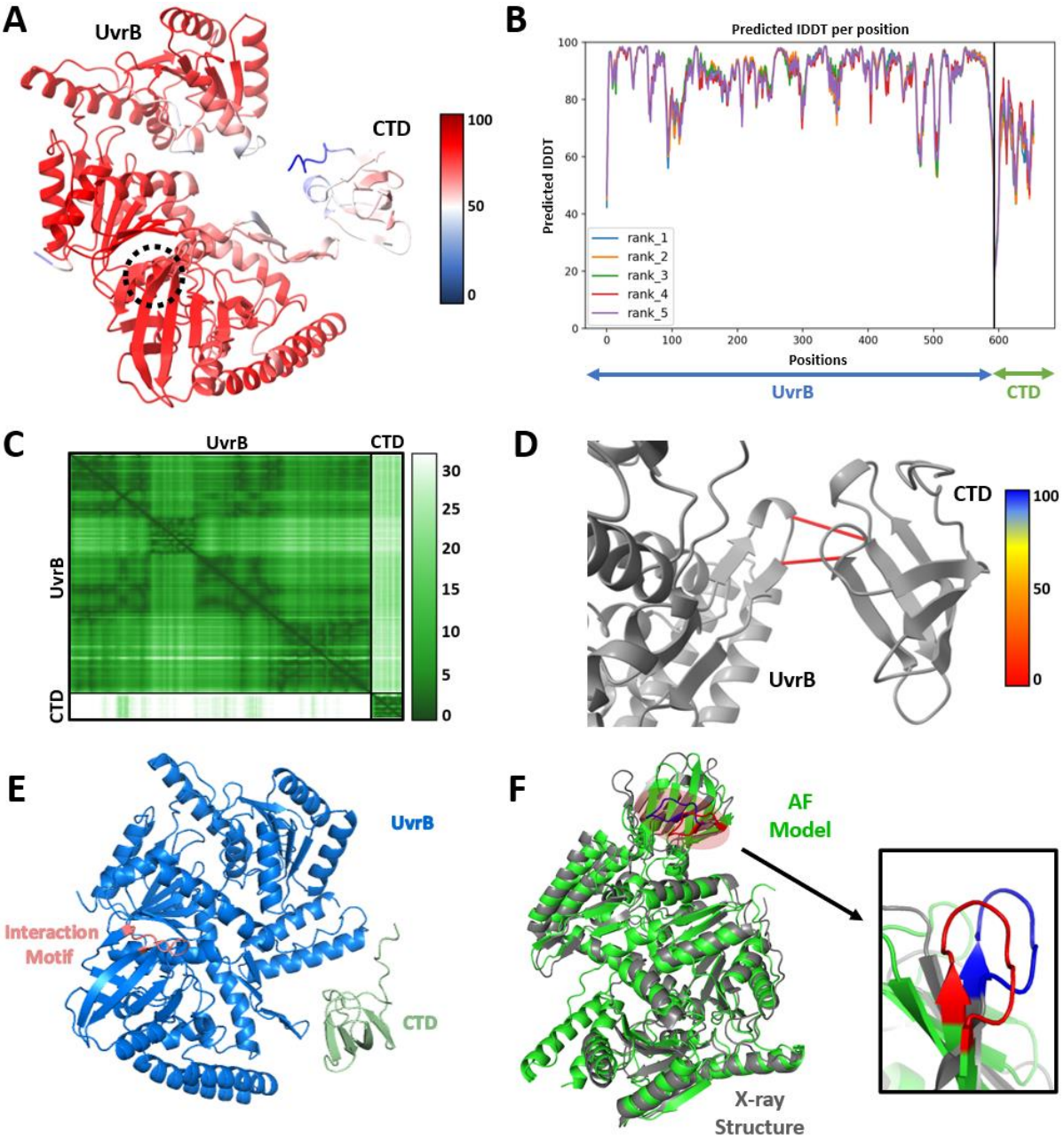

Figure 4-8. UvrB and PcrA CTD interaction prediction by AlphaFold-Multimer. (A) Confidence of prediction according to the pLDDT. pLDDT>90 indicates high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and pLDDT<50 indicates very low confidence. Region within the black dashed circle represents the expected location of interaction within UvrB. (B) Graph showing the predicted IDDT per position for all 5 models produced by AlphaFold-Multimer. Rank 1 is represented in blue. (C) PAE plot for the given structure. Error values are given in Angstroms, where 0 indicates low error and 30 indicates high error. (D) Predicted contacts within 4 Å according to the PAE. Identified contacts are coloured by confidence using the same values as the pLDDT. (E) The helicase interaction motif is present on the opposite side of UvrB showing interaction with the CTD. (F) Alignment of the AlphaFold-Multimer predicted UvrB structure (green) to the structure of UvrB obtained by X-ray crystallography (grey) (PDB: 3V4R). The red circle indicates the region containing the putative HIM. The image within the black square is an enlarged view of the putative HIMs structure. The predicted regions of contact (red) do not directly align to the X-ray structure (blue).

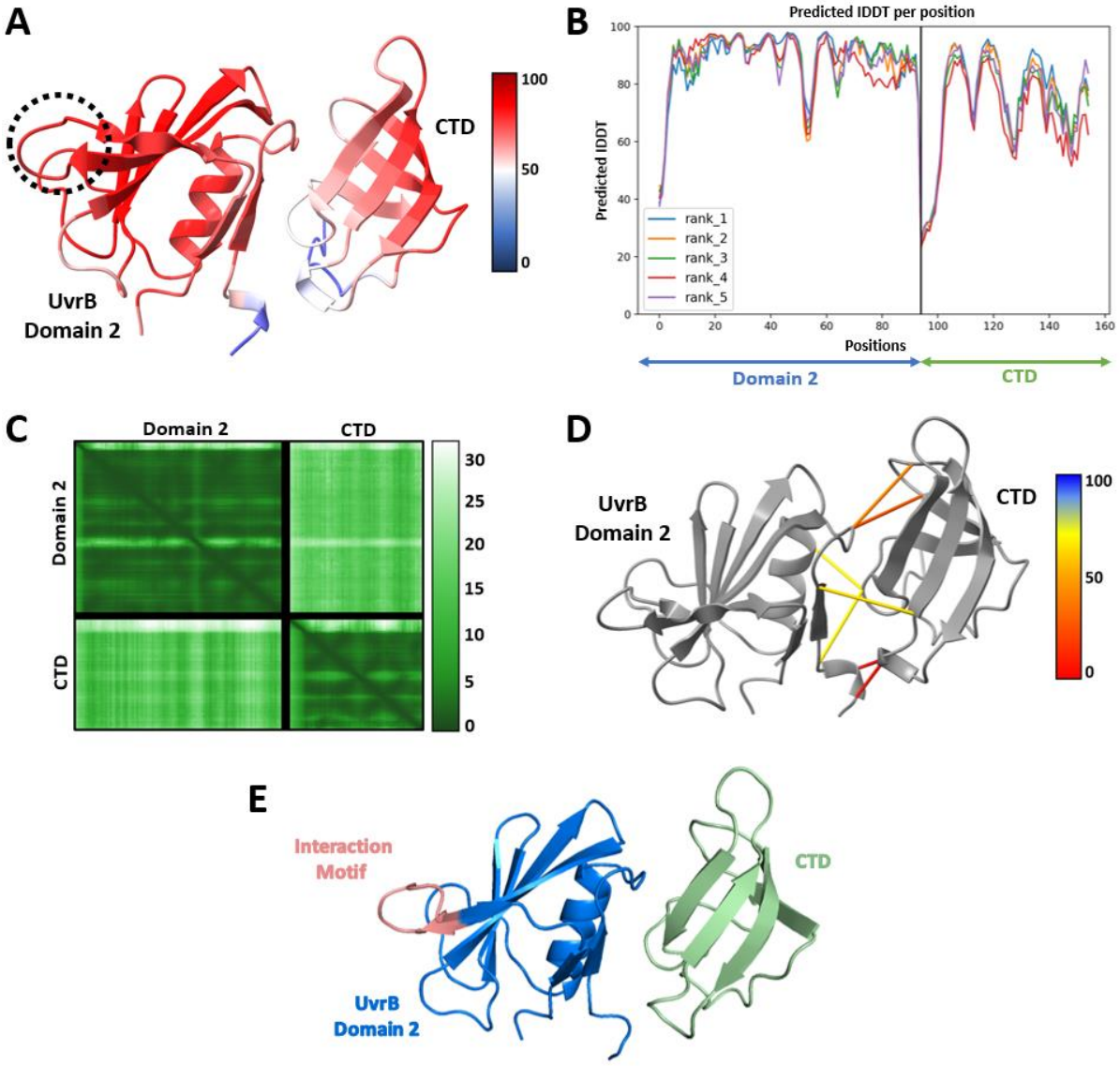

Figure 4-9. UvrB domain 2 and PcrA CTD interaction prediction by AlphaFold-Multimer. (A) Confidence of prediction according to the pLDDT. pLDDT>90 indicates high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and pLDDT<50 indicates very low confidence. Region within the black dashed circle represents the expected location of interaction within domain 2. (B) Graph showing the predicted IDDT per position for all 5 models produced by AlphaFold-Multimer. Rank 1 is represented in blue. (C) PAE plot for the given structure. Error values are given in Angstroms, where 0 indicates low error and 30 indicates high error. (D) Predicted contacts within 4 Å according to the PAE. Identified contacts are coloured by confidence using the same values as the pLDDT. (E) The helicase interaction motif is present on the opposite side of domain 2 showing interaction with the CTD.

4.2.3 AlphaFold-Multimer analysis of PcrA:YwhK interactions

4.2.3.1 YwhK interaction with the PcrA CTD

As shown in Figure 4-10A, YwhK is mostly predicted with very high confidence, particularly within the β-propeller domain, according to the pLDDT scores (Figure 4-10B). However, low confident regions exist towards the N-terminus and within the loops linking the β-strands in the barrel-like domain. The CTD is also predicted with low confidence but again, adopts the same fold as in previous AlphaFold-Multimer submissions. The predicted global structure of the CTD shows low error, but this is not as confident as for YwhK. Whilst AlphaFold-Multimer is confident in its prediction of the YwhK domains, it only has limited confidence in their relative positionings (Figure 4-10C).

The model predicts interaction between the CTD and the barrel-like domain of YwhK (Figure 4-10D). However, as with UvrB, the predicted interactions have low confidence and occur between regions of low structural confidence. In addition, the HIM is present within the β-propeller domain at the opposite side of the protein to the predicted interaction (Figure 4-10E). Thus, the model does not provide support for the hypothesis that PcrA's CTD interacts with the HIM of YwhK. Despite this, the predicted interacting region of YwhK is consistent with the region showing interaction in the Y2H assays, and so there is scope for testing the interaction using site-directed mutagenesis screens.

4.2.3.2 YwhK β-sheet motif interaction with the PcrA CTD

As seen in Figure 4-10A, the β-propeller domain of YwhK is comprised of seven structurally identical β-sheet motif "blades", as predicted by AlphaFold-Multimer. Since the interaction between full-length YwhK and the CTD of PcrA was predicted with low confidence, and did not involve the putative HIM, interaction between the β-sheet containing the HIM and the CTD of PcrA was tested.

As shown in Figure 4-11A, both the β-sheet motif (BSM) of YwhK and the CTD of PcrA are mostly predicted with confidence, particularly high confidence for the CTD, according to their pLDDT scores (Figure 4-11B). However, both have low confidence predictions at their termini. The global structures for both are predicted with low error, and their positioning relative to one another is also predicted with moderate to high confidence (Figure 4-11C).

When contacts within 4 Å were displayed on the model, 12 were predicted with confidence, and 10 were predicted with low confidence (Figure 4-11D). Importantly however, the residues involved in the confident interactions predicted by AlphaFold-Multimer are present towards the core of the β-propeller domain in the complete YwhK structure, and thus would not be

accessible for interaction with the CTD. Interaction between K727 and the conserved glutamate residue is not predicted. In fact, the HIM is oriented away from the CTD (Figure 4-11E). Therefore, the model predicted by AlphaFold-Multimer is unlikely to reflect the true interaction between YwhK and PcrA.

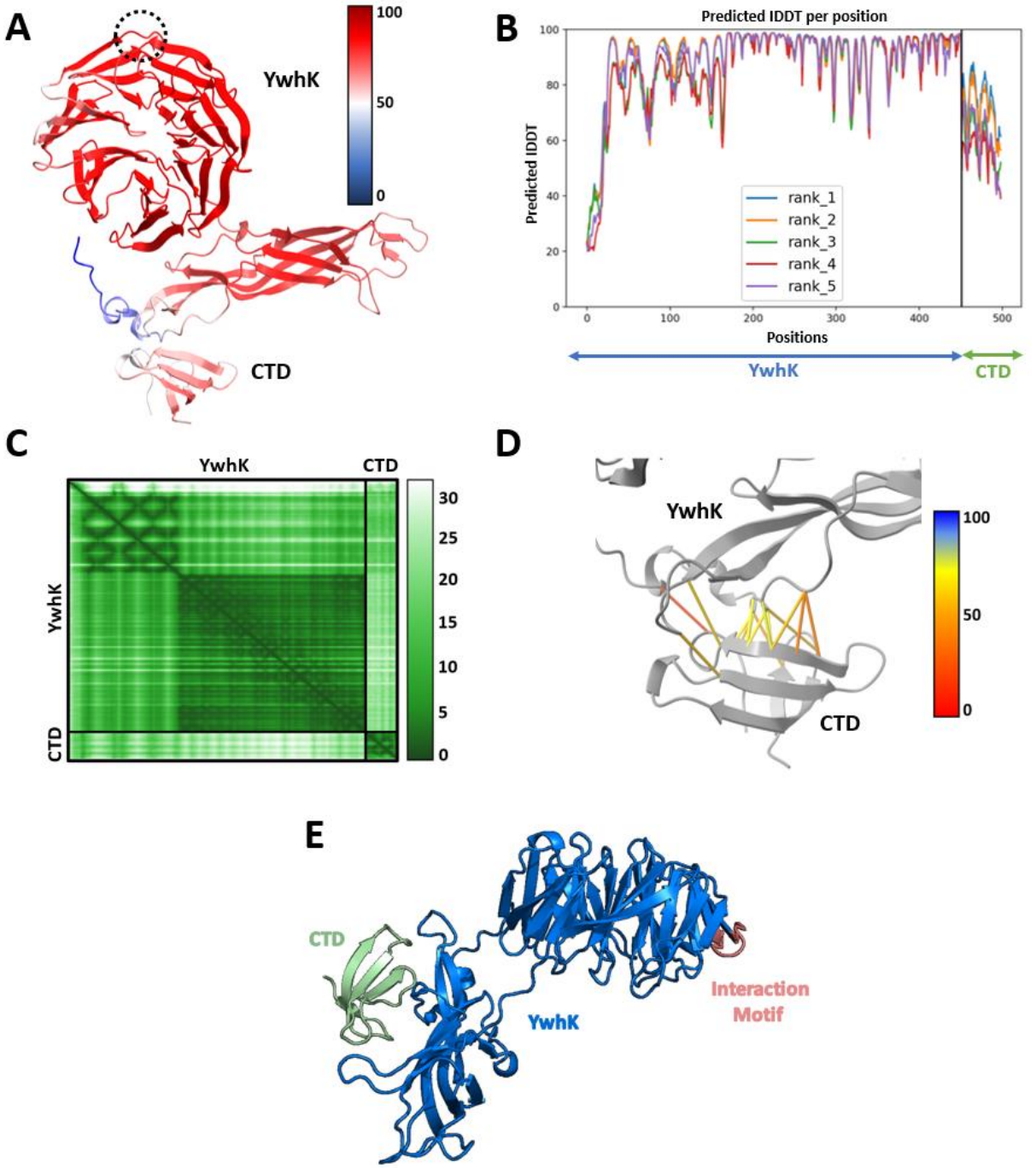

Figure 4-10. YwhK and PcrA CTD interaction prediction by AlphaFold-Multimer. (A) Confidence of prediction according to the pLDDT. pLDDT>90 indicates high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and pLDDT<50 indicates very low confidence. Region within the black dashed circle represents the expected location of interaction within YwhK. (B) Graph showing the predicted IDDT per position for all 5 models produced by AlphaFold-Multimer. Rank 1 is represented in blue. (C) PAE plot for the given structure. Error values are given in Angstroms, where 0 indicates low error and 30 indicates high error. (D) Predicted contacts within 4 Å according to the PAE. Identified contacts are coloured by confidence using the same values as the pLDDT. (E) The helicase interaction motif is present in a different domain to that showing interaction with the CTD.

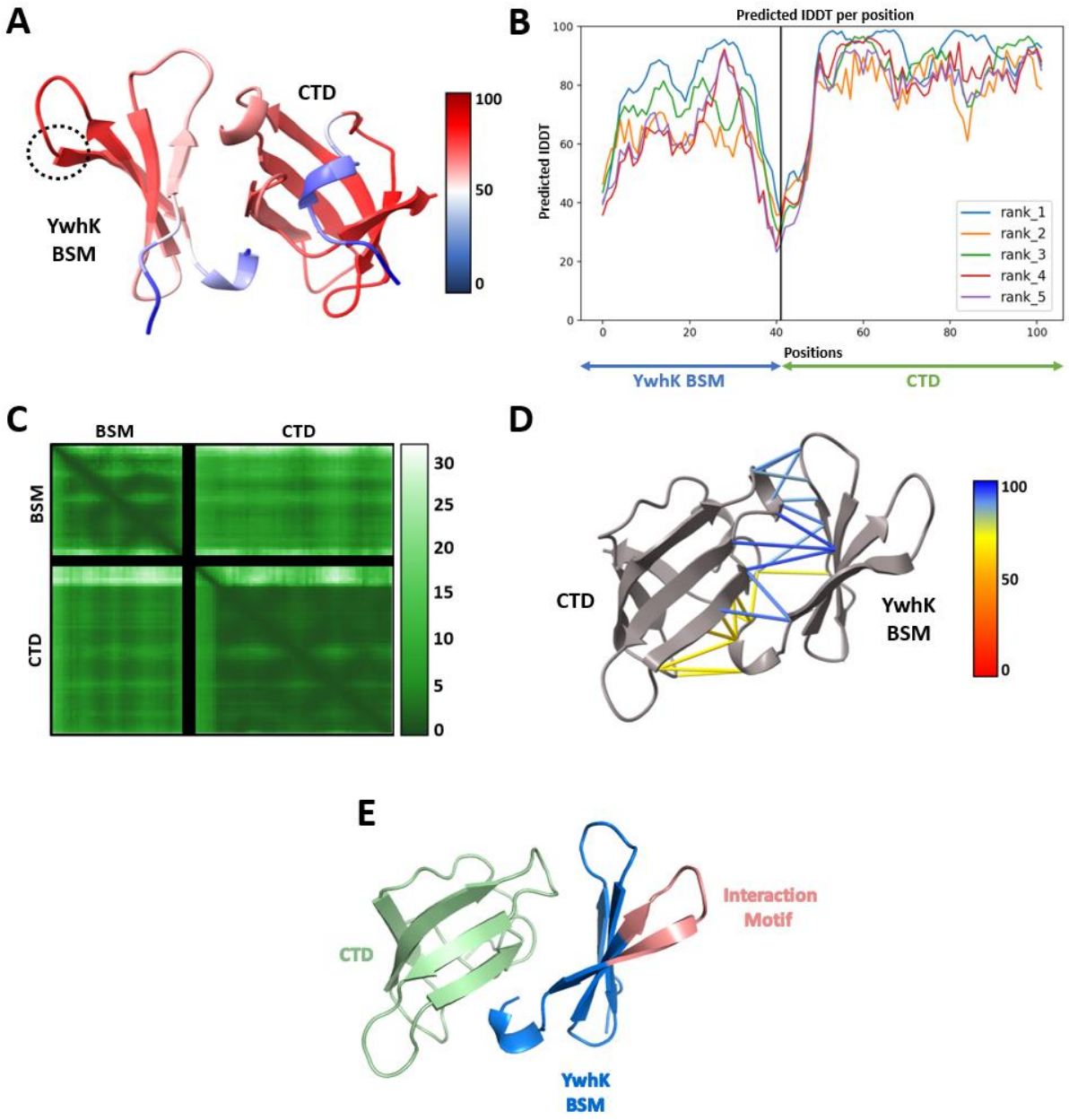

Figure 4-11. YwhK β-sheet motif and PcrA CTD interaction prediction by AlphaFold-Multimer. (A) Confidence of prediction according to the pLDDT. pLDDT>90 indicates high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and pLDDT<50 indicates very low confidence. Region within the black dashed circle represents the expected location of interaction within the YwhK BSM. (B) Graph showing the predicted IDDT per position for all 5 models produced by AlphaFold-Multimer. Rank 1 is represented in blue. (C) PAE plot for the given structure. Error values are given in Angstroms, where 0 indicates low error and 30 indicates high error. (D) Predicted contacts within 4 Å according to the PAE. Identified contacts are coloured by confidence using the same values as the pLDDT. (E) The helicase interaction motif faces away from the CTD and is predicted to not be involved in interactions.

4.2.4 AlphaFold-Multimer analysis of PcrA:YxaL interactions

4.2.4.1 YxaL interaction with the PcrA CTD

As shown in Figure 4-12A, the β-propeller domain of YxaL was predicted with confidence, but the structure of its N-terminal tail has very low confidence, according to the pLDDT scores (Figure 4-12B). This is likely due to its flexibility. Interestingly, only the very N-terminal region of the tail is predicted to be natively disordered (Hu et al., 2021). The CTD of PcrA was again predicted with low confidence but remained identical to that of previous submissions. The global structure of YxaL was mostly predicted with low error, implying correct packing of the BSMs forming the propeller structure (Figure 4-12C). In fact, the predicted structure aligned perfectly to an X-ray structure of YxaL (Figure 4-12D) (Kim et al., 2022).

When contacts between the CTD of PcrA and YxaL were assessed, it was found that all those within 4 Å were predicted with low confidence (Figure 4-12E). As with YwhK, the HIM was found to be present on the solvent-facing side of one of the BSM "blades", whilst the CTD interaction was predicted on a different side of the protein (Figure 4-12F). Thus, this model produced by AlphaFold-Multimer does not provide evidence for interaction between PcrA's CTD and the HIM of YxaL.

4.2.4.2 YxaL β-sheet motif interaction with the PcrA CTD

Since no confident interactions were found between full-length YxaL and PcrA's CTD, the BSM containing the "TGE" triad was tested for interaction with the CTD. As shown in Figure 4-13A, both PcrA's CTD and the BSM of YxaL were mostly predicted with high confidence, but showed low confidence at their respective termini's, according to pLDDT scores (Figure 4-13B). They both also showed low error in the prediction of their global structures, as well as their positioning relative to one another (Figure 4-13C). In particular, the relative positionings of the loop containing the "TGE" triad and the β-strand containing K727 were predicted with very low error (Figure 4-13C, red boxes) and in fact, they are directly facing one another (Figure 4-13D).

When contacts of ≤ 4Å between the two were displayed on the model, they were all shown to be predicted with high confidence (Figure 4-13D). As a result, PDBePISA was used to assess the chemical nature of the interactions as predicted by AlphaFold-Multimer, and it identified a salt bridge between K727 of PcrA and E233 of YxaL (Figure 4-13E). Contacts were also predicted between the CTD and other residues in the HIM (Table 4-3).

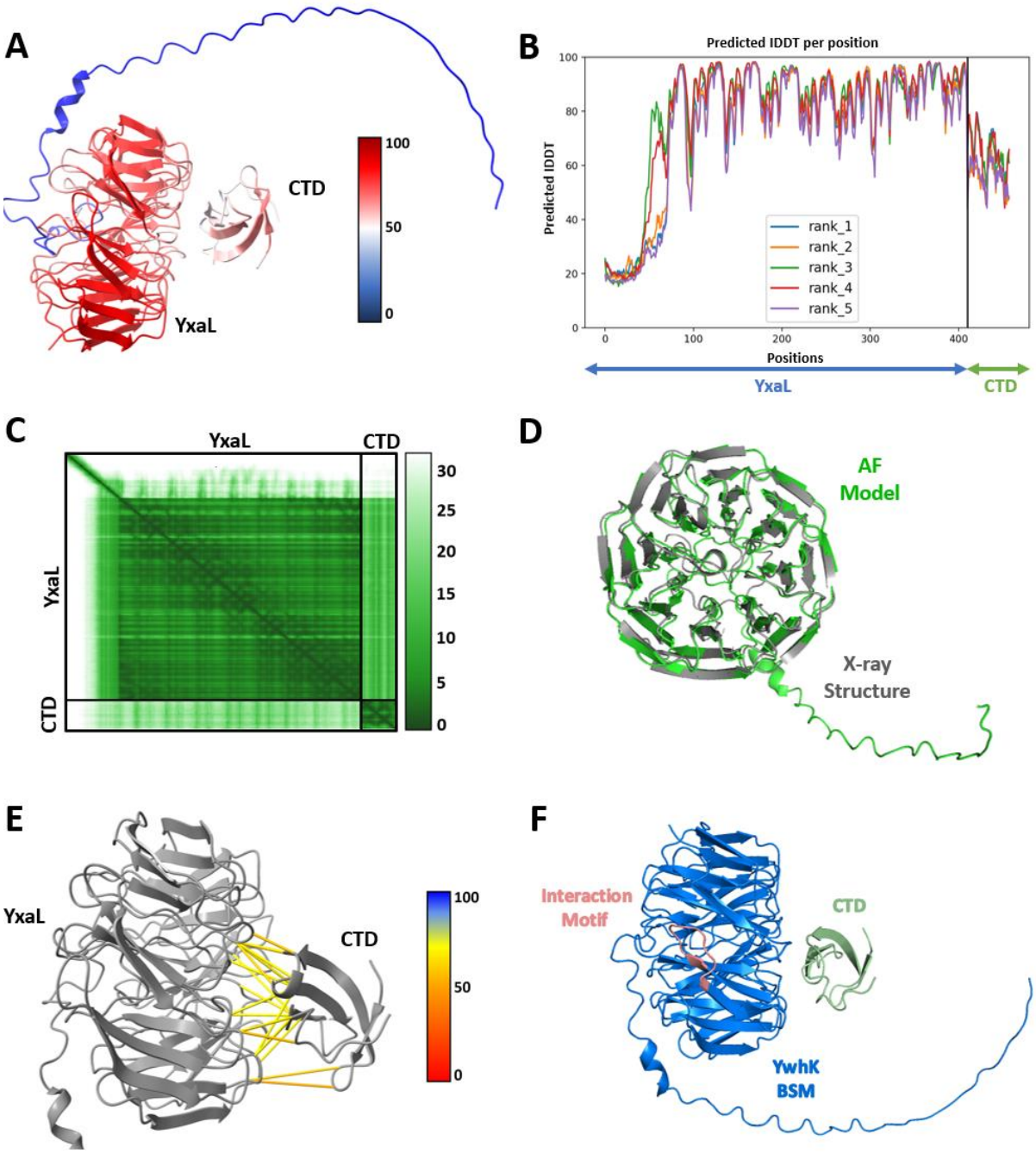

Figure 4-12. YxaL and PcrA CTD interaction prediction by AlphaFold-Multimer. (A) Confidence of prediction according to the pLDDT. pLDDT>90 indicates high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and pLDDT<50 indicates very low confidence. (B) Graph showing the predicted IDDT per position for all 5 models produced by AlphaFold-Multimer. Rank 1 is represented in blue. (C) PAE plot for the given structure. Error values are given in Angstroms, where 0 indicates low error and 30 indicates high error. (D) Alignment of the AlphaFold-Multimer predicted YxaL structure (green) to the structure of YxaL obtained by X-ray crystallography (grey) (PDB: 7DXN). (E) Predicted contacts within 4 Å according to the PAE. Identified contacts are coloured by confidence using the same values as the pLDDT. (F) The helicase interaction motif is present on a different side of YxaL to that showing interaction with the CTD.

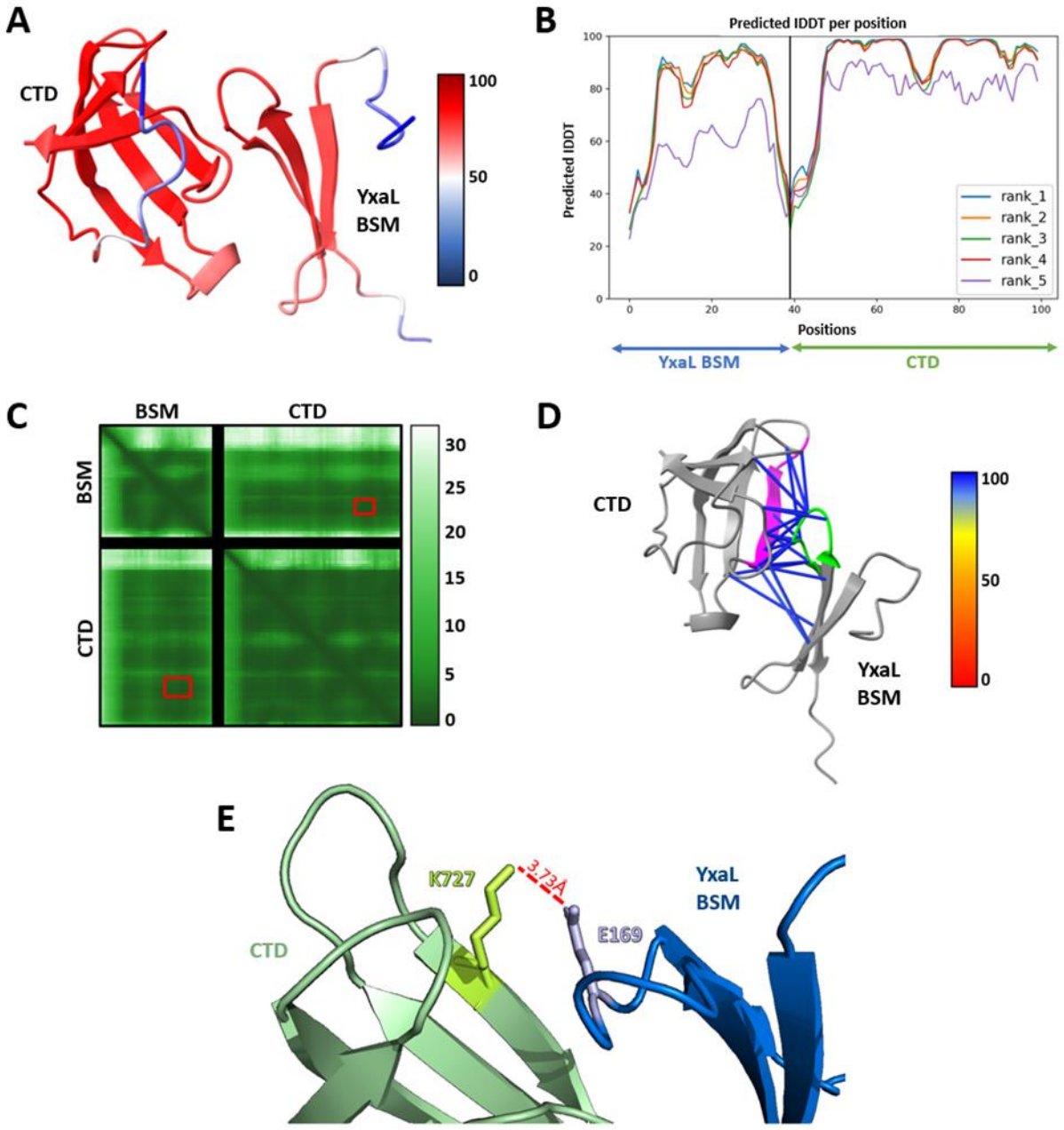

Figure 4-13. YxaL β-sheet motif and PcrA CTD interaction prediction by AlphaFold-Multimer. (A) Confidence of prediction according to the pLDDT. pLDDT>90 indicates high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and pLDDT<50 indicates very low confidence. (B) Graph showing the predicted IDDT per position for all 5 models produced by AlphaFold-Multimer. Rank 1 is represented in blue. (C) PAE plot for the given structure. Regions corresponding to the expected interaction are highlighted by red boxes. Error values are given in Angstroms, where 0 indicates low error and 30 indicates high error. (D) Predicted contacts within 4 Å according to the PAE. Identified contacts are coloured by confidence using the same values as the pLDDT. The loop region containing the putative helicase interaction motif is coloured in green and the β-strand containing K727 of PcrA is coloured in magenta. (E) Salt bridge (red dashed line) of 3.73 Å predicted between E169 of YxaL and K727 of PcrA.

Table 4-4. Interactions between the YxaL BSM and the CTD of PcrA as predicted by PDBePISA.

Bond Type	YxaL Residue	PcrA Residue	Distance (Å)
Salt Bridge	E169	K727	3.73
Hydrogen Bond	T167	H696	3.02
Hydrogen Bond	G168	L730	3.11
Hydrogen Bond	I170	R728	2.86

The remaining three proteins Urrutia-Irazabal et al. (2021) identified as containing the putative HIM have currently not been shown to interact with PcrA through experimental approaches. Thus, they remain only as potential interactors of PcrA. As a result, we were interested to see if AlphaFold-Multimer would predict interactions between these proteins and PcrA's CTD, and if so, whether the HIM was involved. We were also interested to see if the structure of the HIM was consistent with that of the experimentally determined PcrA interactors.

4.2.5 AlphaFold-Multimer analysis of PcrA:QueA interactions

As shown in Figure 4-14A, most of the QueA structure is predicted with moderate to high confidence, but there are regions of low confidence between residues Y134-D160 and T203-F228, according to pLDDT scores (Figure 4-14B). The PAE plot suggests the presence of three distinct domains, however, it is clear from the structure only two exist: a β-barrel domain and an αβα-sandwich domain (Figure 4-14C). There is also low confidence in the positioning of these domains relative to one another. As seen previously, the CTD of PcrA was predicted with low confidence according to pLDDT scores (Figure 4-14B), but its global structure is predicted with mostly low error (Figure 4-14C).

When interactions between QueA and PcrA were analysed, a single contact of very low confidence was present within 4 Å (Figure 4-14D). AlphaFold-Multimer also positions the CTD on the opposite side of QueA to the putative HIM (Figure 4-14E). Furthermore, the predicted QueA structure does not directly align to a structure obtained by X-ray crystallography, including the structure of the HIM, although this is mild (Figure 4-14F) (Grimm et al., 2006). As a result of the error present in the prediction of the individual proteins and also their relative positions, little structural and biological relevance should be applied to the model.

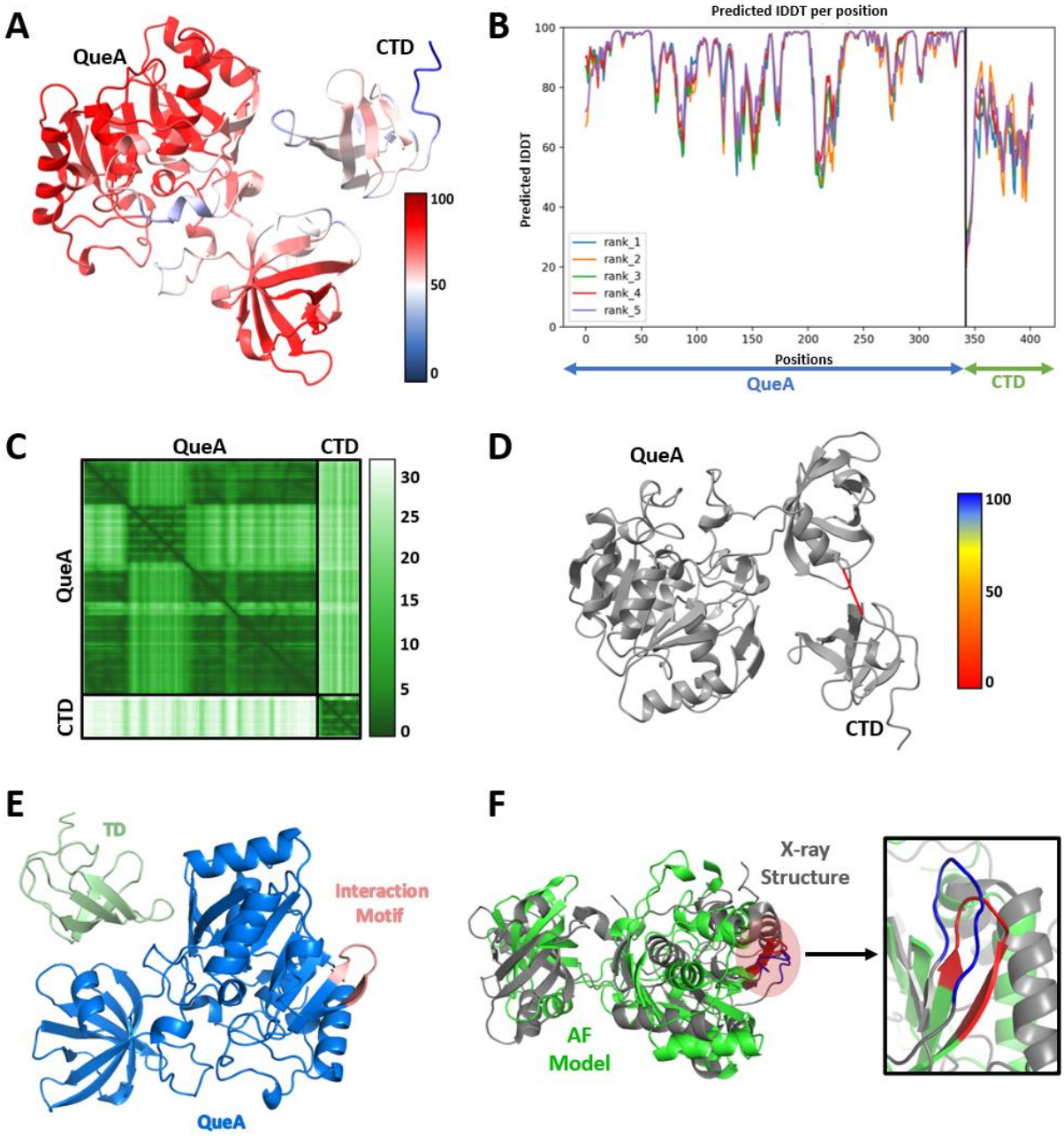

Figure 4-14. QueA and PcrA CTD interaction prediction by AlphaFold-Multimer. (A) Confidence of prediction according to the pLDDT. pLDDT>90 indicates high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and pLDDT<50 indicates very low confidence. (B) Graph showing the predicted IDDT per position for all 5 models produced by AlphaFold-Multimer. Rank 1 is represented in blue. (C) PAE plot for the given structure. Error values are given in Angstroms, where 0 indicates low error and 30 indicates high error. (D) Predicted contacts within 4 Å according to the PAE. Identified contacts are coloured by confidence using the same values as the pLDDT. (E) The helicase interaction motif is present on a different side of QueA to that showing interaction with the CTD. (F) Alignment of the AlphaFold-Multimer predicted QueA structure (green) to the structure of QueA obtained by X-ray crystallography (grey) (PDB: 1YY3). The red circle indicates the region containing the putative HIM. The image within the black square is an enlarged view of the putative HIMs structure. The predicted regions of contact (red) do not directly align to the X-ray structure (blue).

4.2.6 AlphaFold-Multimer analysis of PcrA:RplX interactions

As shown in Figure 4-15A, RplX was mostly predicted with high confidence, including the loop containing the HIM, according to pLDDT scores (Figure 4-15B). However, residues M40-P62 were predicted with low confidence. Despite this, the global structure of RplX was predicted with low error (Figure 4-15C). The CTD of PcrA was predicted with moderate confidence according to pLDDT scores (Figure 4-15B) and with low error in terms of its global structure (Figure 4-15C).

Most of the contacts within 4 Å were predicted with low confidence (Figure 4-15D), and did not involve the HIM, which was present in a loop facing away from the CTD (Figure 4-15E). Due to this, and the lack of experimental evidence for interaction between PcrA and RplX, little structural and biological relevance should be applied to the AlphaFold-Multimer model.

4.2.7 AlphaFold-Multimer analysis of PcrA:YtzB interactions

As shown in Figure 4-16A, the main body of YtzB was predicted with mostly moderate to high confidence, but its helical tail was predicted with very low confidence, according to both the pLDDT and PAE (Figure 4-16, B and C), likely due to its flexibility. As with previous submissions, the CTD is predicted with moderate confidence in terms of pLDDT, but high confidence in terms of overall fold (Figure 4-16, B and C).

The relative positionings of YtzB and the CTD are predicted with high error (Figure 4-16C). As a result, interactions within 4Å are predicted with low confidence (Figure 4-16D) and also do not involve the HIM. As with the other putative interactors of PcrA, the loop is predicted to face away from the CTD (Figure4-16E). Due to the low confidence prediction, little structural and biological relevance should be applied to the model, and any interaction should be validated by experimental approaches.

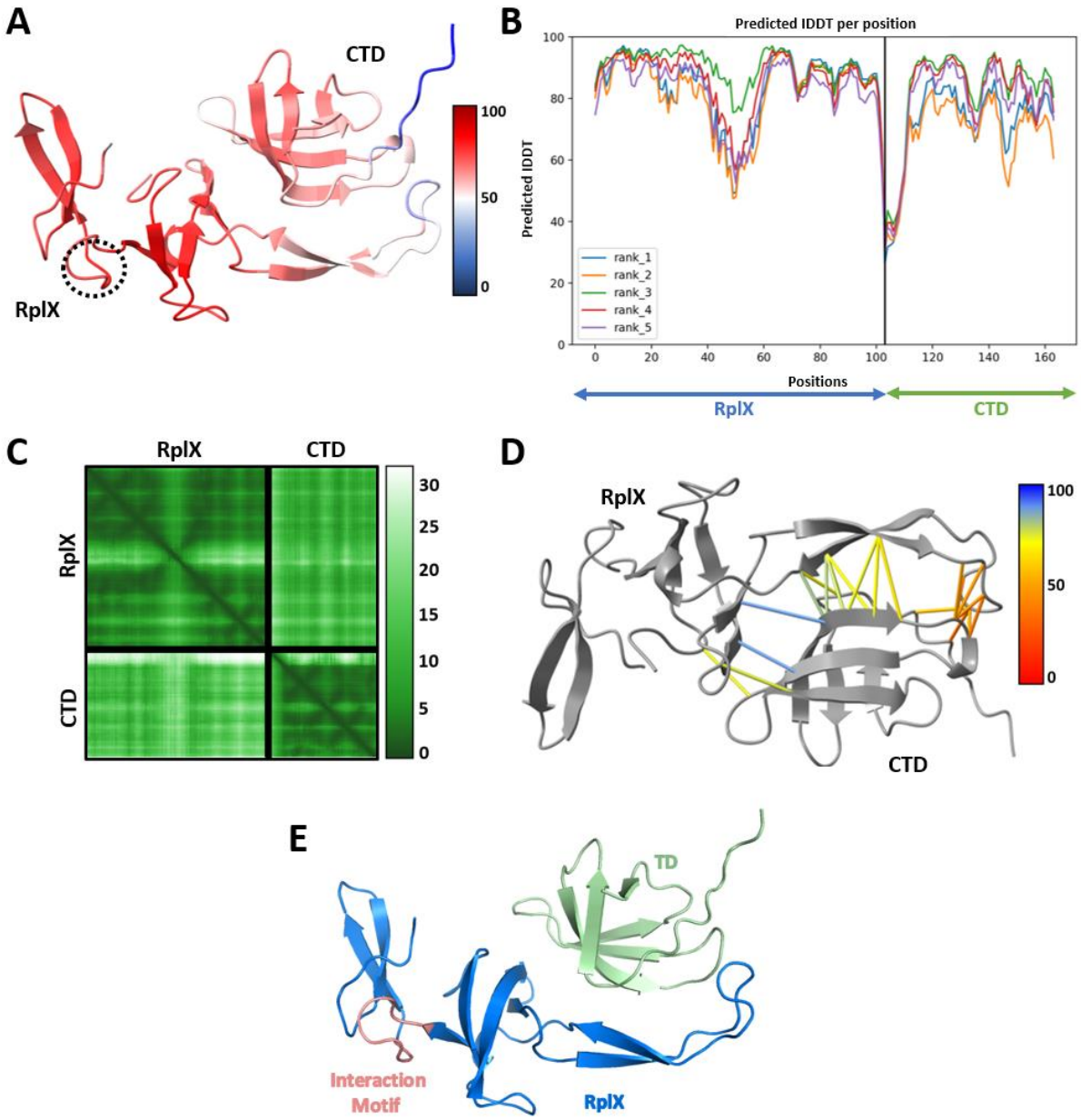

Figure 4-15. RplX and PcrA CTD interaction prediction by AlphaFold-Multimer. (A) Confidence of prediction according to the pLDDT. pLDDT>90 indicates high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and pLDDT<50 indicates very low confidence. Region within the black dashed circle represents the expected location of interaction within RplX (B) Graph showing the predicted IDDT per position for all 5 models produced by AlphaFold-Multimer. Rank 1 is represented in blue. (C) PAE plot for the given structure. Error values are given in Angstroms, where 0 indicates low error and 30 indicates high error. (D) Predicted contacts within 4 Å according to the PAE. Identified contacts are coloured by confidence using the same values as the pLDDT. (E) The helicase interaction motif is present on a different side of RplX to that showing interaction with the CTD.

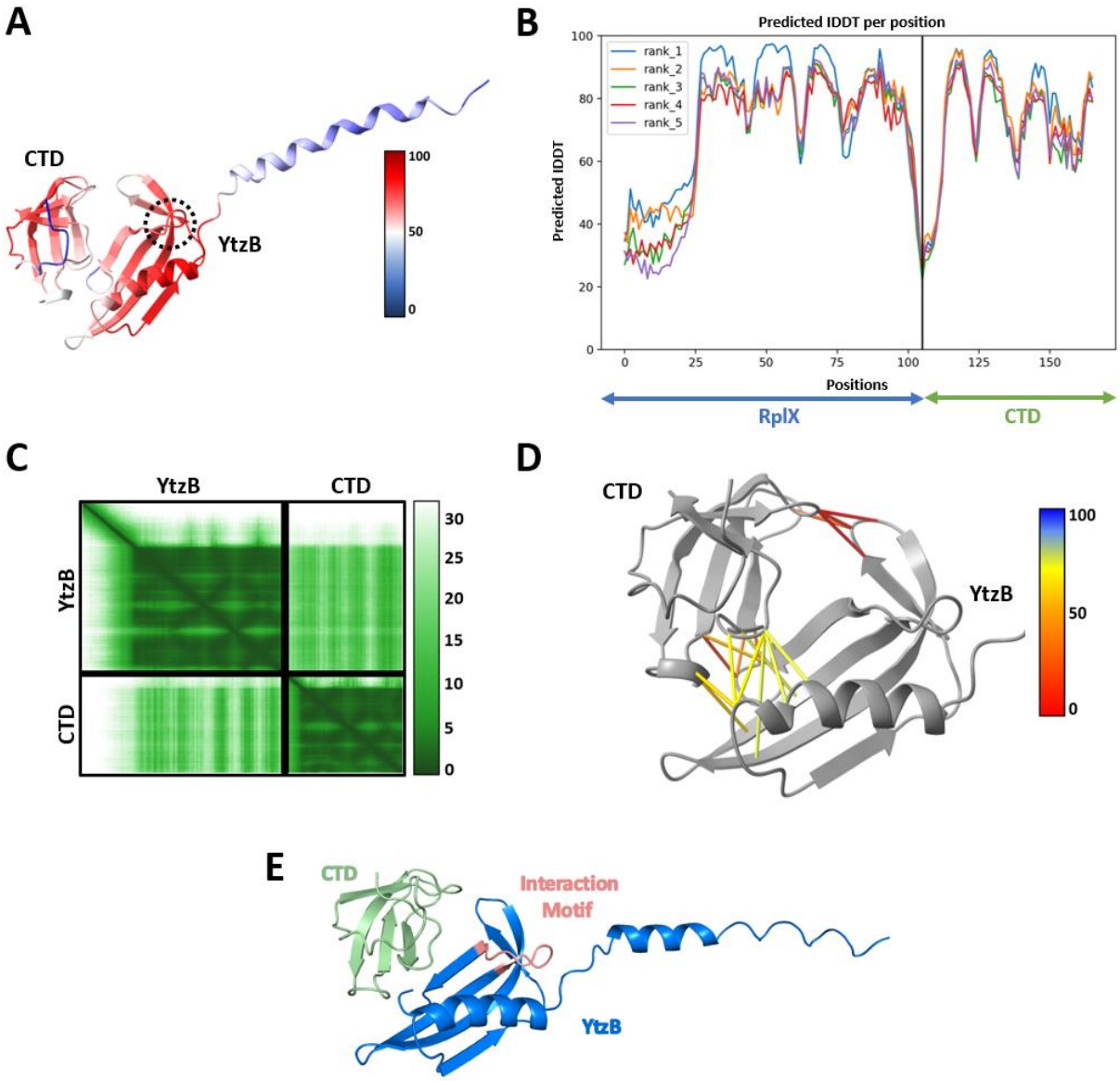

Figure 4-16. YtzB and PcrA CTD interaction prediction by AlphaFold-Multimer. (A) Confidence of prediction according to the pLDDT. pLDDT>90 indicates high confidence, 90>pLDDT>70 indicates moderate confidence, 70>pLDDT>50 indicates low confidence, and pLDDT<50 indicates very low confidence. Region within the black dashed circle represents the expected location of interaction within YtzB. (B) Graph showing the predicted IDDT per position for all 5 models produced by AlphaFold-Multimer. Rank 1 is represented in blue. (C) PAE plot for the given structure. Error values are given in Angstroms, where 0 indicates low error and 30 indicates high error. (D) Predicted contacts within 4 Å according to the PAE. Identified contacts are coloured by confidence using the same values as the pLDDT. (E) The helicase interaction motif faces away from the CTD.

4.3 Discussion and future directions

Although recent HDX-MS and crosslinking studies have highlighted general regions of interaction between PcrA and RNAP, the specific contacts remain unclear. In addition, interactions between PcrA and its experimentally determined and putative partner proteins which contain the HIM await further study. Here, we assessed AlphaFold-Multimer predicted contacts formed between PcrA and it's known and putative interactors.

According to the AlphaFold-Multimer confidence metrics, the structures predicted from the individual protein chains were done so with mostly moderate to high accuracy, which is the aspect of AlphaFold that has repeatedly shown exceptional success (Jumper et al., 2021). However, the PAE plots underestimated the accuracy of several of these predictions: although the plots displayed large regions of uncertainty (white and pale green), the predicted structures mostly aligned directly to experimentally determined structures, where these were available. This was particularly evident for the β′-subunit of RNAP, which perfectly aligned to a cryo-EM structure but was predicted to have a poorly defined inter-domain configuration. Conversely, the plDDT scores provided more accurate representations of the predicted models since regions of low confidence tended not to align to experimentally predicted structures, whilst those with high confidence did. For example, regions showing low plDDT in the predicted β-subunit of RNAP did not align to their respective regions within the cryo-EM structure. It is unsurprising that the predicted structures of proteins which have already been structurally characterised were almost identical to their solved counterparts since AlphaFold was initially trained on over 170,000 protein structures present in the protein data bank (PDB) (Jones and Thornton, 2022).

Despite adopting an almost identical fold across several submissions, PcrA's CTD was predicted with varying confidence. Specifically, the confidence of its prediction decreased as the sequence of the interacting protein chain increased in length. However, preprints have stated that sequence length only minimally impacts the accuracy of AlphaFold2 predictions (Ko and Lee, 2021, Xiong et al., 2022), but the effects of this factor have not been tested for AlphaFold-Multimer. Thus, this could be a potential flaw of the newer system. The same trend was not observed with the other tested domains of PcrA, and in fact, they maintained the same levels of confidence across submissions.

Consistent with experimentally determined structures, the HIMs of the PcrA interactors without any available structural data were all predicted to occupy the tip of a loop present between β-sheets (Eryilmaz et al., 2006, Newing et al., 2020). This provides further support for the proposed conserved interaction between the CTD of PcrA and a number of its partners (Urrutia-Irazabal et al., 2021). Furthermore, AlphaFold-Multimer also predicted similar

structures of YxaL and YwhK to homology models made by Urrutia-Irazabal et al. (2021), providing confidence in their predictions. These predicted structures also supported the suggestion that both YxaL and YwhK are likely to adopt a β-propeller conformation (Noirot-Gros et al., 2002b). In fact, the β-propeller domain of YxaL aligned to a recently solved structure (Kim et al., 2022). In addition, the AlphaFold-Multimer model of UvrB aligned almost directly to an X-ray structure (Webster et al., 2012). Conversely, the predicted QueA structure was mostly misaligned, but still showed an almost identical conformation for the HIM (Grimm et al., 2006).

Many studies have praised the ability of AlphaFold-Multimer to accurately model PPIs (Johansson-Åkhe and Wallner, 2022), particularly for bacterial complexes (Bryant et al., 2022). However, as shown here when known interactions were tested, and by a number of other groups, its accuracy can be variable (Miyazono and Tanokura, 2022, Yin et al., 2022, Gomes et al., 2022). Although the individual protein chains generated mostly confident structures, AlphaFold-Multimer only predicted feasible, high confidence interactions between PcrA and two of its experimentally determined interaction partners (Figures 4-1 and 4-13). In addition, these interactions were only predicted between selected regions of the respective proteins. Despite the fact that interactions between PcrA and both UvrB and YwhK have been experimentally validated, AlphaFold-Multimer failed to predict high confidence interactions (or interactions that made structural sense), either with the full-length proteins or isolated regions containing the HIM. Comparisons between other computational PPI predictors and AlphaFold-Multimer have shown that AlphaFold-Multimer generates a number of false negative results, particularly more so than the number of false positive predictions (Singh et al., 2022). The original publication also found successful heteromeric interface prediction in only 70 % of the cases tested, leaving a considerable percentage of error (Evans et al., 2021). In fact, it's predictions of homomeric interfaces tend to be more accurate than those of heteromeric protein interfaces due to the availability of evolutionary information concerning the protein interfaces (Evans et al., 2021). AlphaFold-Multimer's reliance on paired MSAs is particularly problematic for heteromers since interacting orthologues must be identified. Some identified paralogs may form interactions with different partners through non-conserved interactions, resulting in unclear interpretation and thus, inaccurate predictions (Gao et al., 2022, Chen et al., 2022). This is particularly relevant here since PcrA is closely related to other UvrD-like helicases, such as *E.coli* Rep, which shares significant sequence similarity and similar domain architectures. Therefore, it is likely that Rep will be sampled by AlphaFold in PcrA MSAs. However, Rep does not contain the Tudor domain and is not expected to interact with RNAP and other HIM-containing proteins, possibly leading to false negative results. Conversely, proteins which are not ubiquitous, including YxaL, YwhK and YtzB, will have fewer homologs

to sample from. Thus, it is possible that the lack of interactions shown here, particularly between PcrA and its interactors that have experimental validation, are also false negative results. Optimisation of MSAs can generate more reliable predictions and minimise false positives (Laurents, 2022). In fact, Bryant et al. (2022) found that optimisation of the MSA improved the accuracy of prediction of bacterial heterodimeric complexes. In addition to optimising MSAs, a greater number of highly accurate models can be obtained by forcing AlphaFold to sample a more diverse selection of conformations (Johansson-Åkhe and Wallner, 2022). More accurate protein complex predictions may be generated by combining AlphaFold2 with other docking programs. Ghani et al. (2022) used a combination of AlphaFold2 and ClustPro to substantially improve AlphaFold's accuracy of predictions of PPIs by close to 40 %. Since the individual structures generated here were done so with confidence, the putative interactions may be predicted by combining the AlphaFold results with a docking program.

Although AlphaFold-Multimer did not predict some expected interactions here, and predicted several high-error interactions, its confidence metrics were representative of these inaccuracies, something which was identified by Evans et al. (2021). Thus, our results highlight the importance of assessing the confidence metrics when interpreting AlphaFold-Multimer structures, particularly when assessing novel PPIs. In addition, when testing interactions between segments of proteins, it is essential to be wary of their location within the complete protein structure, since AlphaFold-Multimer treats each sequence as a stand-alone peptide. As a result, it can predict interactions which may not be feasible in nature.

The two interactions showing confidence were between the CTD of PcrA and both the SI1 domain of RNAP and a β-sheet motif of YxaL. The salt bridge predicted to form between K727 of PcrA and a conserved glutamate residue, E301, present within the putative HIM of RNAP's β-subunit has been shown to play a critical role in the PcrA:RNAP interaction (Sanders et al., 2017, Urrutia-Irazabal et al., 2021). Similarly, PcrA and YxaL have been shown to interact experimentally through Y2H and functional assays, but the specific location of the interaction has not been determined (Noirot-Gros et al., 2002b). It was proposed that this interaction occurs between K727 of PcrA and the same conserved glutamate residue (E169) present within the HIM of YxaL (Urrutia-Irazabal et al., 2021). Here, AlphaFold-Multimer predicted interaction between K727 of PcrA and E169 of YxaL, supporting this hypothesis. This result is also consistent with the fragment of YxaL that showed interaction in the Y2H assays. Both predicted interactions also uncovered several stabilising hydrogen bonds between the HIMs of RNAP and YxaL and PcrA's CTD. In fact, these interactions bond the same residues of the CTD and the same residues within the HIM of both interacting proteins. H696 of PcrA, which is predicted to form a hydrogen bond with T299 of RNAP's β-subunit and T167 of YxaL, has

been mutated in a previous study, and found to severely disrupt interaction between PcrA and RNAP (Sanders et al., 2017). However, aside from K727, the other residues present within PcrA's CTD that are predicted to form interactions with its partners have not been assessed experimentally and thus provide novel targets for mutational analysis. This includes R728 and L730 of PcrA. Despite this, similarities between the two predicted interactions provides evidence for a conserved interaction between PcrA's CTD and the HIM of RNAP and YxaL, which has the potential to be extended to other PcrA interactors containing the HIM.

AlphaFold-Multimer predicted a second interaction between RNAP and PcrA, involving the 2B subdomain of PcrA and the β-subunit of RNAP, but this was with very low confidence. Despite this, the region of interaction within the β-subunit is present on the surface of the protein and occupies the 3D space in-between the residues K1129-D1141 of the β-subunit and K372-L412 of the β′-subunit, which showed protection by PcrA in HDX-MS experiments. In fact, when Urrutia-Irazabal et al. (2021) modelled a PcrA:RNAP complex based off of these results, it was the 2B subdomain that was predicted to interact with K1129-D1141. Thus, the close proximity of the predicted interaction to regions already shown to interact with PcrA may provide another region of interaction that had not been identified by HDX-MS.

Aside from the two interactions mentioned, AlphaFold-Multimer did not predict any other interactions between PcrA and RNAP, despite the wealth of experimental evidence for their interaction. However, interactions between the full-length proteins could not be tested due to the approximate 1,400 residue limit of ColabFold. Nevertheless, predictions involving full-length proteins do not guarantee more accurate models, as seen here with YxaL. AlphaFold is further limited by its inability to model nucleic acids, which may become problematic when predicting the structure of nucleic acid binding proteins and their interactions, such as RNAP and PcrA. Since the input only comprises amino acid sequences, the program fails to model other ligands, cofactors and metal ions, and in turn, cannot predict any conformational changes caused by these interactions (Miyazono and Tanokura, 2022). Thus, AlphaFold only outputs single states of proteins, consequently providing little functional information.

Here, only the default parameters of ColabFold were used. However, increasing the recycle count can increase the accuracy of the predicted models, which is particularly useful in cases where little MSA information is available (Mirdita et al., 2022). However, this was not applied here due to lack of the necessary computing power. Accuracy may also be increased in further analyses by introducing a template for one of the interacting pairs, but evidently, this is only possible for proteins with solved structures.

These results, and those obtained by others in the field, highlight that AlphaFold and its protein complex prediction counterpart, Multimer, still remain in their early stages. Whilst AlphaFold-

Multimer provides a high-throughput method of screening for novel PPIs, care should be taken, and predictions assessed with reference to available biological data and their structural viability. Currently, the only possible methods for confidently validating AlphaFold models involve experimental approaches, and thus we are limited to its own confidence estimates. By combining AlphaFold-Multimer with some of the other approaches mentioned, a complete structure of the PcrA:RNAP complex may be predicted. This has the potential to assist with structural determination by fitting to cryo-EM electron density maps. Furthermore, AlphaFold-Multimer has identified a variety of candidates for site-directed mutagenesis screens across its predicted interactions.

4.4 Chapter 4 summary

• AlphaFold-Multimer predictions of the interaction between RNAP's SI1 domain and PcrA's CTD, as well as between a β-sheet motif of YxaL and PcrA's CTD, revealed similar high-confidence contacts involving the same residues of PcrA. This identified novel amino acid targets of PcrA (R728 and L730) for mutagenesis studies.

• Some high-confidence interactions were predicted between a β-sheet motif of YwhK and the CTD of PcrA, however these interactions would not be feasible in the full-length structure of YwhK.

• Lower confidence interactions were predicted between PcrA's CTD and full-length UvrB, UvrB's domain 2, QueA, RplX, YtzB and full-length YxaL, but none of these interactions involved the HIM of the partner protein.

• No interactions were predicted between the 2A domain of PcrA and either the β or β′ subunit of RNAP or between the 2B domain of PcrA and the β′-subunit of RNAP. However, a low confidence interaction, which is consistent with previous structural data, was predicted between the 2B domain and the β-subunit of RNAP. This predicted interaction requires experimental validation.

• No interaction was predicted between the 2A/2B/CTD of PcrA and the β/β′ subunits of RNAP.

• The HIMs of PcrA's partner proteins modelled here were all predicted to occupy a loop structure present between two β-sheets.

Appendix

Table A-1. *E.coli* contaminants present in material from co-expression of PcrA and RNAP, uncovered by mass spectrometry analysis. Results concerning PcrA are highlighted in light grey. PSMs refers to peptide spectrum matches, FDR refers to the false discovery rate, whereby “High” is identified at 1 % FDR.

Accession	Description	Coverage	# Peptides	# PSMs	# Unique Peptides	Protein FDR confidence
A0A140N587	Bifunctional polymyxin resistance protein ArnA OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=arnA PE=3 SV=1	60.45	36	155	36	High
A0A140NE25	Glutamine--fructose-6-phosphate aminotransferase [isomerizing] OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=glmS PE=3 SV=1	22.00	9	11	9	High
A0A140NCN5	DNA polymerase III subunit gamma/tau OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=dnaX PE=3 SV=1	12.91	7	8	7	High
A0A140NFV3	Chaperone protein DnaK OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=dnaK PE=2 SV=1	12.54	6	6	6	High
A0A6M4JKU6	ATP-dependent DNA helicase OS=Bacillus subtilis (strain 168) OX=224308 GN=pcrA PE=3 SV=1	8.66	5	5	5	High

A0A140ND61	Chaperone protein HtpG OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=htpG PE=3 SV=1	12.98	6	6	6	High
A0A140NBA5	30S ribosomal protein S1 OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_2684 PE=3 SV=1	7.36	3	3	3	High

Table A-2. Protein sequences used in AlphaFold-Multimer predictions of interactions between PcrA and its interaction partners containing a HIM.

Protein	Sequence	Figure(s)
PcrA (CTD)	SKTGGDTLNWAVGDKAGHKKWGTGTVVSVKGEGEGTELDI AFPSPVGVKRLLAAFAPIEKQ	4-1 4-7 4-8 4-9 4-10 4-11 4-12 4-13 4-14 4-15 4-16
PcrA (2A)	QNYRSTKRILRAANEVIKNNSNRKPKNLWTENDEGIKISYYR GDNEFGEGQFVAGKIHQLHSTGKRKLSDIAILYRTNAQSRVIE ETLLKAGLNYNIGKDAITLMTLHAAKGLEFPVVFLMGLEEGVF PHSRSLMEEAEMEEERRLAYVGITRAEQELYLTNAKMRTLF GRTNMNPESRFIAEIPDDLLENLNE	4-3 4-4
PcrA (2B)	VGGTKFYDRKEIKDILAYLRLVSNPDDDISFTRIVNVPKRGVG ATSLEKIASYAAINGLSFFQAIQQVDFIGVSAKAANALDSFRQ MIENLTNMQDYLSITELTEEILDKTEYREMLKAEKSIEAQSRLE NIDEFLSVTKNFEQKSEDKTLVAFLTDLALIADIDQLDQKEEES G	4-5 4-6
PcrA (2A & 2B)	QNYRSTKRILRAANEVIKNNSNRKPKNLWTENDEGIKISYYR GDNEFGEGQFVAGKIHQLHSTGKRKLSDIAILYRTNAQSRVIE ETLLKAGLNYNIVGGTKFYDRKEIKDILAYLRLVSNPDDDISFT RIVNVPKRGVGATSLEKIASYAAINGLSFFQAIQQVDFIGVSAK AANALDSFRQMIENLTNMQDYLSITELTEEILDKTEYREMLKA EKSIEAQSRLENIDEFLSVTKNFEQKSEDKTLVAFLTDLALIAD IDQLDQKEEESGGKDAITLMTLHAAKGLEFPVVFLMGLEEGV FPHSRSLMEEAEMEEERRLAYVGITRAEQELYLTNAKMRTLF GRTNMNPESRFIAEIPDDLLENLNE	4-7

QueA	MKVDLFDFELPERLIAQVPLEQRDASRLMVLDKHTGELTDSS FKHIISFFNEGDCLVLNNTRVLPARLFGTKEDTGAKVELLLLK QETGDKWETLAKPAKRVKKGTVVTFGDGRLKAICTEELEHG GRKMEFQYDGIFYEVLESLGEMPLPPYIKEQLDDKERYQTVY SKEIGSAAAPTAGLHFTEEILQQLKDKGVQIEFITLHVGLGTFR PVSADEVEEHNMHAEFYQMSEETAAALNKVRENGGRIISVG TTSTRTLETIAGEHDGQFKASSGWTSIFIYPGYEFKAIDGMIT NFHLPKSSLIMLVSALAGRENILRAYNHAVEEEYRFFSFGDA MLII	4-14
RNAP (β-subunit)	MTGQLVQYGRHRQRRSYARISEVLELPNLIEIQTSSYQWFLD EGLREMFQDISPIEDFTGNLSLEFIDYSLGEPKYPVEESKERD VTYSAPLRVKVRLINKETGEVKDQDVFMGDFPIMTDTGTFIIN GAERVIVSQLVRSPSVYFSGKVDKNGKKGFTATVIPNRGAW LEYETDAKDVVYVRIDRTRKLPVTVLLRALGFGSDQEILDLIG ENEYLRNTLDKDNTENSDKALLEIYERLRPGEPPTVENAKSL LDSRFFDPKRYDLANVGRYKINKKLHIKNRLFNQRLAETLVDP ETGEILAEKGQILDRRTLDKVLPYLENGIGFRKLYPNGGVVED EVTLQSIKIFAPTDQEGEQVINVIGNAYIEEEIKNITPADIISSISY FFNLLHGVGDTDDIDHLGNRRLRSVGELLQNQFRIGLSRMER VVRERMSIQDTNTITPQQLINIRPVIASIKEFFGSSQLSQFMDQ TNPLAELTHKRRLSALGPGGLTRERAGMEVRDVHYSHYGR MCPIETPEGPNIGLINSLSSYAKVNRFGFIETPYRRVDPETGK VTGRIDYLTADEEDNYVVAQANARLDDEGAFIDDSIVARFRG ENTVVSRNRVDYMDVSPKQVVSAATACIPFLENDDSNRALM GANMQRQAVPLMQPEAPFVGTGMEYVSGKDSGAAVICKHP GIVERVEAKNVWVRRYEEVDGQKVKGNLDKYSLLKFVRSNQ GTCYNQRPIVSVGDEVVKGEILADGPSMELGELALGRNVMV GFMTWDGYNYEDAIIMSERLVKDDVYTSIHIEEYESEARDTKL GPEEITRDIPNVGEDALRNLDDRGIIRIGAEVKDGDLLVGKVT PKGVTELTAEERLLHAIFGEKAREVRDTSLRVPHGGGGIIHDV KVFNREDGDELPPGVNQLVRVYIVQKRKISEGDKMAGRHGN KGVISKILPEEDMPYLPDGTPIDIMLNPLGVPSRMNIGQVLEL HMGMAARYLGIHIASPVFDGAREEDVWETLEEAGMSRDAKT VLYDGRTGEPFDNRVSVGIMYMIKLAHMVDDKLHARSTGPY SLVTQQPLGGKAQFGGQRFGEMEVWALEAYGAAYTLQEILT VKSDDVVGRVKTYEAIVKGDNVPEPGVPESFKVLIKELQSLG MDVKILSGDEEEIEMRDLEDEEDAKQADGLALSGDEEPEETA SADVERDVVTKE	4-3 4-5 4-7
RNAP (β′-subunit)	MLDVNNFEYMNIGLASPDKIRSWSFGEVKKPETINYRTLKPE KDGLFCERIFGPTKDWECHCGKYKRVRYKGVVCDRCGVEV TRAKVRRERMGHIELAAPVSHIWYFKGIPSRMGLVLDMSPRA LEEVIYFASYVVTDPANTPLEKKQLLSEKEYRAYLDKYGNKF QASMGAEAIHKLLQDIDLVKEVDMLKEELKTSQGQRRTRAIK RLEVLEAFRNSGNKPSWMILDVLPVIPPELRPMVQLDGGRFA TSDLNDLYRRVINRNNRLKRLLDLGAPSIIVQNEKRMLQEAVD ALIDNGRRGRPVTGPGNRPLKSLSHMLKGKQGRFRQNLLGK RVDYSGRSVIVVGPHLKMYQCGLPKEMALELFKPFVMKELV EKGLAHNIKSAKRKIERVQPEVWDVLESVIKEHPVLLNRAPTL HRLGIQAFEPTLVEGRAIRLHPLVCTAYNADFDGDQMAVHVP LSAEAQAEARILMLAAQNILNPKDGKPVVTPSQDMVLGNYYL TLERAGAVGEGMVFKNTDEALLAYQNGYVHLHTRVAVAANS LKNVTFTEEQRSKLLITTVGKLVFNEILPESFPYMNEPTKSNIE EKTPDRFFLEKGADVKAVIAQQPINAPFKKGILGKIIAEIFKRFH ITETSKMLDRMKNLGFKYSTKAGITVGVSDIVVLDDKQEILEE AQSKVDNVMKQFRRGLITEEERYERVISIWSAAKDVIQGKLM	4-4 4-6 4-7

	KSLDELNPIYMMSDSGARGNASNFTQLAGMRGLMANPAGRI IELPIKSSFREGLTVLEYFISTHGARKGLADTALKTADSGYLTR RLVDVAQDVIIRETDCGTDRGILAKPLKEGTETIERLEERLIGR FARKQVKHPETGEVLVNENELIDEDKALEIVEAGIEEVWIRSA FTCNTPHGVCKRCYGRNLATGSDVEVGEAVGIIAAQSIGEPG TQLTMRTFHTGGVAGDDITQGLPRIQELFEARNPKGQATITEI DGTVVEINEVRDKQQEIVVQGAVETRSYTAPYNSRLKVAEG DKITRGQVLTEGSIDPKELLKVTDLTTVQEYLLHEVQKVYRM QGVEIGDKHVEVMVRQMLRKVRVIDAGDTDVLPGTLLDIHQF TEANKKVLLEGNRPATGRPVLLGITKASLETDSFLSAASFQET TRVLTDAAIKGKRDELLGLKENVIIGKLVPAGTGMMKYRKVKP VSNVQPTDDMVPVE	
RNAP (Si1 domain)	KNRLFNQRLAETLVDPETGEILAEKGQILDRRTLDKVLPYLEN GIGFRKLYPNGGVVEDEVTLQSIKIFAPTDQEGEQVINVIGNA YIEEEIKNITPAD	4-1
RplX	MHVKKGDKVMVISGKDKGKQGTILAAFPKKDRVLVEGVNMV KKHSKPTQANPQGGISNQEAPIHVSNVMPLDPKTGEVTRVG YKVEDGKKVRVAKKSGQVLDK	4-15
UvrB	MKDRFELVSKYQPQGDQPKAIEKLVKGIQEGKKHQTLLGAT GTGKTFTVSNLIKEVNKPTLVIAHNKTLAGQLYSEFKEFFPNN AVEYFVSYYDYYQPEAYVPQTDTFIEKDASINDEIDKLRHSAT SALFERRDVIIIASVSCIYGLGSPEEYREMVVSLRTEMEIERNE LLRKLVDIQYARNDIDFQRGTFRVRGDVVEIFPASRDEHCVR VEFFGDEIERIREVDALTGEILGDRDHVAIFPASHFVTRAEKM EKAIQNIEKELEEQLKVMHENGKLLEAQRLEQRTRYDLEMMR EMGFCSGIENYSRHLTLRPPGSTPYTLLDYFPDDFMIVVDES HVTIPQVRGMFNGDQARKQVLVDHGFRLPSALDNRPLRFEE FEKHMHNIVYVSATPGPYEIEHTDEMVEQIIRPTGLLDPLIDVR PIEGQIDDLIGEIQARIERNERVLVTTLTKKMSEDLTDYLKEIGI KVNYLHSEIKTLERIEIIRDLRLGKYDVLVGINLLREGLDIPEVS LVAILDADKEGFLRSERSLIQTIGRAARNAEGRVIMYADKITKS MEIAINETKRRREQQERFNEEHGITPKTINKEIRDVIRATVAAE DKAEYKTKAAPKLSKMTKKERQKVVEQMEHEMKEAAKALDF ERAAELRDLLLELKAEG	4-8
UvrB (Domain 2)	YREMVVSLRTEMEIERNELLRKLVDIQYARNDIDFQRGTFRV RGDVVEIFPASRDEHCVRVEFFGDEIERIREVDALTGEILGDR DHVAIFPAS	4-9
YtzB	MKLRHLLLGAGLGICTAVVVRQYVMKPYISSEKALRIVKSAFK QRGPIDGSWIYTQPEPYNINGETVQVYKTGITRSAFGELEQY EVMVDAKTGEIVDVIDTIAS	4-16
YwhK	MRKNKLSFKEKAGAEQEECLCFCEGEASREAMFQAPIELPE GFVVDPAEAVANVTWNADSLSCVSDRCLIQTGPEPDEVGVQ FAVRLQGTVTLLVSVSPVRNQYGQGDGAVSVIHNAEIDQVVY YSDESDDCPDISDITIENLVVVPPFYGSPLSVTGTIVLVPEPEP VYAFTANPNDQSVSVIDTNTDTVVTTIALPYNPAGIEITPDKSA VFVLHPNNNVISVIDYDTLTVTATILLDQPPRLIRFIPNHEFAYV FTGTAVYVIGIDTLTVDRSIPVEGYDVAIDPNGLFAYVLNFGIV QKVDLTTGEVTGTIERELIVSTIETNWPERYAYVLEQEFFFNY LTVIDLNTFTISSTQELEYEGEYRMFTSGAEVYLYDGFTGNLY SVSPNGAGVIGNVPQSATDYAFTPNGDFLYATRFIEQSIIVYN TDDYSEETVISLGVSPGAITI	4-10

YwhK (BSM)	EGYDVAIDPNGLFAYVLNFGIVQKVDLTTGEVTGTIERELI	4-11
YxaL	MVKSFRMKALIAGAAVAAAVSAGAVSDVPAAKVLQPTAAYA AETVFSQNNGASGFLPGRYDVQAMAPAMFNWSRESRFAGN TDGTLKWQNDIRTTPQNGAGAVIDGDGTVYLHSRDGEMKAF NPDGSVKWVTGNLGKTYTQSPVLGTNGVIYLASYDKKIYFID KETGEILTTVPLSGGPSSETVIGSDGTLYFSTLDNYVHAIKPT SKSTWTERWKLKTNGVVSSVPVLAKNGTVYVGTYNYVFYAI NSGTGQVKWSRTTSNAFKGYPVIDKDGNIYAGNQDGQLYAY TSTGSLKWTFPLNGFSSSSPAIDHNGNIYIGSGSGELFSISKN GDMNWSFYTDGPVRTAPLIDAKGTVYFGSDDMKVYAADAN GNELWSYQTDSNVVSSPQLAEDGTLYIGSYTKLMAFGK	4-12
YxaL (BSM)	SPVLGTNGVIYLASYDKKIYFIDKETGEILTTVPLSGGP	4-13

References

Adamus, K., Le, S. N., Elmlund, H., Boudes, M. & Elmlund, D. 2019. AgarFix: Simple and accessible stabilization of challenging single-particle cryo-EM specimens through crosslinking in a matrix of agar. Journal of structural biology, 207, 327-331.

Adebali, O., Chiou, Y.-Y., Hu, J., Sancar, A. & Selby, C. P. 2017. Genome-wide transcription-coupled repair in Escherichia coli is mediated by the Mfd translocase. Proceedings of the National Academy of Sciences - PNAS, 114, E2116-E2125.

Aguilera, A. & García-Muse, T. 2012. R Loops: From Transcription Byproducts to Threats to Genome Stability. Molecular cell, 46, 115-124.

Allison, D. F. & Wang, G. G. 2019. R-loops: formation, function, and relevance to cell stress. Cell Stress, 3, 38-46.

Arora, R., Lee, Y., Wischnewski, H., Brun, C. M., Schwarz, T. & Azzalin, C. M. 2014. RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. Nature communications, 5, 5220-5220.

Blattner, F. R., Plunkett, G., Iii, Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., Shao, Y., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K. & Mayhew, G. F. 1997. The Complete Genome Sequence of Escherichia coli K-12. Science (American Association for the Advancement of Science), 277, 1453-1469.

Boguslawski, S. J., Smith, D. E., Michalak, M. A., Mickelson, K. E., Yehle, C. O., Patterson, W. L. & Carrico, R. J. 1986. Characterization of monoclonal antibody to DNA · RNA and its application to immunodetection of hybrids. Journal of Immunological Methods, 89, 123-130.

Bolanos-Garcia, V. M. & Davies, O. R. 2006. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. Biochimica et biophysica acta, 1760, 1304-1313.

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. The EMBO journal, 29, 145-157.

Bryant, P., Pozzati, G. & Elofsson, A. 2022. Improved prediction of protein-protein interactions using AlphaFold2. Nature communications, 13, 1265-1265.

Caron, P. R., Kushner, S. R. & Grossman, L. 1985. Involvement of Helicase II (uvrD Gene Product) and DNA Polymerase I in Excision Mediated by the uvrABC Protein Complex. Proceedings of the National Academy of Sciences - PNAS, 82, 4925-4929.

Carr, S. B., Phillips, S. E. V. & Thomas, C. D. 2016. Structures of replication initiation proteins from staphylococcal antibiotic resistance plasmids reveal protein asymmetry and flexibility are necessary for replication. Nucleic Acids Research, 44, 2417-2428.

Chen, B., Xie, Z., Xu, J., Qiu, J., Ye, Z. & Tang, J. 2022. Improve the Protein Complex Prediction with Protein Language Models. bioRxiv.

Chen, H. & Skolnick, J. 2008. M-TASSER: An Algorithm for Protein Quaternary Structure Prediction. Biophysical journal, 94, 918-928.

Chen, R., Li, L. & Weng, Z. 2003. ZDOCK: An initial-stage protein-docking algorithm. Proteins, structure, function, and bioinformatics, 52, 80-87.

Cohen, S., Puget, N., Lin, Y.-L., Clouaire, T., Aguirrebengoa, M., Rocher, V., Pasero, P., Canitrot, Y. & Legube, G. 2018. Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. Nature communications, 9, 533-14.

Delumeau, O., Lecointe, F. O., Muntel, J., Guillot, A., Guédon, E., Monnet, V. R., Hecker, M., Becher, D. R., Polard, P. & Noirot, P. 2011. The dynamic protein partnership of RNA polymerase inBacillus subtilis. PROTEOMICS, 11, 2992-3001.

Dillingham, M. S., Soultanas, P., Wiley, P., Webb, M. R. & Wigley, D. B. 2001. Defining the Roles of Individual Residues in the Single-Stranded DNA Binding Site of PcrA Helicase. Proceedings of the National Academy of Sciences - PNAS, 98, 8381-8387.

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 (Easton), 39, 205-212.

Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. 2003. HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. Journal of the American Chemical Society, 125, 1731-1737.

Epshtein, V., Kamarthapu, V., McGary, K., Svetlov, V., Ueberheide, B., Proshkin, S., Mironov, A. & Nudler, E. 2014. UvrD facilitates DNA repair by pulling RNA polymerase backwards. Nature (London), 505, 372-377.

Eryilmaz, J., Ceschini, S., Ryan, J., Geddes, S., Waters, T. R. & Barrett, T. E. 2006. Structural Insights into the Cryptic DNA-dependent ATPase Activity of UvrB. Journal of molecular biology, 357, 62-72.

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J. & Hassabis, D. 2021. Protein complex prediction with AlphaFold-Multimer. bioRxiv, 2021.10.04.463034.

Fernandes, R. V., Feretzaki, M. & Lingner, J. 2021. The makings of TERRA R-loops at chromosome ends. Cell cycle (Georgetown, Tex.), 20, 1745-1759.

Gao, M., Nakajima An, D., Parks, J. M. & Skolnick, J. 2022. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. Nature communications, 13, 1744-1744.

García-Muse, T. & Aguilera, A. 2016. Transcription-replication conflicts: how they occur and how they are resolved. Nature reviews. Molecular cell biology, 17, 553-563.

Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H. & Ferrin, T. E. 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis: UCSF ChimeraX Visualization System. Protein science, 27, 14-25.

Gomes, P. S. F. C., Gomes, D. E. B. & Bernardi, R. C. 2022. Protein structure prediction in the era of AI: Challenges and limitations when applying to in silico force spectroscopy. Frontiers in Bioinformatics, 2, 983306-983306.

Gorbalenya, A. E. & Koonin, E. V. 1993. Helicases: amino acid sequence comparisons and structure-function relationships. Current opinion in structural biology, 3, 419-429.

Grimm, C., Ficner, R., Sgraja, T., Haebel, P., Klebe, G. & Reuter, K. 2006. Crystal structure of Bacillus subtilis S-adenosylmethionine:tRNA ribosyltransferase-isomerase. Biochemical and biophysical research communications, 351, 695-701.

Guy, C. P., Atkinson, J., Gupta, M. K., Mahdi, A. A., Gwynn, E. J., Rudolph, C. J., Moon, P. B., van Knippenberg, I. C., Cadman, C. J., Dillingham, M. S., Lloyd, R. G. & McGlynn, P. 2009. Rep Provides a Second Motor at the Replisome to Promote Duplication of Protein-Bound DNA. Molecular cell, 36, 654-666.

Gwynn, E. J., Smith, A. J., Guy, C. P., Savery, N. J., McGlynn, P. & Dillingham, M. S. 2013. The conserved C-terminus of the PcrA/UvrD helicase interacts directly with RNA polymerase. PloS one, 8, e78141-e78141.

Halazonetis, T. D., Negrini, S. & Gorgoulis, V. G. 2010. Genomic instability - an evolving hallmark of cancer. Nature reviews. Molecular cell biology, 11, 220-228.

Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T. & Cimprich, K. A. 2017. Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. Cell, 170, 774-786.e19.

Hasanova, Z., Klapstova, V., Porrua, O., Stefl, R. & Sebesta, M. 2022. Human senataxin is a bona fide R-loop resolving enzyme and transcription termination factor. bioRxiv, 2022.08.25.505353.

Hawkins, M., Dimude, J. U., Howard, J. A. L., Smith, A. J., Dillingham, M. S., Savery, N. J., Rudolph, C. J. & McGlynn, P. 2019. Direct removal of RNA polymerase barriers to replication by accessory replicative helicases. Nucleic acids research, 47, 5100-5113.

Herold, M. & nierhaus, K. H. 1987. Incorporation of six additional proteins to complete the assembly map of the 50 S subunit from Escherichia coli ribosomes. The Journal of biological chemistry, 262, 8826-8833.

Hu, G., Katuwawala, A., Wang, K., Wu, Z., Ghadermarzi, S., Gao, J. & Kurgan, L. 2021. flDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions. Nature communications, 12, 4438-4438.

Johansson-Åkhe, I. & Wallner, B. 2022. Improving peptide-protein docking with AlphaFold-Multimer using forced sampling. Frontiers in Bioinformatics, 2, 959160-959160.

Jones, D. T. & Thornton, J. M. 2022. The impact of AlphaFold2 one year on. Nature methods, 19, 15-20.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. 2021. Highly accurate protein structure prediction with AlphaFold. Nature : International weekly journal of science, 596, 583-589.

Keskin, O., Nussinov, R. & Gursoy, A. 2008. Prism: Protein-Protein Interaction Prediction by Structural Matching. Functional Proteomics, 484, 505-521.

Kim, J., Pham, H., Baek, Y., Jo, I., Kim, Y.-H. & Ha, N.-C. 2022. Structure of the plant growth-promoting factor YxaL from the rhizobacterium Bacillus velezensis and its application to protein engineering. Acta crystallographica. Section D, Biological crystallography., 78, 104-112.

Ko, J. & Lee, J. 2021. Can AlphaFold2 Predict Protein-Peptide Complex Structures Accurately? bioRxiv, 2021.07.27.453972.

Krissinel, E. & Henrick, K. 2007. Inference of Macromolecular Assemblies from Crystalline State. Journal of molecular biology, 372, 774-797.

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. & Moult, J. 2019. Critical assessment of methods of protein structure prediction (CASP)—Round XIII. Proteins, structure, function, and bioinformatics, 87, 1011-1020.

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. & Moult, J. 2021. Critical assessment of methods of protein structure prediction (CASP)—Round XIV. Proteins, structure, function, and bioinformatics, 89, 1607-1617.

Lahue, R. S., Au, K. G. & Modrich, P. 1989. DNA Mismatch Correction in a Defined System. Science (American Association for the Advancement of Science), 245, 160-164.

Lang, K. S., Hall, A. N., Merrikh, C. N., Ragheb, M., Tabakh, H., Pollock, A. J., Woodward, J. J., Dreifus, J. E. & Merrikh, H. 2017. Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. Cell, 170, 787-799.e18.

Laurents, D. V. 2022. AlphaFold 2 and NMR Spectroscopy: Partners to Understand Protein Structure, Dynamics and Function. Frontiers in molecular biosciences, 9, 906437-906437.

Leah, E. M., Marcy, E. L. & Steven, W. M. 1999. A Region Near the C-Terminal End ofEscherichia coli DNA Helicase II Is Required for Single-Stranded DNA Binding. Journal of Bacteriology, 181, 2519.

Lee, C., Bo-Han, S. & Tseng, Y. J. 2022. Comparative studies of AlphaFold, RoseTTAFold and Modeller: a case study involving the use of G-protein-coupled receptors. Briefings in bioinformatics, 23.

Lee, J. Y. & Yang, W. 2006. UvrD Helicase Unwinds DNA One Base Pair at a Time by a Two-Part Power Stroke. Cell, 127, 1349-1360.

Lieber, M. R., Yu, K., Chedin, F., Hsieh, C.-L. & Wilson, T. E. 2003. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nature immunology, 4, 442-451.

Machón, C., Lynch, G. P., Thomson, N. H., Scott, D. J., Thomas, C. D. & Soultanas, P. 2010. RepD-mediated recruitment of PcrA helicase at the Staphylococcus aureus pC221 plasmid replication origin, oriD. Nucleic acids research, 38, 1874-1888.

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, 8, 1300-1310.

Marchadier, E., Carballido-López, R., Brinster, S., Fabret, C., Mervelet, P., Bessières, P., Noirot-Gros, M.-F., Fromion, V. & Noirot, P. 2011. An expanded protein-protein interaction network in Bacillus subtilis reveals a group of hubs: Exploration by an integrative approach. Proteomics (Weinheim), 11, 2981-2991.

Matson, S. W. 1989. Escherichia coli DNA helicase II (uvrD gene product) catalyzes the unwinding of DNA.RNA hybrids in vitro. Proceedings of the National Academy of Sciences - PNAS, 86, 4430-4434.

Mechanic, L. E., Hall, M. C. & Matson, S. W. 1999a. Escherichia coli DNA Helicase II Is Active as a Monomer. The Journal of biological chemistry, 274, 12488-12498.

Mechanic, L. E., Latta, M. E. & Matson, S. W. 1999b. A region near the C-terminal end of Escherichia coli DNA helicase II is required for single-stranded DNA binding. Journal of bacteriology, 181, 2519-2526.

Merrikh, C. N., Brewer, B. J. & Merrikh, H. 2015. The B. subtilis Accessory Helicase PcrA Facilitates DNA Replication through Transcription Units. PLoS genetics, 11, e1005289-e1005289.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. & Steinegger, M. 2022. ColabFold: making protein folding accessible to all. Nature methods, 19, 679-682.

Miyazono, K.-I. & Tanokura, M. 2022. New era in structural biology with the AlphaFold program. Translational and Regulatory Sciences, 4, 48-52.

Moreno-Del Álamo, M., Carrasco, B., Torres, R. & Alonso, J. C. 2021. Bacillus subtilis PcrA Helicase Removes Trafficking Barriers. Cells, 10.

Moreno-Del Álamo, M., Torres, R., Manfredi, C., Ruiz-Masó, J. A., Del Solar, G. & Alonso, J. C. 2020. Bacillus subtilis PcrA Couples DNA Replication, Transcription, Recombination and Segregation. Frontiers in molecular biosciences, 7, 140-140.

Mukherjee, S. & Zhang, Y. 2011. Protein-Protein Complex Structure Predictions by Multimeric Threading and Template Recombination. Structure (London), 19, 955-966.

Newing, T. P., Oakley, A. J., Miller, M., Dawson, C. J., Brown, S. H. J., Bouwer, J. C., Tolun, G. & Lewis, P. J. 2020. Molecular basis for RNA polymerase-dependent transcription complex recycling by the helicase-like motor protein HelD. Nature communications, 11, 6420-11.

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. The Journal of biological chemistry, 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. Proceedings of the National Academy of Sciences - PNAS, 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. Molecular genetics and genomics : MGG, 267, 391-400.

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. Molecular microbiology, 29, 261-273.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. & Ferrin, T. E. 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein science, 30, 70-82.

Pohjoismäki, J. L. O., Holmes, J. B., Wood, S. R., Yang, M.-Y., Yasukawa, T., Reyes, A., Bailey, L. J., Cluett, T. J., Goffart, S., Willcox, S., Rigby, R. E., Jackson, A. P., Spelbrink, J. N., Griffith, J. D., Crouch, R. J., Jacobs, H. T. & Holt, I. J. 2010. Mammalian Mitochondrial DNA Replication Intermediates Are Essentially Duplex but Contain Extensive Tracts of RNA/DNA Hybrid. Journal of molecular biology, 397, 1144-1155.

Qi, Y. & Hulett, F. M. 1998. PhoP~P and RNA polymerase σA holoenzyme are sufficient for transcription of Pho regulon promoters in Bacillus subtilis: PhoP~P activator sites within the coding region stimulate transcription in vitro. Molecular microbiology, 28, 1187-1197.

Richard, P. & Manley, J. L. 2017. R Loops and Links to Human Disease. Journal of molecular biology, 429, 3168-3180.

Rocha, E. P. C., Danchin, A. & Viari, A. 1999. Translation in Bacillus subtilis: roles and trends of initiation and termination, insights from a genome analysis. Nucleic acids research, 27, 3567-3576.

Runyon, G. T., Wong, I. & Lohman, T. M. 1993. Overexpression, purification, DNA binding, and dimerization of the Escherichia coli uvrD gene product (Helicase II). Biochemistry (Easton), 32, 602-612.

Sanders, K., Lin, C.-L., Smith, A. J., Cronin, N., Fisher, G., Eftychidis, V., McGlynn, P., Savery, N. J., Wigley, D. B. & Dillingham, M. S. 2017. The structure and function of an RNA polymerase interaction domain in the PcrA/UvrD helicase. Nucleic acids research, 45, 3875-3887.

Sidorenkov, I., Komissarova, N. & Kashlev, M. 1998. Crucial Role of the RNA:DNA Hybrid in the Processivity of Transcription. Molecular cell, 2, 55-64.

Singh, R., Devkota, K., Sledzieski, S., Berger, B. & Cowen, L. 2022. Topsy-Turvy: integrating a global view into sequence-based PPI prediction. Bioinformatics (Oxford, England), 38, i264-i272.

Singleton, M. R., Dillingham, M. S. & Wigley, D. B. 2007. Structure and mechanism of helicases and nucleic acid translocases. Annual review of biochemistry, 76, 23-50.

Skourti-Stathaki, K., Proudfoot, Nicholas J. & Gromak, N. 2011. Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. Molecular cell, 42, 794-805.

Soultanas, P., Dillingham, M. S. & Wigley, D. B. 1998. Escherichia coli ribosomal protein L3 stimulates the helicase activity of the Bacillus stearothermophilus PcrA helicase. Nucleic acids research, 26, 2374-2379.

Soultanas, P., Dillingham, M. S., Wigley, D. B., Papadopoulos, F., Phillips, S. E. V. & Thomas, C. D. 1999. Plasmid replication initiator protein RepD increases the processivity of PcrA DNA helicase. Nucleic acids research, 27, 1421-1428.

Stark, H. 2010. GraFix: Stabilization of Fragile Macromolecular Complexes for Single Particle Cryo-EM. Methods in Enzymology, 481, 109-126.

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

Thomas, M., White, R. L. & Davis, R. W. 1976. Hybridization of RNA to Double-Stranded DNA: Formation of R-Loops. Proceedings of the National Academy of Sciences - PNAS, 73, 2294-2298.

Urrutia-Irazabal, I. 2021. The regulation of transcription by helicases. University of Bristol.

Urrutia-Irazabal, I., Ault, J. R., Sobott, F., Savery, N. J. & Dillingham, M. S. 2021. Analysis of the PcrA-RNA polymerase complex reveals a helicase interaction motif and a role for PcrA/UvrD helicase in the suppression of R-loops. eLife, 10.

Vakser, Ilya A. 2014. Protein-Protein Docking: From Interaction to Interactome. Biophysical journal, 107, 1785-1793.

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Žídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D. & Velankar, S. 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic acids research, 50, D439-D444.

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. The EMBO journal, 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.

Webster, M. P. J., Jukes, R., Zamfir, V. S., Kay, C. W. M., Bagnéris, C. & Barrett, T. 2012. Crystal structure of the UvrB dimer: insights into the nature and functioning of the UvrAB damage engagement and UvrB-DNA complexes. Nucleic acids research, 40, 8743-8758.

Wolak, C., Ma, H. J., Soubry, N., Sandler, S. J., Reyes-Lamothe, R. & Keck, J. L. 2020. Interaction with single-stranded DNA-binding protein localizes ribonuclease HI to DNA replication forks and facilitates R-loop removal. Molecular microbiology, 114, 495-509.

Xiong, H., Han, L., Wang, Y. & Chai, P. 2022. Evaluating the Reliability of AlphaFold 2 for Unknown Complex Structures with Deep Learning. bioRxiv, 2022.07.08.499384.

Yancey, J. E. & Matson, S. W. 1991. The DNA unwinding reaction catalyzed by Rep protein is facilitated by an RHSP-DNA interaction. Nucleic acids research, 19, 3943-3951.

Yang, X. & Lewis, P. J. 2008. Overproduction and purification of recombinant Bacillus subtilis RNA polymerase. Protein expression and purification, 59, 86-93.

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

Yin, R., Feng, B. Y., Varshney, A. & Pierce, B. G. 2022. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. Protein science, 31, e4379-n/a.

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. Journal of molecular biology, 371, 336-348.