Regulation of Transcription through the Secondary Channel of RNA Polymerase

Amber Riaz-Bradley

Centre for Bacterial Cell Biology
Institute for Cell and Molecular Biosciences

Thesis submitted for the degree of Doctor of Philosophy
September 2018

Abstract

The bacterial transcription factor Gre performs the highly conserved function of stimulating the endonucleolytic activity of RNA polymerase for efficient RNA cleavage, thereby promoting transcription elongation and assisting in transcript fidelity. This dynamic factor-mediated cleavage has been extensively studied, except for within the unusual Cyanobacteria where, notably, no Gre homologues have yet been identified.

To investigate this apparent absence of Gre factor, the RNA polymerase of two cyanobacterial species, *Synechococcus elongatus* 7942 and *Synechocystis sp* 6803, were purified and tested for their *in vitro* transcription activity. Rates of intrinsic RNA cleavage were 20-90x greater for the cyanobacterial polymerases than rates found for *Escherichia coli*. Further study revealed differences in the bridge helix and trigger loop structural elements of the active site as the possible cause of this increased activity. Mutational analyses indicate a reduced flexibility of these elements which may fix the active site into a closed and more hydrolytically competent state.

We propose that in cyanobacteria, the lack of Gre factor is compensated by the unique composition and endogenous ability of the polymerase itself to perform fast and efficient transcript cleavage thus eliminating the need for additional factors. In this work a Gre factor homologue, Gfh1, of *Thermus aquaticus* is also examined. Gfh1 is known to stimulate transcriptional pausing at a wide variety of pause signals and we present further evidence of preferential activity towards the inhibition of transcription from a pre-translocated state.

Acknowledgments

The work described in this thesis would not have been possible without the support and assistance of the following people, to whom I would like to express considerable gratitude:

- Dr Yulia Yuzenkova, without whom I would have never had the opportunity to undertake this research. Under her supervision I have learned much; about science, about 'good old times', and the excitement of a fun side project. Thank you for your patience and guidance throughout this project.
- All members of the Yuzenkova and Zenkin lab groups for their support, vast knowledge of the research area, and seemingly unlimited ability to supply snacks
- Past and present members of ICAMB and the CBCB as their welcoming attitudes and technical competence make this building a great place to work
- For specific contributions: Professor Conrad Mullineaux and Professor Nigel Robinson for the donation of the cyanobacterial strains. Dr Pamela Gamba for the donation of several *E. coli* strains used throughout this thesis. Professor Irina Artsimovitch for the donation of the pVS10, pVS14, pRL663 and pIA349 plasmids. Dr Katherine James, Dr Achim Treumann, and Dr Kevin Waldron for respective RNAseq, Mass spectrometry, and ICP-MS analysis. Dr Flint Stevenson-Jones for the donation of the his-tagged *E. coli* GreA construct. Dr Soren Nielsen for always ensuring smooth running of the lab. The BBSRC for funding this project.

Declarations

a) I declare that this thesis is my own work and that I have correctly acknowledged the work of others. This submission is in accordance with University and School guidance on good academic conduct.

b) I certify that no part of the material offered has been previously submitted by me for a degree or other qualification in this or any other University.

c) I confirm that the word length is within the prescribed range as advised by my school and faculty

d) Does the thesis contain collaborative work, whether published or not?
Yes. Data obtained by others is clearly acknowledged in the thesis.

Signature of candidate: A. Riaz-Bradley

Date 18/09/2018

Table of Contents

List of Figures

Introduction

Results and Discussions

List of Tables

List of Abbreviations

ATP – Adenosine triphosphate

BG-11 – Blue-Green medium 11

BH – Bridge Helix

Bp – base pair

CTP – Cytosine triphosphate

DNA – Deoxyribonucleic acid

dNTP – Deoxy-nucleoside triphosphate

DPBB – Double psi beta barrel

EC – Elongation complex

*Ec*RNAP – *E. coli* RNAP

EDTA – Ethylenediaminetetraacetic acid

GTP – Guanosine triphosphate

ITPG – Isopropyl β-D-1-thiogalactopyranoside

kb – kilo base

kDa – kilo dalton

LB – Luria-Bertani medium

MEC – Mis-incorporated elongation complex

mRNA – messenger RNA

NAC – Nucleotide addition cycle

ncRNA – non-coding RNA

NMP – nucleoside monophosphate

nt – nucleotide

NTP – nucleoside triphosphate

PAGE – Polyacrylamide gel electrophoresis

PNK – Polynucleotide kinase

PPi – Pyrophosphate

RNA – Ribonucleic acid

RNAP – RNA polymerase

rNTP – Ribonucleoside triphosphate

S.*el* – *Synechococcus elongatus* PCC 7942

S.sp – *Synechocystis sp* PCC 6803

SB – Storage buffer

SCBFs – Secondary Channel binding factors

SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis

*Sel*RNAP - *Synechococcus elongatus* PCC 7942 RNAP

*Ssp*RNAP – *Synechocystis sp* PCC 6803 RNAP

T.aq – *Thermus aquaticus*

*Taq*RNAP – *Thermus aquaticus* RNAP

TH – Trigger Helices

TL – Trigger Loop

UTP – Uridine triphosphate

WT – Wild-type

1. Introduction

1.1 Thesis Scope

Investigation into the regulation of gene expression is a continuing concern within the current climate of antibiotic resistance, environmental change, and need for production of high value products. Interest in cyanobacterial gene expression is therefore multiplied as the diverse cyanobacteria are capable of providing key information or possible solutions for each of these issues.

Adaptation to variation in environmental conditions relies heavily on the ability of an organism to modify and regulate gene expression. The photosynthetic cyanobacteria are one of the most versatile groups and inhabit nearly all aquatic and terrestrial environments. They are responsible for much of the marine oxygen production and share similarities with their chloroplast descendants. Their wide distribution and diversity has resulted in the identification of a number of useful secondary metabolites, and the cyanobacteria are also very amenable for use in industrial production. A greater understanding of the mechanisms employed by cyanobacteria during gene expression would allow better manipulation of this production.

Transcription, the first stage of gene expression, is carried out by RNA polymerase. There are a number of atypical features employed in cyanobacterial transcription. The RNA polymerase of this organism has an extra subunit, γ, caused by a split of the *rpoC* gene and also encodes a large lineage-specific insertion within the β' subunit. Furthermore, cyanobacteria are unusual as the highly conserved GreA factor is absent. This proofreading RNA polymerase accessory factor greatly enhances hydrolysis of the RNA transcript in the event of a mis-incorporated nucleotide or stalled complex, allowing the mistake to be removed and the enzyme to be reset. The only other groups lacking GreA homologues are the small taxons of *Aquificaceae*, *Dictyoglomaceae*, and *Fusobacteriaceae*. It is currently unknown how this loss of GreA is compensated for.

This thesis attempts to characterise the *in vitro* transcription of cyanobacterial RNAP with particular regard to the absence of the proof-reading GreA factor and the effects of this absence on the hydrolysis and fidelity of the nascent RNA transcript.

1.2 Cyanobacteria

Cyanobacteria are one of the largest groups of gram-negative prokaryotes, with exceptional variation displayed among roughly 2700 described species (Nabout *et al*., 2013). Members of this taxon are capable of inhabiting nearly all aquatic and terrestrial environments and encompass both unicellular and filamentous morphology (Kulasooriya, 2011). Thought to have been the first group to develop oxygenic photosynthesis, cyanobacteria have played a central role in the history of our planet and its atmosphere.

Between 3.8 and 2.8 billion years ago, the arrival of oxygenic photosynthesis in cyanobacteria led to the gradual increase in atmospheric O_2 (Schopf, 1993). Initially, this oxygen was able to be neutralised by iron deposits and the methane-rich environment of Earth (Daines and Lenton, 2016). However, the competitive energy advantage provided by oxygenic photosynthesis, along with the toxic nature of oxygen to the anaerobes of the time, resulted in an overall gain of oxygen and ozone UV shield allowing conversion towards an aerobic atmosphere (Olsen, 2006; Blaustein, 2016). This transformation, termed the Great Oxidation Event, is believed to have occurred approximately 2.45 billion years ago. A thinner and less-insulating atmosphere ensued which launched Earth's longest ice age, the Huronian glaciation (Kopp *et al*., 2005). In the time following, adapted aerobes became the predominant organisms leading to the emergence and subsequent evolution of protists and higher life forms (Och and Shields-Zhou, 2012; Van der Geizen and Lenton, 2012).

Cyanobacteria were again of great importance roughly 1.5 billion years after this glaciation event in the formation of chloroplasts. Mereschkowsky (1905) was the first to hypothesise that chloroplasts were derived from cyanobacteria as a result of an endosymbiotic event. It is now thought that the primary endosymbiosis, where a eukaryotic ancestral cell engulfed and integrated a cyanobacterial-like prokaryote, gave rise to three major lineages: glaucophytes, rhodophytes (red algae), and the direct ancestors of all modern plants, chloroplastida (green algae) (Gould *et al.*, 2008; Ball *et al*., 2011). Upon gradual progression to a true organelle, more than 90% of the cyanobacteria-derived genome of chloroplasts have either been lost or undergone endosymbiotic gene transfer to the nuclear genome of the host (Timmis *et al*., 2004). Key proteins involved in oxygenic photosynthesis, translation, and the division of the symbiont make up the majority of retained genes (Daniell *et al*., 2016). Recent phylogenetic analyses of these nuclear and plastid-encoded genes by Ponce-Toledo *et al.* (2017) propose a freshwater cyanobacterium, *Gloeomargarita lithophora*, as being the closest living relative able to most accurately represent the primary chloroplast endosymbiont.

Modern day cyanobacteria continue to perform vital functions. Oceanic phytoplankton, which include members of cyanobacteria, lead production of total atmospheric oxygen at an estimated 50-70% (Sekerci and Petrovskii, 2015; Harris, 1986). Nitrogen-fixing cyanobacteria help in the maintenance of soil fertility and enhancement of agricultural production, especially important considering the predicted food requirements and limited land resources of our growing population (Singh *et al*., 2016). The wide distribution and diversity of cyanobacteria has resulted in a number of secondary metabolites of ecological, medicinal and industrial interest which are reviewed by Mazard *et al*. (2016). Cyanobacteria are also widely amenable for use in industrial chemical production, particularly in the production of sustainable biofuels and biodegradable plastics (Troschl *et al*., 2017; Nozzi *et al*., 2013).

The regulation of gene expression in cyanobacteria is therefore of particular interest. A greater understanding of the mechanisms employed during transcription would allow better manipulation for industrial production. The transcription machinery of two major cyanobacterial model species are studied in this thesis, *Synechococcus elongatus* PCC 7942 (**S.*el***) and *Synechocystis sp*. PCC 6803 (***S.sp***). Both are freshwater, polyploid, unicellular cyanobacteria with fully sequenced genomes. The genome of *S.sp* was the first cyanobacterial genome to be fully sequenced and annotated, whereas *S.el* was the first cyanobacterium to show reliable transformation by exogenous DNA (Kaneko *et al*., 1996; Shestakov and Khyen, 1970). The genome for *S.el* is roughly 2.7 Mb comprised of a single chromosome and one plasmid encoding 2,657 proteins in total. The genome of *S.sp* is slightly larger with a single chromosome and four endogenous plasmids at 3.57 Mb, and a total protein count of 3,507. Each carry a complete set of genes for oxygenic photosynthesis, and while some strains can use alternate carbon sources such as glucose for growth, others like *S.el* strictly require photosynthetically derived energy and are known as obligate photoautotrophs (McEwen *et al*., 2013). Primarily due to interest in this photosynthetic machinery and natural competency, many methods have now been established for the study of their molecular biology. This is particularly true in the area of synthetic biology, where the ability of cyanobacteria to respond to environment signals such as light intensity or CO_2 concentration make them attractive as engineered biosensors (Sun *et al*., 2018).

1.3 Transcription in Bacteria

The first stage of gene expression is carried out by DNA-dependent RNA polymerase (**RNAP**) which produces messenger or non-coding RNA (**m/ncRNA**) from a DNA template in the cyclical process of transcription. This primary RNA transcript can then be used as a template for protein synthesis or perform regulatory roles in the case of most ncRNA. Transcription by RNAP is a repeated multi-stage process consisting of initiation, promoter escape, elongation, and termination.

1.3.1 The Transcription Cycle

Initiation of transcription begins with binding of RNAP holoenzyme to promoter regions upstream of the intended gene or operon to be transcribed. Holoenzyme is formed through binding of RNAP with a dissociable sigma factor which allows recognition and binding of consensus DNA elements such as -35 TTGACA and -10 TATAAT (Harley and Reynolds, 1987; Gruber and Gross, 2003).

Sigma factors are promoter-specific and the number of different sigma factors available varies greatly between species. They are organised into two evolutionarily distinct families, σ^{70} and σ^{54}, based on structure and sequence homology (Merrick, 1993). The principal sigma factors, RpoD/σ^{70} in *E. coli* and SigA/RpoD1 in cyanobacteria, use variations of the motifs above to initiate transcription of the housekeeping genes essential for cell viability (Imamura and Asayama, 2009). These factors can tolerate considerable sequence divergence from the consensus whereas alternative sigma factors, which tend to promote transcription of very specific regulons, exhibit a very high stringency for precise promoter sequences (Feklistov and Darst, 2009). Cyanobacteria only encode σ^{70}-type sigma factors, and these are further divided into four groups (Paget and Helmann, 2003). Group 1 comprises the essential primary sigma factor for transcribing exponential growth phase genes. Group 2 factors also use the above consensus promoters but are non-essential (Goto-Seki *et al*., 2002). *E. coli* possesses only one group 2 factor, σ^{S}, thought to be a master regulator of stress response while cyanobacteria typically have at least three and up to seven (Osanai *et al*., 2008). It is also common to have multiple alternative-type group 3 and group 4 factors which direct transcription of environmental-adaptation genes such as those for motility, sporulation, iron transport, and further stress responses (Paget, 2015; Huckauf *et al*., 2000).

Cyanobacteria are the only known species able to perform oxygenic photosynthesis and to synchronise cell activities with the rotation of Earth in a circadian rhythm (Johnson, 2007). Light-induced gene expression is achieved through the circadian KaiABC clock proteins which transmit environmental light signals to response regulators through the

transfer of phosphorylation status (Bell-Pedersen *et al*., 2005). Markson *et al*. (2013) found phosphorylation of the master response regulator RpaA results in the activation of some group 2 sigma factors and the *KaiBC* operon itself, creating a positive feedback loop. Altered expression of the group 2 sigma factors or RpaA can interfere with downstream circadian gene expression and dysregulate circadian timing (Nair *et al*., 2002; Puszynska and O'Shea, 2017; Diamond *et al*., 2017). It is currently suggested that different members of the group 2 cyanobacterial sigma factors have important roles in activating transcription at different points during the light cycle (Summerfield and Sherman, 2007). Access to the DNA is also affected by genomic compaction and DNA supercoiling which adds further layers of regulation to the transcription process of cyanobacteria (Vijayan *et al*., 2009; Smith and Williams, 2006).

Once the holoenzyme (RNAP + σ) has bound the promoter region, the DNA duplex is melted from the A/T-rich -10 element to the transcription start site to form the transcription bubble. The separated DNA strands are held within the main channel of RNAP to position the +1 transcription start site of the template strand at the active center for transcription initiation. RNAP tends to be retained at the promoter region due to the strong binding of σ, resulting in cycles of abortive initiation (Samanta and Martin, 2013). Short RNA transcripts are transcribed before the productive elongation phase. A conformational strain known as DNA scrunching is produced by the accumulation of unwound DNA within the RNAP which cannot move forward (Revyakin *et al*., 2006). Escape of RNAP from the promoter requires energy generated by this stressed DNA and the synthesis of a nascent RNA of a significant size (Kapanidis *et al*., 2006; Dulin *et al*., 2018). Full dissociation of σ is common but not necessary for productive elongation, and either continued binding or re-attachment of σ may permit further regulatory activities such as promoter- proximal pausing (Mooney *et al*., 2005).

During the elongation phase, the RNAP extends the RNA transcript using another repeated process known as the nucleotide addition cycle. RNAP is able to maintain strong contact with the DNA, moving the transcription bubble along the duplex while preserving DNA strand separation and a 9-10 bp RNA:DNA hybrid positioned at the active site, allowing highly efficient and processive elongation until a termination signal is encountered (Zuo and Steitz, 2017). The main focus of this thesis revolves around the elongation phase of transcription, and so the specific activities of RNAP in NTP incorporation and the regulatory mechanisms at this stage are discussed in more detail in the following chapters.

Terminator signals cause the RNA transcript to be released from the RNAP. There are two major types of termination signal in bacteria; Rho-dependent and Rho-independent termination. Rho is an ATP-dependent helicase, essential in *E. coli*, which binds to a Rho-binding site on the nascent mRNA and dissociation of the elongating complex with full release of the RNA transcript (Ciampi, 2006). Although roughly half of all transcription terminators identified in *E. coli* are Rho-dependent, no full Rho homologue has yet been identified in cyanobacteria (de Hoon *et al*., 2005). Rho-independent or intrinsic termination depends on a palindromic sequence encoded at the terminator which results in a hairpin structure in the transcript with a strong G:C stem followed by a stretch of consecutive U bases (Rédei, 2008; Farnham and Platt, 1981). RNAP stalls due to the poly-U region which gives the hairpin time to form and block the RNA exit channel, subsequent displacement of the transcript from the DNA:RNA hybrid and weak Watson-Crick base pairing of the U stretch result in dissociation of the RNAP (Peters *et al*., 2011). Vijayan *et al*. (2011) found Rho-independent termination to be widespread in cyanobacteria, along with occasional incomplete termination events. Interestingly, inefficient termination is also a recognised feature of chloroplast gene expression. It was thought that in chloroplasts, the intrinsic termination signals are used mainly as structures to stabilise the transcript for further RNA processing by additional factors (Stern and Gruissem, 1987). In 2014, Chi *et al*. described Rho-like termination activity for the RNA-binding RHON1 chloroplast protein. RHON1 relies on an ATPase domain with a completely distinct structure from that of Rho, and binds sites on the RNA transcript which are sparse in C residues in comparison to the binding sites for Rho which tend to be enriched with C (Chi *et al*., 2014). The mode of action is therefore distinctive and not identical to that of Rho. They also suggest RHON1 may only found in the dicot branch of flowering plants. A BLAST search for RHON1 in cyanobacteria returns only 15 species. The proteins aligned are much shorter than RHON1, mostly described as hypothetical, and have roughly 50% identity aligning to the N-terminal domain of Rho. Termination is the least understood transcriptional stage, particularly in cyanobacteria, and further study into these potential Rho-like proteins may prove useful in future application to synthetic engineered systems.

1.3.2 The Transcription Machinery

All cellular organisms use multi-subunit RNAPs to carry out the transcription cycle. Bacterial RNAP has a typical core subunit composition of $\alpha_2\beta\beta'\omega$ encoded by the *rpoA*, *rpoB*, *rpoC*, and *rpoZ* genes, respectively (Minakhin *et al*., 2001). There is a great deal of both structural and sequence conservation between bacterial RNAPs and the RNAPs of archaea and eukaryotes, suggesting divergence from a common ancestor. This has been demonstrated in a variety of published RNAP crystal structures from all three domains (Zhang *et al*., 1999; Korkhin *et al*., 2009; Cramer, 2002). Bacterial RNAP represents the 'core component', or the simplest form required for the conserved mechanism of RNAP, with homologues for each of the bacterial subunits identified in archaea and eukaryotes (Sweetser *et al*., 1987; Ebright, 2000; Lane and Darst, 2009). Transcription in archaea and bacteria also relies on just a single RNAP complex whereas eukaryotes can have multiple multi-subunit RNAPs, each specialised for the synthesis of a particular type of RNA transcript. The RNAPs of the archaea-eukaryote lineage manifest increased complexity through additional subunits, many of which have functions carried out by general transcription factors for other RNAPs (Carter and Drouin, 2010). Structural features of RNAP are highlighted in **Figure 1**.

Assembly of bacterial RNAP first involves formation of a homodimer by two α subunits via the N-terminal domains (**NTD**) which then recruit the β subunit to form $\alpha_2\beta$, the αNTD does not participate in catalysis and functions solely as an aid for assembly. However, the C-terminal domains (**CTD**) of the α dimer have important roles in regulating the interaction with σ factor and holoenzyme recruitment to the promoter region (Chen *et al*., 2003). The ω subunit forms a β'ω complex by simultaneously binding the NTD and CTD of the β' which is thought to promote formation of RNAP by preventing aggregation of the separate parts (Ghosh *et al*., 2001). Although ω appears to be a non-essential subunit under normal conditions, deletion mutants have identified altered growth phenotypes and a potential non-structural role in the regulation of RNAP to the alarmone ppGpp (Vrentas *et al*., 2005; Mathew *et al*., 2006). Gunnelius *et al*. (2014a, b) found ω to be essential in cyanobacteria when under heat stress, and also for the efficient recruitment of σ factor to form holoenzyme. The β and β' subunits form the bulk of the enzyme and make extensive interactions with each other to shape the RNAP active site. β and β' are comprised of a number of distinct domains which provide the elements for catalytic activity. Full assembly is then achieved upon joining of the two sub-assemblies $\alpha_2\beta + \beta'\omega$ to give an approximate RNAP molecular mass of 400 kDa (Ishihama, 1981; Minakhin *et al*., 2001).

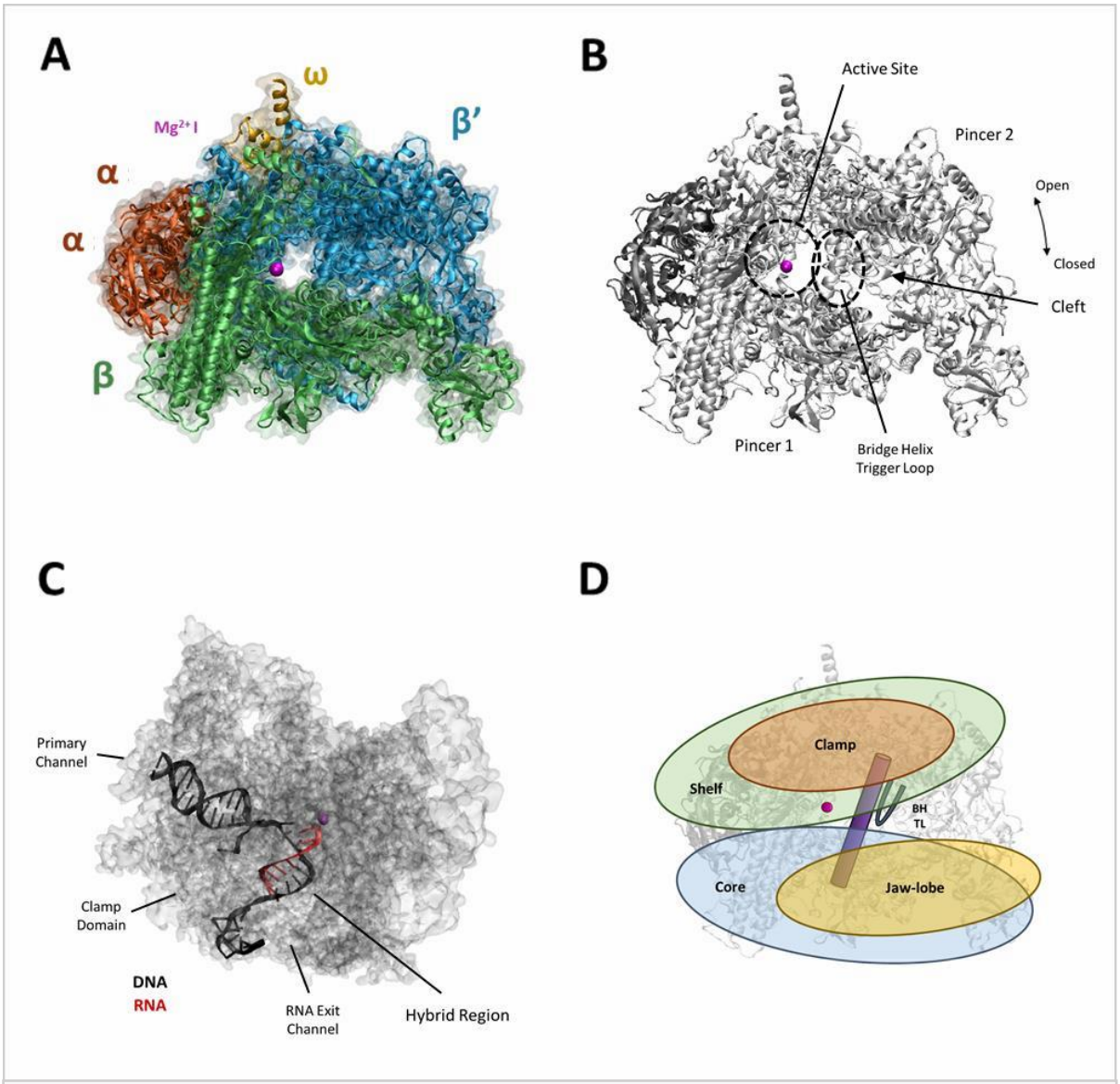


Figure 1. Structural Overview of Bacterial RNAP. (A) *E. coli* RNAP subunit composition and arrangement. **(B)** Schematic of RNAP with the major structural and catalytic elements labelled. **(C)** Space fill of RNAP showing positioning and separation of the nucleic acids, and RNA:DNA hybrid with the 3'-OH of RNA at the active site ready to receive an incoming NTP. **(D)** Approximate division of the four mobile RNAP modules, with the active-site spanning bridge helix (BH) and catalytic trigger loop (TL) components. Adapted from Sekine *et al*., 2015 and Tagami *et al*., 2010. All structures are *E. coli* RNAP, made using data of an elongating complex from PDB 6ALH (Kang *et al*., 2017).

Bacterial RNAP has an overall 'crab claw'-like structure split by a positively charged cleft for nucleic acid binding, with the largest subunits β and β' each providing a 'pincer'. The general architecture can be divided into four mobile modules formed from a combination of subunits; the core, the shelf, the clamp, and the jaw-lobe (Zhang *et al*., 1999; Cramer *et al*., 2001). The core and shelf modules connect to form the central part of the claw and shape the primary, secondary and RNA exit channels. These two modules are linked by a long α-helix, the bridge helix, which spans the RNAP cleft and is next to the deep catalytic site formed at the junction of the pincers. The core and shelf modules can shift relative to each other to open the active site and affect conformation of the nucleic acids in RNAP transcription bubble (Sekine *et al*., 2015). The clamp and jaw-lobe modules extend from the core and shelf modules to form the claw pincers. These modules also complete the primary or main channel, through which the nucleic acids are positioned at the active site. The jaw-lobe helps guide the DNA through the RNAP whereas the clamp has three loops which extend into the active site. These loops, known as the rudder, lid, and zipper, help facilitate separation of DNA:RNA hybrid to direct the nascent RNA transcript through the RNA exit channel (Gnatt *et al*., 2001). The secondary channel provides access to the active site for incoming NTPs and is also relied on heavily during transcription as a site of regulation through the action of secondary channel binding factors (Nickels *et al*., 2004).

The catalytic subunits of RNAP are β and β' encoded by the *rpoB* and *rpoC* genes. These two genes tend to be part of an operon in many bacteria and are synthesised as a single transcription unit. They are thought to be the result of an ancient duplication event where each subunit subsequently diverged and developed specialised functions (Iyer *et al*., 2003). In cyanobacteria, however, *rpoC* has undergone a split into two parts, an N-terminal γ subunit (***rpoC1***) and a C-terminal β′ subunit (***rpoC2***) (Schneider *et al*., 1987; Xie *et al*., 1989). Although no published cyanobacterial RNAP structures are available, the high degree of conservation between multi-subunit RNAPs from all three domains suggests a similar 'crab claw'-like structure is likely. Pollari *et al*. (2008) constructed models for the RNAP holoenzyme of *S.sp* using a *Thermus thermophilus* published structure as a template. In these models, the arrangement of RNAP is comparable to that of other bacterial RNAP, with two recognisable 'pincers' which in this instance, are formed from the β and γ+β' subunits. This split of the *rpoC* gene must have occurred before the endosymbiotic event which gave rise to chloroplasts, as two subunits are also employed in chloroplast RNAP which correspond to the N and C-terminal domains of the usually singular bacterial *rpoC* gene. Chloroplasts rely on two RNAPs, a nuclear encoded polymerase (**NEP**) of the host genome and the cyanobacterial descended chloroplast/plastid-encoded polymerase (**PEP**) both of which are required (Smith

and Purton, 2002). The terminology used in PEP is different to that for cyanobacterial RNAP, however, as chloroplast *rpoC1* is termed the β' subunit and *rpoC2* is β''. It is interesting that PEP has been retained when so many other cyanobacterial-derived genes have been lost or transferred to the nuclear genome (Timmis *et al*., 2004). Both PEP and NEP are essential for chloroplast transcription and both appear able to transcribe the complete chloroplast genome. The only gene which undergoes NEP-exclusive transcription is PEP *rpoB*, which may be some form of quantity regulation since PEP is the predominating RNAP in green leaves (Börner *et al*., 2015).

The sequences of the large β and β' (or γ+β') subunits are highly conserved among the multi-subunit RNAPs of all three domains of life and encode a number of distinct elements which assist in catalysis (Jokerst *et al*., 1989; Allison *et al*., 1985). The availability of sequence data from an increasing variety of species has allowed further refinement of these conserved element regions (Zhang *et al*., 1999; Lane and Darst, 2010a). Differences between the *E. coli* β' and the cyanobacterial γ+β' feature prominently in this thesis and these conserved regions are depicted in a linear map along with the structural domains they form in **Figure 2**.

Using nomenclature from Lane and Darst (2010a), cyanobacterial γ maps to the N-terminal sections b1-b5 of single bacterial β', leaving the cyanobacterial β' subunit to align to the C-terminal b6-b11 regions of conservation. Regions b1 to b3 inclusive form the majority of the clamp module with the zipper encoded in b1, and both the lid and rudder in section b3. Part of the clamp is also encoded by section b11. An evolutionarily conserved Zn-ribbon (ZNR) is found in b1 which has an important role in maintaining the structural stability of RNAP (Chanfreau, 2013). β-β' module, BBM, domains form a curved hairpin structure restricted by two α-helices, however, the function they provide is not currently understood. It has been suggested that they work in protein:protein interactions with transcriptional regulatory factors (Opalka *et al*., 2010). BBM1 is found after the ZNR and notably inserts itself into the conserved domain of the β subunit which forms part of the active site (Iyer *et al*. 2004). Two conserved AT-hook like modules are at the end of region b3 which are small positively charged flap structures used in recognition and binding of nucleic acids (Iyer *et al*., 2003). Regions b4 and b5 make up the precise site of catalysis in RNAP where NTPs are incorporated into the RNA transcript. These regions are further examined in the next section but encode a number of critical residues which contact the DNA, RNA, and NTP, and also chelate a Mg^{2+} ion required for catalysis. NTPs and accessory transcription factors access the active site of RNAP through the secondary channel. The secondary channel rim helices are the binding platform for these external transcription factors and are formed from the common

b6 region. The *rpoC* split in cyanobacteria is also in the vicinity of this region which may form part of the reason for the absence of secondary channel binding factors, i.e. GreA, in these organisms. Two integral active site components known as the bridge helix and the trigger loop are denoted by regions b7 and b8-9, respectively. These two mobile elements are recurrently discussed in later chapters. In between these lies the conserved BBM2. Again, the specific function is not understood, however, Iyer *et al*. (2004) has identified several BBM2 variations in other proteins, including in cyanobacteria where a BBM2-type domain has fused to the C-terminal of Ferredoxin which also suggests protein:protein or regulatory interactions. After the trigger loop, are the final conserved sequences of b10 and b11 which provide further structural elements for the jaw and clamp modules (Lane and Darst, 2010b).

Comparison of the RNAP sequences not only led to detailed recognition of conserved domains, but also identified several regions which had undergone non-conserved insertion events characteristic of different evolutionary lineages (Zhang *et al*., 1999; Iyer *et al*., 2003). These insertion sequences are found towards the surface of the enzyme, away from the highly conserved active site, and constitute the main differences between the RNAP of different bacterial species. The largest and most studied of these insertions occur in the β and β' subunits of proteobacteria, known as sequence insertions 1, 2, and 3 **(SI3)**. SI1 and SI2 are found within the β subunit whereas SI3 is inserted in the middle of the highly conserved trigger loop element of β' (Artsimovitch *et al*., 2003) (also in **Figure 2**).

Lane and Darst (2010a) have since been able to identify several smaller lineage-specific insertions in bacteria using alignment of a greater number of β and β' sequences to give a current total of 12 β inserts and 7 β' inserts (not shown). Inserts differ in size, sequence, and location but are all predicted to be surface-exposed which the authors suggest grant species-specific properties on core RNAP. Artsimovitch *et al*. (2003) likens the pattern of surface-exposed inserts to that of the additional surface subunits of eukaryotic RNAPs which while dispensable, modulate RNAP activity and response to regulatory factors. The SI1 and SI2 can tolerate large alterations and even deletion without affecting core function. In addition, SI1 has been discovered to regulate activity through the action of termination factor Alc, through clamping of downstream duplex DNA, and by acting in concert with the β' rim helices as the binding site for the transcription factor DksA (Parshin *et al*., 2015; Saecker *et al*., 2011; Severinov et al., 1994). The function of SI2 is still largely unknown but it appears to play a role in RNAP conformation changes during pausing (Kang *et al*., 2018). Alterations of the large β'SI3 of *E. coli*, however, are much more detrimental to RNAP activity, likely due to its position within the integral trigger loop (Zakharova *et al*., 1998). The SI3 of the

cyanobacterial lineage is unique in that it is much larger at >600 amino acids and harbours at least seven sandwich-barrel-hybrid-motifs (**SBHM**) as opposed to the two encoded in *E. coli* SI3 (Iyer *et al*., 2003). Several suggestions have been made for the function of these motifs and the role of the SI3 in general which are discussed in the next section.

These SI regions also show only slightly more sequence variability between related bacteria than the remaining subunit regions, further suggesting a functional role. The SIs display 60-70% conservation between *E. coli* and other γ-proteobacteria whereas conservation of the remaining regions is >80% (Artsimovich *et al*., 2003). The total length of β' or γ+β' is 1407 amino acids for *E. coli* and 1942/1943 for *S.el* and *S.sp* respectively. Sequence conservation (excluding SI3) between these organisms is ~42% for identical positions and ~70% if including positions with similar amino acids. Comparison of the subunit as a whole shows a small increase in negatively-charged residues to ~15.5% and a decrease in positively-charged residues to ~12% for cyanobacteria over the equivalent 13.5% and 13.2% for *E. coli*. These percentages also hold true when calculated without the SI3. This is reflected in the lower theoretical pI for the combined cyanobacterial subunits of 5-5.1 as opposed to 6.8 for *E. coli* β'. The β' subunit is unstable and prone to aggregation, a lower pI further from the intracellular pH may assist in maintaining solubility (Ghosh *et al*., 2001). Additional differences in composition between cyanobacterial and *E. coli* subunits include increases of at least 1% in aspartic acid and glutamine, with decreases in alanine, lysine, and leucine. These composition differences were calculated excluding SI3 and so represent differences in the predominantly conserved regions of the subunits. Once a crystal structure has been established for cyanobacterial RNAP, it would be interesting to find out whether these composition differences in amino acids are grouped together, form surface exposed pockets, or are in functionally relevant areas of the polymerase. Sequence conservation analysis of just the SI3 domains of both cyanobacterial strains gave 60.1% identity, consistent with the percentages of SI conservation between other related bacteria but also indicating a potential variability in function or binding partners of this insert.

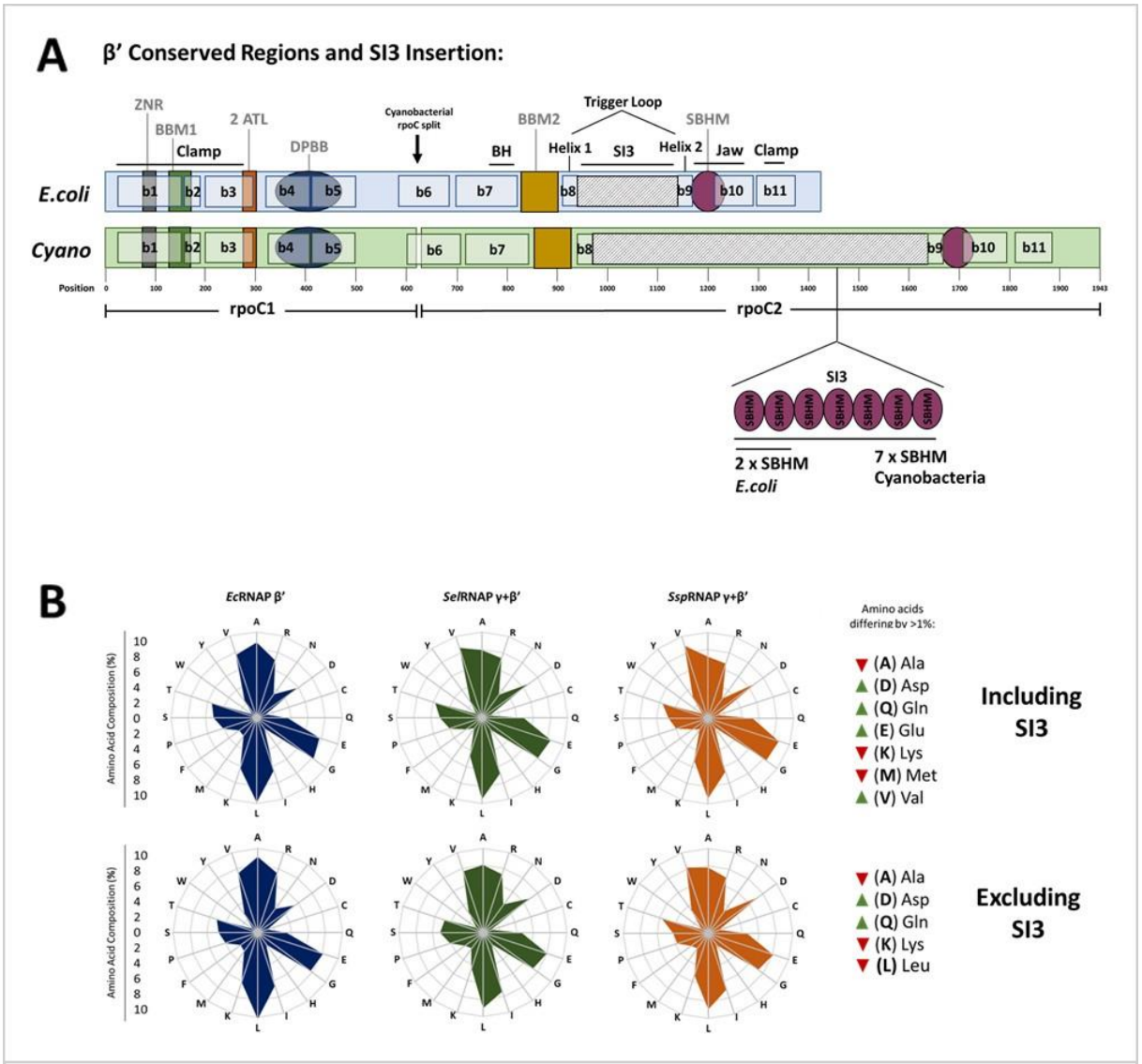


Figure 2. Comparison of *E. coli rpoC* and Cyanobacterial *rpoC1/2*. (A) Linear map of conserved regions (b1-b11) and important elements of the large catalytic β' subunit, or γ+β' in the case of cyanobacteria. Approximate positioning of these regions was obtained using alignment of *E. coli*, *S.el*, and *S.sp* sequences. Adapted from (Lane and Darst, 2010a; Weilbaecher *et al*., 1994; Imashimizu *et al*., 2011). ZNR, zinc ribbon; BBM, β-β' module; ATL, AT-hook-like motif; DPBB, double-Ψ β-barrel; BH, bridge helix; SBHM, sandwich-barrel-hybrid-motif. **(B)** Radar charts showing amino acid composition percentages of the stated *E. coli* and cyanobacterial subunits with and without the SI3 domain. Amino acids differing by >1% between *E. coli* and cyanobacteria are shown to the right. Calculated using ExPASy ProtParam tool.

1.3.3 Elements of the Active Site

The active site of RNAP is positioned in the cleft formed between the β and β' subunits and is buried in the structural center of the enzyme surrounded by a number of entry/exit channels to the polymerase surface. All enzymatic activities catalysed by RNAP occur in the same active site via a two-metal-ion mechanism to direct phosphoryl transfer. This mechanism is supported by two mobile elements of the β' subunit, the trigger loop and the bridge helix, which play essential roles in catalysis and translocation.

1.3.3.1 Metal Chelation Site and Two-Metal-Ion Catalysis

The site of RNAP catalysis is formed not only at the interface of the β and β' subunits but specifically between a six-stranded double-Ψ β-barrel (**DPBB**) motif in each (Iyer *et al*., 2003). These motifs are encoded by the conserved b4 and b5 regions of β', a similar conserved region from β, and involve several nearby residues to perform efficient catalysis.

Depicted in **Figure 3**, the DPBB of the β' subunit provides a universal heptapeptide motif of NADFDGD for metal coordination whereas the diverged DPBB of the β subunit contributes two lysine residues which are involved in substrate binding (Cramer, 2001). In cyanobacteria the "β'-DPBB" is located in the N-terminal region of the rpoC gene split and so is actually encoded by the γ subunit. These two DPBBs are oriented asymmetrically to form the positively-charged catalytic cleft for nucleic acid binding.

The principal activity of RNAP is to synthesise an RNA transcript via nucleotidyl transfer between the α-phosphate of the NTP substrate and the 3'-OH of the growing RNA with release of pyrophosphate as a by-product. This reaction can also be reversed, where the pyrophosphate stimulates degradation of the RNA with release of 3'-terminal NTP, known as pyrophosphorolysis. The RNAP active site can additionally perform further exo- and endo-nuclease reactions important in maintaining fidelity of the transcript. All the reactions are catalysed via a two-metal-ion mechanism (Mg^{2+}) which was first proposed for *E. coli* DNA polymerase by Freemont *et al.* (1988) and then as a universal mechanism for most nucleotidyl-transfer enzymes (Steitz and Steitz, 1993; Steitz, 1998). Sosunov *et al*. (2003, 2005) later expanded the details of this mechanism with regard to the conserved architecture and many activities of the single multi-subunit RNAP active site.

The universally conserved aspartate triad of the NADFDGD motif tightly chelates Mg^{2+}I with a binding affinity of ~100 μM (Sosunov *et al*., 2003). Substitution of any of these aspartates dramatically reduces Mg^{2+} binding and all activities catalysed by RNAP (Zaychikov *et al*., 1996). The second Mg^{2+} ion, Mg^{2+}II, is only weakly bound with a Kd ~10 mM as it contacts

just one aspartate residue of the triad and instead relies on several water-mediated contacts with surrounding residues (Nudler, 2009; Sosunov *et al*., 2003). Mg^{2+}I is permanently sequestered by RNAP and held the junction of the main and secondary channel, whereas Mg^{2+}II is delivered to the active site via contact with three triphosphate oxygen atoms of the incoming NTP and is released upon completion of the reaction.

The general mechanism of two-metal-ion dependent catalysis uses these ions to orientate proper alignment of the reaction substrates in order to form a nucleophile and then stabilise nucleophilic attack. The role of Mg^{2+}I in phosphodiester bond formation is activation of the RNA 3'-OH and co-ordination of the incoming NTP α-phosphate. RNA 3′ O^- attack on this α-phosphate results in the formation of a new phosphodiester bond and release of a protonated pyrophosphate leaving group (Castro *et al*., 2007). Mg^{2+}II stabilises the pentacovalent transition state by co-ordinating the β and γ phosphates of the NTP substrate and may promote product formation by ligating the pyrophosphate leaving group (Yang *et al*., 2006). The roles of each ion are reversed in pyrophosphorolysis as Mg^{2+}II instead activates nucleophilic attack of pyrophosphate on the phosphodiester bond of the 3' terminal NMP, and Mg^{2+}I stabilises the transition state (Sosunov *et al*., 2003). The two-metal-ion mechanism is based on common S_N2-type nucleophilic attack reaction kinetics where bond formation and breakage occur simultaneously (Yee *et al*., 1979). This mechanism also allows both the formation and degradation of phosphodiester bonds with no specific requirement for a general base or acid. However, the existence of a general base for deprotonation of the RNA 3' hydroxyl and general acid for protonation of the pyrophosphate have been suggested for RNAP, in addition to utilisation of water from the surrounding solution, and are proposed to reside in the trigger loop (Castro *et al*., 2007; Svetlov and Nudler, 2013; Yuzenkova *et al*., 2010).

The geometry of the two metal ions has also been proposed to promote catalysis and result in the energetically favourable protonated leaving group even in lieu of a general acid (Yang *et al*., 2006). The position and separation distance between the two metal ions may vary during the reaction according to the changes in the coordination environment. It has been suggested that the 4 Å separation of the two metal ions described by Steitz and Steitz (2003) represents a "resting" state and that shortening of metal–metal distance may be necessary for phosphoryl transfer reactions to occur. Both Nowonty *et al*. (2006) and Yang *et al*. (2006) used other two-metal-ion mechanism dependent enzymes to substitute the Mg ions for other divalent cations to show that while all could bind the active site, catalytic activity required the ability of ions to form closer transition interatomic distances of <3.5 Å adding evidence for this shortening.

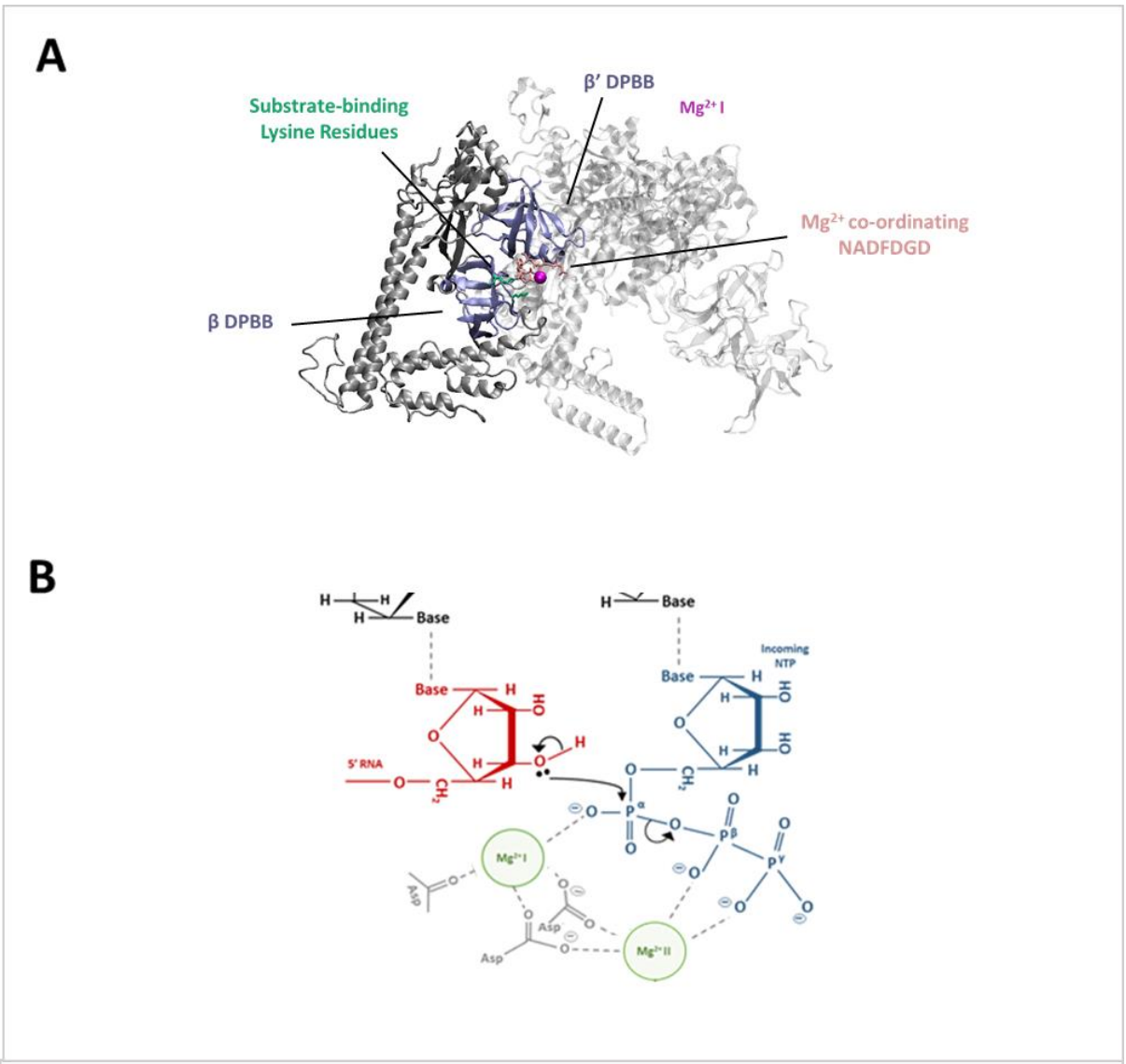


Figure 3. The Metal Co-ordination Site. (A) *E. coli* RNAP PDB 6ALH (Kang *et al*., 2017) showing the DPBB motifs, Mg^{2+} co-ordinating residues of NADFDGD of β', and lysine residues of β. **(B)** Schematic of the phosphodiester bond formation reaction. The two metal ions are chelated by the aspartate residues and also interact with the reacting groups. The RNA 3'-OH attacks the α- phosphate of the NTP substrate. Based on Zhu *et al*. (2009).

1.3.3.2 Trigger Loop

The β' trigger loop (**TL**) is a flexible domain of the active site which assists the two-metal-ion dependent mechanism described previously and is required for efficient catalysis. It lies at the end of the secondary channel, next to the bridge helix, and is encoded by the b8 and b9 conserved regions of β'. While five distinct conformations of this element have been observed in crystal structures of eukaryotic RNAPs, only two functionally important conformations representing the beginning and end points of this spectrum will continue to be discussed. These are known as the 'open/unfolded' and 'closed/folded' TL states (Vassylyev *et al*., 2007; Wang *et al*., 2006). TL folding is one of the major conformational changes of RNAP during transcription and plays important roles in both nucleotide addition and RNA cleavage activities.

The 'folded' conformation is adopted upon binding of the correct substrate NTP and results in an α-helical hairpin-like formation of two α-helices, known as the trigger helices (**TH**), separated by the SI3 insert (Vassylyev *et al*., 2007). The NTP is then enclosed in the active site via a TH and bridge helix three-helical bundle. Two residues, β' H936 and R933 (*E. coli* numbering), of the TL contact the NTP triphosphate to ensure proper positioning of the electrophile and facilitate attack by the RNA 3'-OH leading to determination of the TL as a positional catalyst (Mishanina *et al*., 2017). The orthologous β' H936 residue in *Thermus aquaticus* has also been shown to promote catalysis by acting as a general base (Yuzenkova *et al*., 2010). In this 'folded' conformation, the secondary channel is blocked therefore refusing additional NTP entry and transcript disengagement from the active site. The release of pyrophosphate is thought to destabilise the TH interactions and return the TL to an 'unfolded/inactive' state (Da *et al*., 2012). This is coupled with conformations of the bridge helix which translocates or moves the RNAP along the template DNA to position the next base for synthesis. The 'folded/unfolded' TL states are illustrated in **Figure 4E** using *Thermus thermophilus* RNAP as it does not possess the flexible SI3 insert which usually hinders structural determination in this region.

Substitutions or deletions of the TL significantly affect RNAP catalytic activity, but the presence of the TL is not absolutely required (Temiakov *et al*., 2005; Zhang *et al*., 2010). However, folding of the TL into the TH enhances catalysis rates roughly 10^4-fold (Vassylyev *et al*., 2007). Yuzenkova *et al*. (2010) found that the TL also has roles in transcript fidelity with the 'folded' conformation being sterically hindered by the base of a non-cognate NTP in the active site. The TL is suggested to therefore couple nucleotide recognition to catalysis in

an induced-fit mechanism which helps prevent incorporation of the wrong NTP (discussed in further detail in later chapters).

1.3.3.3 SI3

Sequence Insertion 3 (**SI3**) is the large non-conserved lineage-specific amino acid insertion of the β' subunit which separates the two TH of the TL. SI3 is located on the surface of RNAP and rests next to the entrance of the secondary channel. In *E. coli*, it is a 188 amino acid insert consisting of two sandwich-barrel-hybrid motifs (**SBHM**) motifs whereas in cyanobacteria, it is much larger at ~640 amino acids which encode at least seven SBHMs (Chlenov *et al*., 2005; Iyer *et al*., 2003). The flexible nature of this motif has impeded structural determination in this region, particularly at the N- and C-terminal linker ends of the SI3 which connect to the TH (**Figure 4**). This large cyanobacterial-like SI3 is also present in chloroplasts but completely absent in gram-positive bacteria and the *Deinococcus-Thermus* phylum (Artsimovich *et al*., 2003).

The retention of SIs in related bacteria suggests a functional role such as the ability to adapt RNAP activity to different cellular environments through interaction with extrinsic regulatory factors (Severinov, 2000). The placement of SI3 within the integral TL may also influence catalytic activity and several mutational studies have implicated SI3 involvement in a number of RNAP functions across a variety of species. The SI3 has been proposed to play a role in stabilising interactions with downstream DNA, and with the secondary channel binding factors due to its structural location. Zhang *et al.* (2010) determined SI3 to be essential for the activity of transcription factor GreB, and Furman *et al*. (2013) suggests conformational changes of this flexible insert are responsible for the ability of another factor, DksA, to bind the secondary channel. Artsimovich *et al*. (2003) showed an SI3 deletion in *E. coli* to substantially decrease pause recognition and increase escape from paused RNAP states with Ray-Soni *et al*. (2017) adding that the presence of SI3 promotes termination by increasing pausing at terminator signals. The SI3 has also demonstrated modulation of TL folding into TH thereby affecting rates of nucleotide addition and transcript cleavage by Windgassen *et al*. (2014) who posit that insertions within active site elements like the TL could allow quicker accumulation of advantageous adaptive mutations to new environmental conditions. Deletion mutants of SI3 reduce cell viability and show defects in transcript cleavage and elongation, and partial ΔSI3 RNAP also inhibits correct enzyme assembly in *E. coli* (Zakharova *et al*., 1998).

In cyanobacteria, the large SI3 has been implicated in recognition of specific promoter nucleotides but this is yet to be proven (Imashimizu *et al*., 2003). Chlenov *et al*. (2005) also

note an interesting correlation between presence of the β'SI3 and another SI in the β subunit suggesting that effects of these domains are functionally co-ordinated, however, they also mark cyanobacterial RNAP to be the only exception as they do not encode an equivalent β SI (Iyer *et al*., 2003). The role of the repeated SBHMs within the SI3 is also yet to be determined. These motifs appear to be evolutionary accrued in other lineage-specific structural elements of RNAP (Iyer *et al*., 2004). Only a few have been characterised such as the SBHM of the β domain which assists in σ-factor and promoter binding and may function as the binding site for small RNAP-binding proteins, such as for mycobacterial RbpA whose role is also to stabilise the holoenzyme (Kuznedelov *et al*., 2002; Hu *et al*., 2012).

1.3.3.4 Bridge Helix

The bridge helix (**BH**) is a long α-helix which spans the RNAP active site and separates the main nucleic acid-binding channel from the secondary channel. It is a roughly 35 amino acid mobile element which has a central role in catalysis and the nucleotide addition cycle, encoded by conserved region b7. The N-terminal sequence of b7 forms one wall of the secondary channel and part of the secondary channel rim helices before forming the central BH (Lane and Darst, 2010b).

The BH is proposed to co-ordinate the movements of several domains during the catalytic cycle by transitioning between 'straight' and 'kinked' conformations. Weinzierl (2011) describes two glycine hinges within the BH as functionally important in these conformational changes and in RNAP activity. Amino acid substitutions just before each of these hinges, and at positions interacting with the TL or nucleic acid, can drastically affect rates of nucleotide incorporation (Weinzierl, 2010; Tan *et al*., 2008). The BH works in concert with the folded TH to enclose the substrate NTP in the active site and allow catalysis with the aspartate triad. β′ T790 and β′ G794 interact with the template DNA at the +1 position and with T928 of the TL, whereas β′ R798 interacts with the DNA at +2 (Tan *et al*., 2008). Epshtein *et al*. (2002) correlated BH conformation change with positioning of the RNA 3' terminus suggesting an additional role of the BH in translocation. After NTP incorporation, the RNAP must move or translocate along the template DNA to position the next base for addition. This is currently thought to occur via a Brownian ratcheting mechanism including both the BH and TL and is described further in the next section.

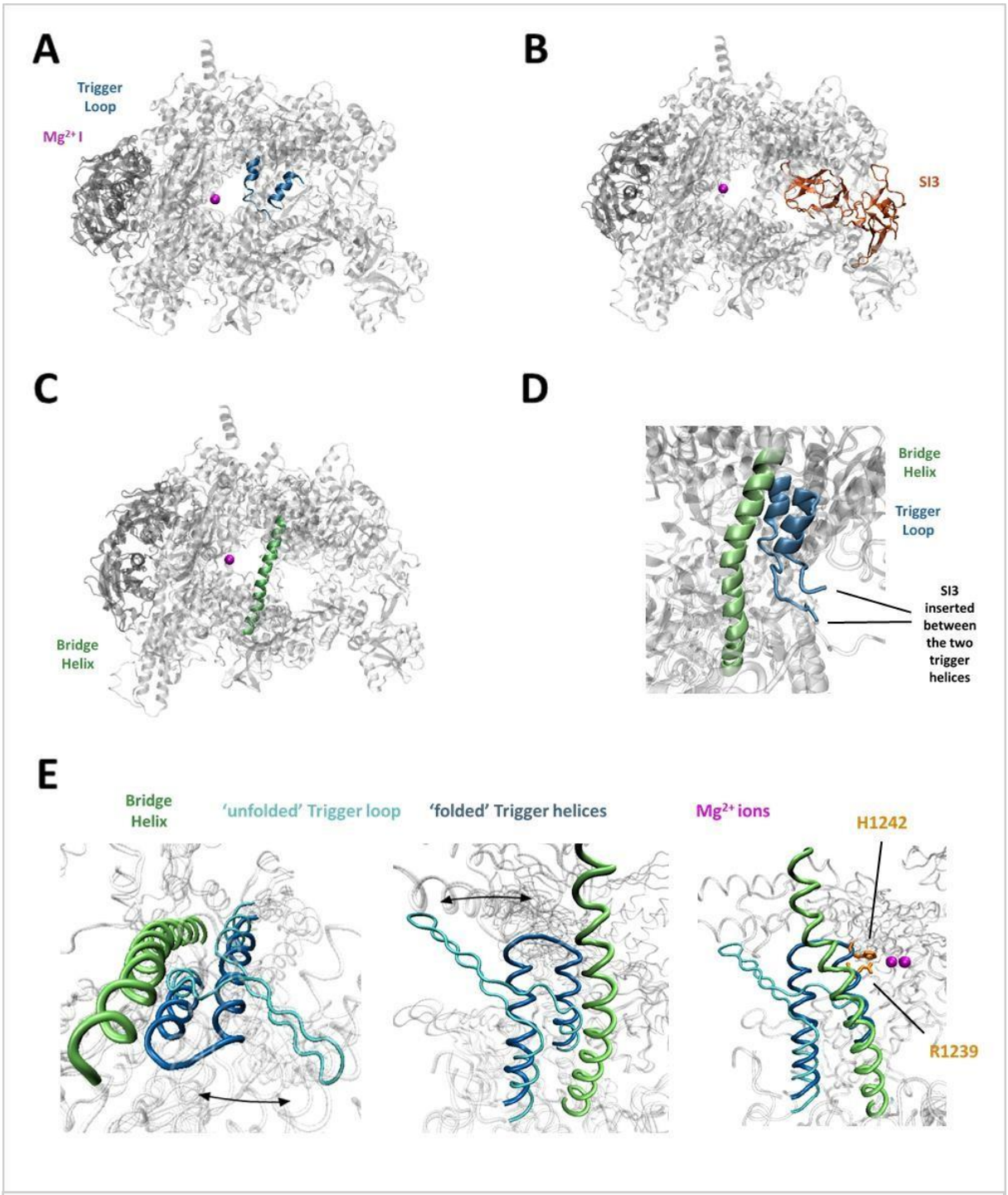


Figure 4. The Active Site Elements of RNAP. (A-D) *E. coli* RNAP PDB 6ALH (Kang *et al*., 2017) showing structural locations of the trigger loop (blue), SI3 (orange), bridge helix (green). The BH/TL is depicted in D, exhibiting the missing linker regions from the TL to SI3 domain. **(E)** *Thermus thermophilus* RNAP structure highlighting the 'folded' and 'unfolded' conformations of the TL/TH and distance of the NTP co-ordinating residues (yellow) to the BH and metal ions of the active site. Folded TL, PDB 205J (Vassylyev *et al*., 2007). Unfolded TL, PDB 4WQS (Murayama *et al*., 2015).

1.4 Regulation and Modulation of Activity

RNAP is subject to a multitude of regulatory mechanisms at all stages of the transcription cycle including elongation. The cycle of RNA synthesis has four steps consisting of nucleotide addition, TL unfolding, pyrophosphate release, and translocation. Each step measurably limits the rate of transcription and provides a point of control (Malinen *et al*., 2012). Regulation also occurs via the template DNA sequence which can encode intrinsic signals for the RNAP to pause or backtrack, allowing time for the binding of an assortment of accessory factors with various modifying activities (Bochkareva *et al*., 2012; Esyunina *et al*., 2016). The ability to control the rate of elongation is an important step of gene expression and can be extremely beneficial for the cell in terms of ensuring transcript fidelity and co-ordination with other cell processes.

1.4.1 Translocation and The Nucleotide Addition Cycle

Synthesis of an RNA transcript requires dynamic conformational changes of the RNAP and involvement of all active site elements described previously in a repeating nucleotide addition cycle (**NAC**). Each round of the cycle, energy derived from triphosphate hydrolysis is used to add one nucleotide to the growing RNA chain and then translocate RNAP along the template DNA to align the next position. There are two important sites in the RNAP active center, the product site (**i site**) where the RNA 3′-OH group is activated for nucleophilic attack on substrate NTP, and the substrate site (**i+1 site**) where the new RNA 3′ terminal end is generated after phosphodiester bond formation (Malinen *et al*., 2012).

Translocation is the one-base-pair (**1 bp**) movement of RNAP along the DNA template. Three main RNAP translocation states affect the alignment of the 3'-OH of RNA with these i sites which are; post-translocated, where the RNA 3' end occupies the i site leaving the i+1 site free; pre-translocated, where the RNA 3' occupies the i+1 site and therefore also blocks the i site; and backtracked, where the RNA 3' end is disengaged from both sites and ejected into the secondary channel (**Figure 5**). Translocation involves conformational changes of the BH and TL, and also requires melting and reformation of 1 bp of upstream and downstream DNA in order to maintain the transcription bubble.

The first step of NAC involves translocation into the post-translocated state in order to position the 3'-OH of the RNA in the i site, leaving the i+1 site vacant and ready to accept an incoming NTP. Binding of the correct substrate NTP in the i+1 site induces a conformational change of the TL into the 'folded' TH. The formation of a new phosphodiester bond is catalysed with pyrophosphate as a by-product. The RNA transcript 3' terminus occupies the i+1 site and the RNAP is in the pre-translocated state. Pyrophosphate release destabilises the

TH interactions and returns the TL to an ‘unfolded/inactive’ state (Da *et al*., 2012). This conformation is coupled with translocation of the RNAP by 1bp to align the i+1 site at the next DNA-template position. This returns RNAP to a post-translocated state where the i+1 site is again free to accept an incoming NTP.

The specific mechanism of translocation is still inadequately understood. Malinen *et al*. (2012) have shown that translocation occurs shortly after or concurrently with pyrophosphate release, is independent of the rate of NTP incorporation, and is co-ordinated with melting of the RNA–DNA hybrid and the downstream DNA. The proposal with the most concurring data is a two-step Brownian ratchet model suggesting RNAP rapidly and reversibly transitions between the post-translocated and pre-translocated states prior to NTP binding (Martinez-Rucobo and Cramer, 2013) Step one involved conformational change of the BH and TL to move the terminal RNA 3’ to the i site, and the second step required relaxation of the BH to allow the next DNA template base to twist into the active site. Recent developments by Kireeva *et al*. (2018) and Nedialkov *et al*. (2018) instead propose a restrained thermal ratchet where slightly slower forward translocation occurs and is poorly reversible. This is potentially supported by the many crystal structures where RNAP is seen predominantly in the post/forward-translocated register (Sekine *et al*., 2015). Nedialkov *et al*. (2018) additionally suggest an alternate mechanism of conformation-generated force-based translocation using molecular simulations. In their model, breakage of downstream i+2 bases cause tightening of the TL against the i+1 site via a TL glycine hinge and compensation by the rest of the TL/BH generate a force resulting in forward translocation of the transcription bubble, with the clamp region and β’ ZNR strongly engaging with the exiting nucleic acids.

1.4.2 Mechanisms to Maintain Transcript Fidelity

In order to ensure correct downstream function of mRNA or ncRNA, RNAPs must transcribe with high accuracy. This is accomplished using two main approaches, NTP substrate selection and transcript proofreading, which both take place in the same RNAP active site.

1.4.2.1 Error Prevention

Strong substrate selection acts as a method of error prevention by reducing the number of erroneous incorporations and therefore the need to perform additional error correction. RNAPs must be able to discriminate rNTPs from dNTPs, and also must ensure selection of the correct rNTP for complementary pairing to the DNA template.

Initial selection is thought to begin with NTP diffusion through the secondary channel to bind an entry, or 'E' site, which overlaps the i+1 site (Westover *et al*., 2004). A cognate NTP would then be able to rotate around Mg^{2+}II to the occupy the i+1 site and form correct pairing with the template DNA and surrounding active site residues. A non-cognate NTP would not form these interactions. The less stable binding and steric clashes with active site elements like the TL would result in transition back to the E site and likely expulsion through the secondary channel (Wang *et al*., 2017).

Mishanina *et al*. (2017) established the TL as a positional catalyst of the active site which possess residues that when in the 'folded' conformation assist co-ordination of the substrate NTP. Yuzenkova *et al*. (2010) has shown that the TL is additionally involved in recognition of the correct NTP and couples this nucleotide recognition to catalysis in an induced-fit mechanism. Binding of the correct incoming NTP in the i+1 site induces folding of the TL into the TH which actively participates in phosphodiester bond formation. Substrate selection is achieved through steric collision of the 'folded' TH conformation with the occupancy of a non-cognate NTP in the i+1 site. This prevention of TL folding delays catalysis and slows progression of the reaction. The delay either allows time for the non-cognate NTP to vacate the active site or results in a slowly performed mis-incorporation event. In this way, the TL acts as a kinetic rather than affinity-based mechanism of maintaining fidelity by providing roughly 3 additional orders of magnitude of NTP discrimination (Yuzenkova *et al*., 2010; Kaplan *et al*., 2008; Wang *et al*., 2015).

The steric hindrance is caused by the altered base pairing geometry to the template of either a non-Watson-Crick pairing for rNTPs, or with an incorrect sugar moiety such as in 2'- and 3'-deoxy NTPs. TL residues β'M932 and Q929 play key roles in discrimination by hydrophobic interactions requiring a precise distance for optimal positioning of the NTPs (Huang *et al*.,

2010). Mutations of these residues or deletion of the TL lead to significantly compromised substrate discrimination (Yuzenkova *et al*., 2010; Zhang *et al*., 2010).

1.4.2.2 Proof-reading and Error Correction

Despite high discrimination at the stage of substrate selection, the mis-incorporation of an incorrect NTP occasionally occurs. This mis-incorporation event results in the upstream movement or 'backtracking' of the RNAP which forces the RNA 3' end to disengage from the catalytic site and become 'frayed'. Mis-incorporated RNAP complexes are thermodynamically more likely to adopt a 1 bp backtracked state (frayed 3' end), however, backtracking can also continue for many bp depending on the surrounding sequence, and in this case the RNA 3' end is projected into the secondary channel (Zenkin *et al*., 2006; Cheung and Cramer, 2011). In the backtracked state, the terminal RNA 3' is away from the catalytic site. The NAC cannot proceed, and the transcript cannot be extended which results in paused RNAP complexes.

These paused complexes provide time for transcription proof-reading mechanisms to resolve the mis-incorporation via endonucleolytic cleavage either with or without the help of an extrinsic factor (Komissarova and Kashlev, 1997). Hydrolysis of the transcript phosphodiester bonds excises the mis-incorporated nucleotides and generates a new RNA 3' end in the catalytic site which can then be extended. Most RNAPs are intrinsically able to perform this hydrolysis using the two-metal-ion mechanism described earlier and folding of the TL (Orlova *et al.*, 1995; Yuzenkova and Zenkin, 2010). In RNA hydrolysis, it is the Mg^{2+}II ion which activates the attacking water molecule in order to cleave the phosphodiester bond. In this instance, the TL β' H936 is thought to play a role in facilitating the backtracked state and in some species may also participate in the reaction as a general base during proton transfer (Mishanina *et al*., 2017; Yuzenkova and Zenkin, 2010). Additionally, Miropolskaya *et al*. (2017) found RNAPs from several species to notably differ in rates of intrinsic RNA cleavage, and determined these differences depended on the TL and disappeared when the TL was deleted, also confirming an essential role in the cleavage reaction. Zenkin *et al.* (2006) describe a supporting intrinsic mechanism of 'transcript-assisted catalysis' where the 3' end of the RNA transcript participates in catalysis by greatly stimulating cleavage of the penultimate phosphodiester bond. In the 1 bp backtracked state, it is this penultimate bond of the transcript which is placed in the catalytic site. It is thought that the erroneous 3' NMP directly interacts with Mg^{2+}II, potentially from the E site, aiding co-ordination and activation of the attacking water molecule and resulting in the release of a 2 bp 3′-RNA fragment (Zenkin *et al.*, 2006; Nielsen and Zenkin, 2013). This factor-independent mechanism

hydrolyses the transcript in a ribozyme-like manner and may represent an ancient mechanism of self-repair used by the RNAP last universal common ancestor before the evolution of diverged cleavage accessory factors (Poole and Logan, 2005; Zenkin *et al.*, 2006). However, rates of intrinsic RNAP hydrolysis tend to be very weak and require stimulation by external factors in order to perform efficient cleavage and error correction (Roghanian *et al*., 2011; Thomas *et al*., 1998). These factors and their mode of action are further explained in later sections. Together these mechanisms act to maintain fidelity of the nascent transcript, ensuring mis-incorporation of an incorrect NTP only occurs approximately once every 10^4–10^5 nucleotides (Imashimizu *et al*., 2013; Blank *et al*., 1986).

1.4.3 Paused and Stalled Polymerases

Transcription in bacteria occurs on a dynamic nucleoid which undergoes dramatic changes in compaction and organisation during growth that largely correlate with the distribution of RNAP and other genomic machineries (Cagliero *et al*., 2014). These machineries must work in unison to carry out their important cellular functions on the same template DNA, thus a high degree of regulation is required.

Transcription is a tightly regulated process where the activity of RNAP can be modulated through conformational changes of internal catalytic domains, intrinsic signals within the template DNA, and the binding of external accessory factors (Landick, 2006; Esyunina *et al*., 2016). Intrinsic signals cause elongation to be a discontinuous process where RNAP dwells at certain sequence positions induced by properties of the DNA sequence itself (Bochkareva *et al*., 2012; Herbert *et al*., 2006). Entry into a paused state is thought to occur with formation of an initial elemental pause where the NAC is inhibited, which can then stabilise and become long-lived pauses through secondary structure hairpin formation or backtracking to disengage the RNA 3’ end (Toulokhonov *et al*., 2007; Weixlbaumer *et al*., 2013). Gabizon *et al*. (2018) recently describe this elemental pause formation as due to sequence-dependent inhibition of forward translocation which provides time for stabilisation to a long-lived pause or stalled complex by slowing transition to the post-translocated state required for NTP addition.

Pausing plays critical roles in co-ordinating co-transcriptional processes such as RNA folding and processing (Wong *et al*., 2007). Additionally, frequently spaced pause signals are thought to help maintain the coupling between the transcription and translation machineries in bacteria by slowing the rate of RNAP (Landick, 2006). Pausing also allows time for the recruitment of specific interactors at key sites such as the association of Rho at Rho-dependent terminators (Ciampi, 2006). However, long-lived pauses can be highly consequential. The stabilisation of a pause into a backtracked state, or a mis-incorporation event leading to a backtracked state,

can result in pauses lasting for several minutes (Shaevitz *et al*., 2003). This can have major consequences in bacteria where DNA replication and transcription use the same template and occur simultaneously as the replication machinery can collide in either a head-on or a co-directional manner with stalled RNAPs. These collisions have been shown to severely impact cells by resulting in DNA double strand breaks, replication fork arrest, and genomic instability (Merrikh *et al*., 2011; Dutta *et al*., 2012; Mirkin and Mirkin, 2007). A further consequence of stalled RNAPs is that a queue or 'traffic-jam' of RNAPs may form behind and block transcription of that particular operon causing changes in gene expression (Yuzenkova *et al*., 2014). As these stalled RNAPs are in the backtracked state, they can be rescued through cleavage of the RNA to produce a new 3'-OH end at the active site as described above. This activity is greatly stimulated through the action of extrinsic transcription factors which bind the secondary channel of RNAP.

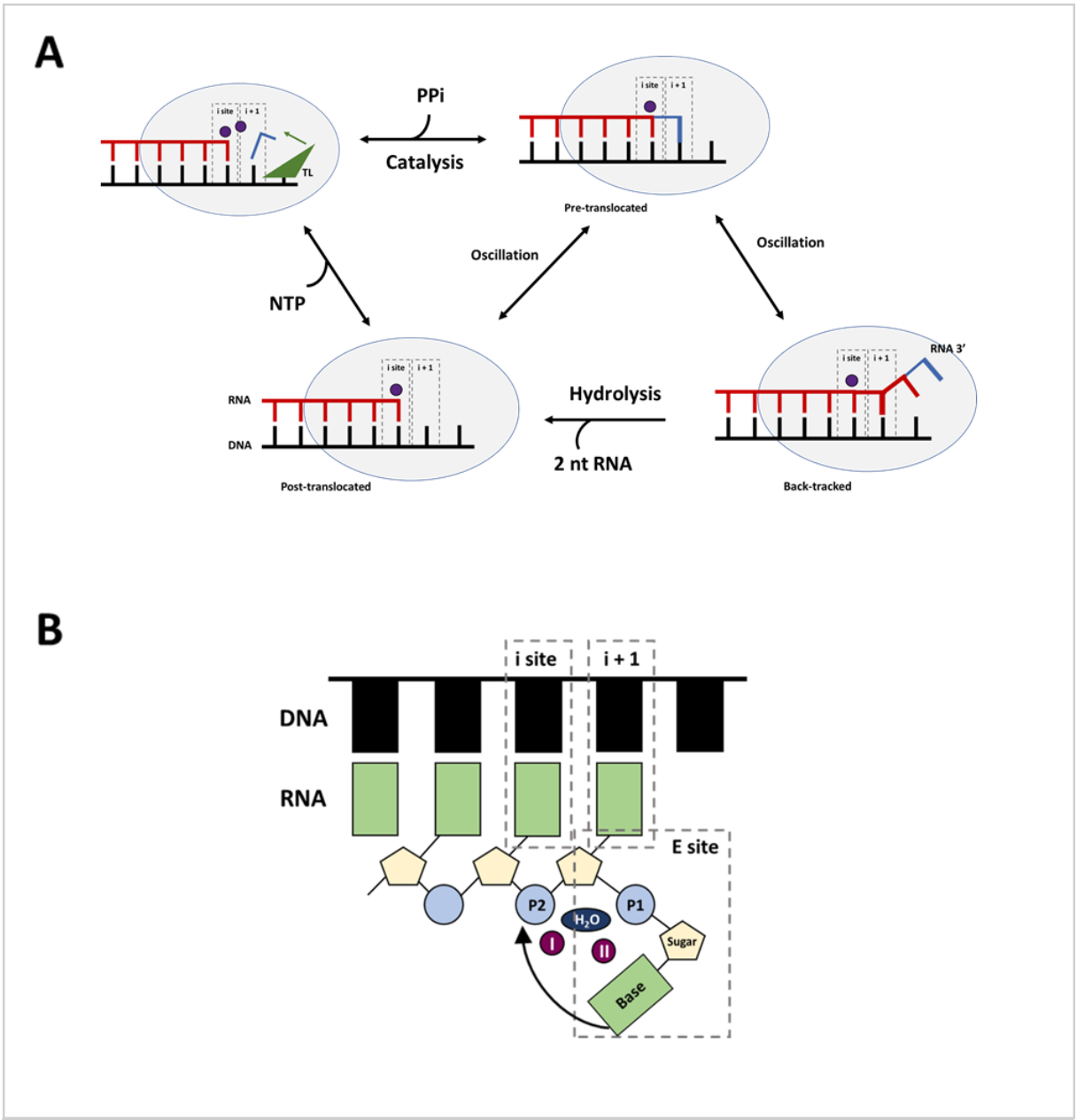


Figure 5. The Nucleotide Addition Cycle and Activities of RNAP. (A) Incoming NTP (blue) binds the i+1 site to begin the NAC. The catalytic Mg^{2+} ions are shown as purple circles. The TL (green) folds upon binding of the NTP in order to catalyse NTP incorporation. The scheme also shows the various translocation states of RNAP during the NAC, and the position of the 3'RNA within the active site for each translocation state. Adapted from Roghanian *et al.* (2011). **(B)** Schematic of the catalytic site of RNAP during transcript-assisted cleavage. The flipped out base is proposed to help co-ordinate the reactant groups to perform cleavage of the penultimate phosphodiester bond.

1.4.4 Secondary Channel Binding Factors

The secondary channel serves as a binding site for several regulatory factors in addition to providing NTP entry to the active site and accommodating the transcript 3' end during backtracking (Esyunina *et al*., 2016). Several small molecules also enter through this channel and affect RNAP activity, such as the alarmone ppGpp (Nickels *et al*., 2004; Perederina *et al*., 2004).

Bacterial secondary channel binding factors (SCBFs) share a similar structural fold of a globular C-terminal RNAP-binding domain and a coiled-coil N-terminal domain encoding important acidic residues at the tip, which is inserted into the secondary channel to contact the active site (Borukhov *et al*., 1993) (**Figure 6**). Five SCBFs have been identified in *E. coli*; GreA, GreB, DksA, RnK, and YacL (Borukhov *et al*., 1993; Nickels *et al*., 2004). These factors are not essential, however there is increased fitness-cost with increased SCBF deletion indicating that some physiological functions may be redundant (Orlova *et al*., 1995). *E. coli* appears to have the greatest number of identified SCBFs, with many species possessing only GreA (Yuzenkova *et al*., 2014; Glass *et al*., 2006).

1.4.4.1 The Gre Factors

RNA hydrolysis is the main mechanism of co-transcriptional proof-reading which corrects transcriptional errors during RNA synthesis (Sosunova *et al*., 2013). Following nucleotide mis-incorporation, the RNAP backtracks to position the internal phosphodiester bonds at the catalytic i+1 site to allow cleavage. The same two-metal-ion mechanism and mobile TL are used to co-ordinate the reactants. Mg^{2+}II activates the attacking water molecule while the TL is thought to assist in positioning the frayed RNA 3' end into the E site where the NMP would provide chemical groups for stabilising the Mg^{2+}II and facilitating catalysis (Zenkin *et al*., 2006; Yuzenkova and Zenkin, 2010). Gre factors are able to displace the TL in the active site and greatly stimulate this reaction by donating catalytic residues which better chelate Mg^{2+}II and co-ordinate more efficient hydrolysis (Sosunova *et al*., 2003; Laptenko *et al*., 2003).

These factors are universally conserved across the domains of life for the function of cleavage stimulation and have evolved independently to encompass the Gre factors in bacteria, and TFS/TFIIS factors in archaea and eukaryotes (Hausner *et al*., 2000; Fish and Kane, 2002). *E. coli* possesses two cleavage factors, GreA and GreB, which act on complexes backtracked by 1 bp and several bp respectively (Borukhov *et al*., 1993). The Gre factors are formed from a long N-terminal coiled-coil domain and a C-terminal globular domain (Stebbins *et al*., 1995). The conserved acidic residues for metal ion co-ordination, D41 and E44, lie at the tip of this coil which is inserted into secondary channel of RNAP to contact the active site (Sosunova *et al*., 2003).

Whereas, the globular domain binds at the entrance to this channel (Laptenko *et al*., 2003). The acidic residues of GreA simultaneously assist with Mg^{2+}II co-ordination, stabilise the attacking water via hydrogen bonding with the water hydrogen atom, and to accelerate water ionisation by lowering the water pK_a by acting as a general base. (Sosunova *et al*., 2013). The SI3 has been shown to be essential for the action of cleavage factor GreB in *E. coli* and flexibility in this domain is required to allow binding of SCBF DksA (Zhang *et al*. 2010; Furman *et al*., 2013). The TL is thought to play an indirect role in Gre-dependent cleavage by assisting to position GreA for catalysis. The TL becomes trapped in an unfolded state by insertion of GreA, which enlarges the secondary channel and induces kinking of the BH (Sekine *et al*., 2015; Miropolskaya *et al*., 2017).

The increased hydrolytic activity provided by these factors increases the efficiency of proof-reading and error correction in the synthesised transcripts. However, James *et al*. (2017) analysed actively transcribing elongation complexes in *E. coli* and found 7% of complexes had undergone a mis-incorporation in cells lacking GreA and GreB whereas wild-type cells with these proof-reading factors had 3% of complexes with mis-incorporations. They suggest that as mis-incorporation induces backtracking of the RNAP, and these pauses can be detrimental to cells, that the major role of cleavage factors may be in resolving stalled RNAPs with increased fidelity of the transcript as a by-product. Yuzenkova *et al*. (2014) has determined this to be the case for *Streptococcus pneumoniae* which only encodes one Gre factor, GreA.

Cyanobacteria are one of the largest bacterial taxons, however no homologue to GreA has been identified. Other small groups in which GreA appears absent are the extremophiles *Aquificaceae* and *Dictyoglomaceae*, and anaerobic gram-negative *Fusobacteriaceae*. It is yet unidentified how these species cope with the absence of cleavage factors considering their important functions regarding the resolution of mis-incorporated and backtracked complexes.

1.4.4.2 The Gfh factors

Gre-factor-homologues are members of the Gre family of SCBFs and are lineage-specific factors of the *Deinococcus/Thermus* extremophile phylum (Hogan *et al*., 2002). Gfh1 of *Thermus aquaticus* (***T.aq***) shows a high degree of sequence similarity with *T.aq* GreA and also shares the N-terminal coiled-coil and C-terminal globular domain structure (**Figure 6**). Lamour *et al*. (2006) solved the crystal structure of *T.aq* Gfh1 to add specific characteristics to the Gfh1 protein. These included a straighter coiled domain with a more flexible tip, and significant structural flexibility with large interdomain rotation indicating the possible use of a conformational switch for activity (Lamour *et al*., 2006).

Laptenko *et al*. (2006) confirmed this switch by showing pH-induced conformational change of Gfh1 into an active state. At a pH above 7, Gfh1 is in an inactive conformation which prevents binding to RNAP whereas at a lower pH, the globular domain of Gfh1 is orientated much more comparably to that of GreA which enables Gfh1 to bind and modulate RNAP activity (Laptenko *et al*., 2006). The activity of Gfh1 is also in contrast to that of GreA. Gfh1 of *Thermus thermophilus* has been shown to inhibit transcript cleavage and NTP addition (Hogan *et al*., 2002; Symersky *et al*., 2006). This difference in activity is thought to be due to four acidic residues on the tip of the Gfh1 coiled-coil, as opposed to the two for GreA. The flexible tip of Gfh1 encodes four aspartate residues in the formation of DDYDD (Symersky *et al*., 2006). Gfh1 is proposed to inhibit RNAP activity through competition with the NTP substrates for co-ordination of Mg^{2+}II which results in mis-alignment of the ion and stabilises the active site of RNAP in a catalytically inactive conformation (Laptenko *et al*., 2006; Symersky *et al*., 2006). However, Esyunina *et al*. (2016) has demonstrated that the Gfh1 of *Deinococcus radiodurans* does not inhibit cleavage but does increase RNAP pausing and termination. They also showed that the TL is involved in Gfh1 binding and that use of Mn^{2+} as the second metal ion significantly increases Gfh1 activity.

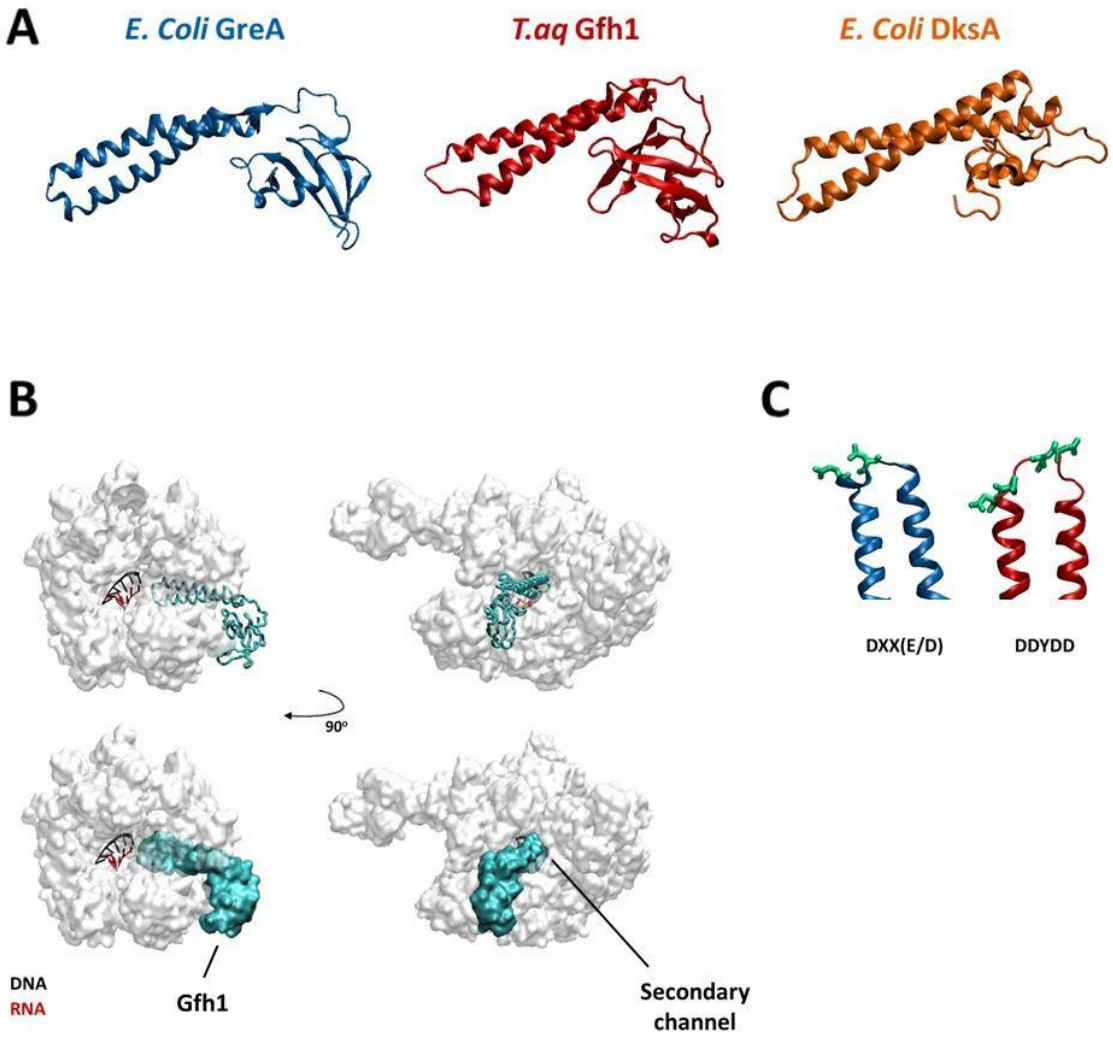


Figure 6. The Secondary Channel Binding Factors. (A) Structure comparison of the main SCBFs showing the N-terminal coiled coil and C-terminal globular domain. *E. coli* GreA PDB 1GRJ (Stebbins *et al*., 1995), *T. aq* Gfh1 PDB 2ETN (Lamour *et al*., 2006), *E. coli* DksA PDB ITJL (Perederina *et al*., 2004). **(B)** Space filled *Thermus thermophilus* RNAP showing binding of a GreA/Gfh1 chimeric protein, cyan, in the secondary channel PDB 4WQT (Murayama *et al*., 2015). **(C)** Close up of the flexible tip regions of the coiled-coil domains of GreA and Gfh1. Acidic metal- chelating residues are highlighted green. There are two for GreA and four on the tip of Gfh1.

1.5 Statement of Aims

Transcription in cyanobacteria involves several interesting features. Cyanobacteria comprise one of the very few groups in which no proof-reading GreA factor has been identified. GreA increases efficiency of RNA cleavage, therefore helping to maintain the fidelity of the RNA transcript and assists in the resolution of stalled RNAPs to prevent genome damage. The vast majority of bacterial species encode at least one of these highly conserved factors and so their absence in cyanobacteria is intriguing. It was thought that as GreA is absent, a compensatory mechanism which carries out the important functions performed by this factor may exist.

The aims of this thesis were to;

1. Investigate this hypothesis, primarily through analysis of *in vitro* transcription activity of *Escherichia coli* RNAP, which is the best characterized RNAP, and the RNAPs of *Synechocystis sp* PCC 6803 and *Synechococcus elongatus* PCC 7942.
2. If such a compensatory mechanism were found, to attempt determination of the details and efficiency of this mechanism as opposed to that of GreA in *E. coli*
3. Characterise the *in vitro* transcription behaviours of cyanobacterial RNAPs and identify any species-specific transcriptional activity.

4. Gfh1 of *Thermus aquaticus* was also investigated to determine the effect of translocation state on the inhibitory activity demonstrated by this factor.

2. Materials and Methods

2.1 Microbiology

2.1.1 Strains

Wild-type (WT) strains of *Synechocystis sp* PCC 6803 (*S.sp*) and *Synechococcus elongatus* PCC 7942 (*S.el*) were kindly donated by Professor Nigel Robinson of Durham University and Professor Conrad Mullineaux of Queen Mary University, respectively. Three *E. coli* strains were generously supplied by Dr. Pamela Gamba of Newcastle University of *E. coli: rpoC*-FLAG, a double detetion strain *E. coli:* Δ*greA*Δ*greB*, and triple deletion strain *E. coli:* Δ*greA*Δ*greB*Δ*dksA*. Commercial strains used in the construction and purification of cloned proteins are identified in the appropriate forthcoming sections. A full list of all strains can be found in **Appendix 5.1**.

2.1.2 Growth Media and Conditions

Both cyanobacterial strains were subject to the same conditions. Cells were grown under a 12-hour light/dark cycle at 30ºC with constant air flow in an Algaetron AG 130-ECO light box from Photon Systems Instruments. Light intensity was set at 100 µE $m^{-2}s^{-1}$ for plates and cultures. Spectrophotometry of cyanobacterial liquid cultures were performed at OD_{730} to account for light absorbance from chlorophyll and pigment proteins (BioChrom Libra S35 spectrophotometer). Cyanobacterial cultures were grown in modified BG-11 media as described by (Stanier *et al*., 1971), or on BG-11 plates (BG-11, 1% agar). *E. coli* cells were grown in in Luria-Bertani (LB) media (10 gL^{-1} Tryptone, 10 gL^{-1} NaCl, 5 gL^{-1} yeast), or on LB plates (LB, 1.2% agar) and incubated at 37ºC unless stated otherwise. Spectrophotometry of *E. coli* liquid cultures were performed at OD_{600}. For long term storage, cells were grown to an OD of 0.5-0.7 and stored at -80°C with 20% glycerol.

2.1.3 Extraction of Total RNA and RNASeq Analysis

To extract total RNA, fresh cultures of *E. coli* and *S.sp* were grown to an OD of 0.5 before being harvested. The protocol of the FastRNA pro Blue Kit was followed initially. 10 ml of culture was frozen in liquid nitrogen and centrifuged at 2800 rpm for 15 min at 4ºC. The supernatant was decanted and 1 ml of RNApro Solution was added and the cells were resuspended. 1 ml of these resuspended cells were transferred to the lysing matrix B powder tube provided with the kit. The tube was then placed in a FastPrep-24 homogeniser (MP biomedicals) on factory setting 3. After homogenisation, the tubes were centrifuged at 12,000g for 5 min at 4ºC and roughly 750 µl of supernatant was transferred to a fresh tube.

This was incubated at room temperature for 5 minutes after which 300 µl of chloroform were added (without isoamyl alcohol) and the tube was vortexed for 10 seconds, followed by another 5 minute room temperature incubation. After centrifugation at 12,000g 4ºC for 5 min, the upper aqueous phase was transferred to a new tube. The Norgen Biotek Corp Total RNA purification plus kit was then followed, starting in section 2. This protocol was followed to completion, resulting in eluted RNA. After verifying the purity of the eluted RNA using the Bioanalyser from Agilent, the purified RNA samples were kept at -80 ºC and aliquots were sent to vertis Biotechnologie AG for sequencing. Analysis of the RNAseq data was performed by Dr Katherine James of Newcastle University. In effect, total error rates were calculated as the percentage of total reads with a mismatched base at each read position in the alignment, and specific error rates were calculated as the percentage of total reads with a specific mismatch, for example an A incorporated instead of a G (G > A misincorporation), at each read position. In this work we present the calculated specific error rates.

2.1.4 Transformation of Cyanobacterial Strains

The dark transformation protocol, as described by Golden *et al*. (2004) was used to transform *S.el* to encode a C-terminal FLAG-tag on the *rpoC2* gene via homologous recombination. Cells were grown to an OD_{730} of 0.6-0.7 and harvested by centrifugation for 10 minutes at 6000 g. The cell pellet was resuspended in 10 ml of 10 mM NaCl and spun again for 10 minutes at 6000 g. This pellet was resuspended in 0.3 ml of BG-11 medium. 50 ng-5 µg of plasmid DNA which encoded the C-terminal end of *rpoC2* with an attached his-tag and homologous regions at either end, was added to the resuspension. To shield the cells from light, they were wrapped in aluminium foil and incubated overnight at 30ºC with gentle agitation. The entire 0.3 ml suspension was then plated on BG-11 containing the appropriate antibiotic (10 $\mu g ml^{-1}$ Spectinomycin and 5 $\mu g ml^{-1}$ Streptomycin) and incubated under the standard 12-hour light/dark cycle at 30ºC. Colonies needed to be picked and re-streaked several times for full segregation of the mutant. Fully segregated *S.el rpoC2*-FLAG was verified by sequencing (GATC Biotech).

2.1.5 Molecular Cloning of Gfh1 and Gfh1-D4A

Gfh1 was amplified from a previous lab construct in order to transfer it to pET28 to include a C-terminal his-tag. Plasmid and insert DNA were restricted using the appropriate restriction enzymes and purified using the protocol provided with the Qiagen Gel Extraction kit. 3 times

molar excess of insert DNA to plasmid DNA was mixed and ligated overnight at 4ºC with T4 DNA ligase (NEB). 5 µl of ligation mix were then transformed into 50 µl of DH5α competent *E. coli* cells (NEB) using the heat shock protocol provided. Cells were plated on LB with the 30 $\mu g ml^{-1}$ kanamycin selective antibiotic. Positive clones were first identified through colony PCR. Plasmid DNA of potential clones was isolated using the Qiagen Mini Prep Kit and checked by subsequent restriction digest. Positive plasmids were then sent for sequencing (GATC Biotech). Once a verified plasmid of pET28 Gfh1 had been constructed, site-directed mutagenesis was used (as described for the *Ec*RNAP mutants) to construct the inactive Gfh1-D4A. All four aspartates of the flexible tip were substituted with alanines (Laptenko *et al*., 2006).

2.1.6 Mutagenesis

Site-directed mutagenesis was performed on his-tagged *Ec*RNAP *rpoC* (β') encoding pRL663 plasmid to construct the mutant *Ec*RNAPs with residue substitutions to match the corresponding residue position of *Ssp*RNAP (pRL663 was donated by Professor Irina Artsimovitch). The QuikChange II Site-Directed Mutagenesis kit from Agilent was used to construct these mutants with no deviation from the protocol provided with the kit. Mutated plasmids were transformed into Agilent Technologies XL10-Gold Ultracompetent cells following the provided protocol. Colonies were picked and grown overnight in LB supplemented with 80 $\mu g ml^{-1}$ ampicillin and plasmid DNA was extracted using the Qiagen Mini Prep Kit. Positive clones were identified through sequencing, using primers of the *E. coli rpoC* gene (GATC Biotech).

2.1.7 Production of Competent Cells

Plasmids encoding the correct substitution from the mutagenesis above were then transformed into the *E. coli* Δ*greA*Δ*greB* strain using the Inoue method of competent cell production (Sambrook and Russell, 2006). A loop of cells were taken from a freshly streaked overnight plate of *E. coli* Δ*greA*Δ*greB* and resuspended in 500 µl LB and incubated at 37ºC for 1 hour. This liquid culture was then spun at 4ºC 3000 g for 2 minutes. The pellet was resuspended in Inoue salts buffer (55 mM $MnCl_2$, 15 mM $CaCl_2$, 250 mM KCl, 10 mM PIPES, H_2O to 1 L). 50 µl of this mix was used per transformation. Roughly 50 µg plasmid were added and left to incubate on ice for 30 minutes. This mix was then heat shocked at 42ºC for 20 seconds and again incubated on ice for 2 minutes. 800 µl of LB were added and the tubes were placed at 37ºC for 1 hour with shaking. The mix was then plated on LB with 80 $\mu g ml^{-1}$ ampicillin. Positive clones were verified by sequencing (GATC Biotech).

2.2 Protein Purification and Analysis

2.2.1 Purification of Native RNAP

2.2.1.1 EcRNAP

Native purification of RNAP was performed using two-step column chromatography. Native core and holoenzyme RNAP were purified from pelleted exponential cultures resuspended in lysis buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, protease inhibitor cocktail tablet) and subjected to 6 minutes of 2 second on/off sonication at 40 % amplitude. The lysates underwent two centrifugation steps of 15 minutes at 15K rpm and 20 minutes at 18K rpm until a clear supernatant was obtained. This supernatant was passed through a 0.45 µM PVDF filter (Merck) before being put through a 5 ml HiTrap heparin column (GE Healthcare). The column was pre-equilibrated in buffer A (20 mM Tris pH 7.9, 5% glycerol) and protein was eluted with an increasing gradient of buffer B (20 mM Tris pH 7.9, 5% glycerol, 1 M NaCl) in fractions of 0, 100, 300, 600, and 1000 mM salt. The eluate determined to contain RNAP was then substantially diluted in buffer MA (10 mM Tris pH 7.9, 5% glycerol, 0.1 mM EDTA, 0.1 mM DTT) for further purification via gradient elution on an ion-exchange MonoQ 5/50 GL using an increasing percentage of buffer MB (10 mM Tris pH 7.9, 5% glycerol, 0.1 mM EDTA, 0.1 mM DTT, 1 M NaCl). SDS-PAGE was used to determine which fractions contained RNAP. These fractions were then concentrated using 100 kDa amicon centrifuge filters (Merck) and dialysed overnight in storage buffer (50% glycerol, 200 mM KCl, 40 mM Tris pH 7.9, 1 mM EDTA, 1 mM DTT). The samples were then continually stored at -20^{o}C. Concentration of the purified protein was determined by Bradford assay, with each purification producing 1 mg of protein on average for *Ec*RNAP. *Ec*RNAP subunits should present around 155 kDa for β', 150 kDa for β, 36 kDa for α, and 10 kDa for ω.

2.2.1.2 Cyanobacterial RNAP

Purification of both cyanobacterial RNAPs was carried out in a similar manner to the above but included a 4^{o}C overnight 0.35 mgml^{-1} ammonium sulphate precipitation step. This precipitation was performed on the eluted RNAP fraction after heparin chromatography. After overnight incubation, the precipitate was spun down at 18K rpm for 20 minutes before resuspension in buffer MA and continued purification via MonoQ chromatography. The average amount of cyanobacterial RNAP obtained per purification was roughly 80 µg. Cyanobacterial RNAP subunits should present around 144 kDa for β', 123 kDa for β, 71 kDa for γ, 34 kDa for α, and 8 kDa for ω.

2.2.2 Purification of his-tagged RNAP

This included WT *Ec*RNAP with a his-tag and all mutant *Ec*RNAPs. IPTG-induced T7 express LB cultures were subjected to the same protocol as above up to the end of the heparin chromatography. The eluted RNAP fraction from the heparin column was passed through a 5 ml nickel chromatography column (GE Healthcare) which had been pre-equilibrated with buffer NA (20 mM Tris-HCl pH 7.9, 600 mM NaCl, 5% glycerol). Protein was eluted with an increasing gradient of buffer NB (20 mM Tris-HCl pH 7.9, 600 mM NaCl, 5% glycerol, 200 mM Imidazole). His-tagged RNAP yield and purity were then assessed via SDS-PAGE. Fractions of WT his-tagged *Ec*RNAP were then subsequently subjected to further MonoQ purification, as described above. Clean fractions of mutant his-tagged *Ec*RNAP were concentrated using 100 kDa amicon centrifuge filters (Merck) and dialysed overnight in storage buffer.

2.2.2.1 TaqRNAP

His-tagged *Taq*RNAP was purified from plasmid pTaqABC, which had been constructed previously by Mohammad Roghanian. Purification from this plasmid followed the protocols described in previous studies (Kuznedelov et al., 2003; Yuzenkova and Zenkin, 2010). pTaqABC was transformed into T7 express cells (NEB) using the manufacturer's protocol and plated on LB with 30 $\mu g ml^{-1}$ Kanamycin. 2 L LB cultures of these cells were grown overnight at 37^{o}C + 30 $\mu g ml^{-1}$ Kanamycin but did not require IPTG induction. These cultures were pelleted and resuspended in T-Lysis buffer (50 mM Tris-HCl pH7.9, 5% glycerol, 300 mM NaCl, protease inhibitor cocktail tablet). Cells underwent sonication at 60% amplitude for 10 minutes before centrifugation at 15k rpm for 30 minutes. The supernatant was then incubated at 70^{o}C for 1 hour before another centrifugation step of 15k rpm for 30 minutes. This supernatant was passed through a 0.45 µM PVDF filter (Merck) before undergoing Nickel chromatography as described above. Fractions determined to contain *Taq*RNAP were then further purified via MonoQ ion-exchange chromatography as described previously. Clean fractions of his-tagged *Taq*RNAP were concentrated using 100 kDa amicon centrifuge filters (Merck) and dialysed overnight in storage buffer. Purified fractions shown in **Figure 7**.

2.2.3 Purification of FLAG-tagged RNAP

A *S.el rpoC2*-FLAG was constructed during this project as described above. 500ml *S.el*-FLAG cultures were grown under a 12 hour Light/Dark cycle 100 µE $m^{-2}s^{-1}$ and 30ºC until an OD of ~0.5 was reached. These cultures were spun down at different phases of the circadian cycle which correspond to differences in gene expression activity (Woelfle *et al*., 2006). These were relative dawn, dusk and midnight which are D11, L11, and D5 respectively. D/L represent dark/light and the number is the number of hours that have passed out of the 12-hour cycle. WT *S.el* was also subjected to growth under these conditions and was pelleted at the dusk phase as a control. Similar to the above, the pellets were sonicated for 6 minutes at 40% amplitude for 2 seconds on/off periods in standard lysis buffer. The lysates then underwent a 30 minute 37ºC incubation with 20 µl micrococcal nuclease (Promega) and 5mM Ca^{2+} before an initial centrifugation of 15 minutes at 15K rpm. The supernatant from this was then ultracentrifugated for 1 hour at 100K g and the supernatant was collected and concentrated to 5ml using a 100K amicon centrifugal filter unit (Merck). 1ml of this concentrated ultra-supernatant was incubated at 4ºC for 20 hours in the presence of 20 µl FLAG resin beads (Sigma). After incubation, the beads were washed twice with (50 mM Tris-HCl pH 7.4, 150 mM NaCl) and RNAP was eluted by competition with 2 $mgml^{-1}$ 3X FLAG peptide.

2.2.4 Purification of E. coli GreA

pET21-GreA with a C-terminal his-tag was provided by Dr. Flint Stevenson-Jones. Purification of this protein followed the method described by Koulich *et al*. (1997). T7 express cells transformed with pET21*Ec*GreA were grown at 37˚C to an OD_{600} of 0.4 and then induced for 4 hrs using 0.4 mM IPTG. The cells were then spun down at 5000 rpm for 7 minutes at 4˚C and resuspended in 40 ml of denaturing buffer (7 M Guanidine-HCl, 40mM Tris-HCl pH 7.5, 0.8 M NaCl, 1 mM EDTA, 1 mM DTT). The cells were then centrifuged at 25K rpm for 20 minutes and the supernatant was added to Ni-NTA beads (GE Healthcare). The supernatant was incubated with the beads for 20 minutes at room temperature before being washed three times with 10 ml buffer A and then five times with a refolding buffer (40 mM Tris-HCl pH 7.5, 0.8 M NaCl, 1 mM EDTA, 1 mM DTT). GreA could then be eluted with three rounds of 2 ml washes with elution buffer (40 mM Tris-HCl pH 7.5, 0.8 M NaCl, 0.6 M imidazole, 1 mM EDTA, 1 mM DTT). Presence of GreA was determined by SDS-PAGE and pure fractions were dialysed overnight in high salt GreA storage buffer (40 mM Tris-HCl pH 7.5, 0.8 M NaCl, 1 mM EDTA, 1 mM DTT, 50% Glycerol). Purified fractions shown in **Figure 7**.

2.2.5 Purification of T. aq Gfh1 and Gfh1-DA4

pET28 plasmids encoding Gfh1 or Gfh1-DA4 were transformed into T7 express cells (NEB) using the manufacturer's protocol. 2 L LB cultures of these cells were grown at 37ºC until an OD_{600} of 0.4, at which point protein expression was induced with addition of 1 mM IPTG, and the cultures were moved to 30 ºC to incubate overnight. Cells were pelleted and resuspended in lysis buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, protease inhibitor cocktail tablet) and subjected to 6 minutes of 2 second on/off sonication at 40 % amplitude. The lysates were then centrifuged at 18k rpm for 20 minutes. The supernatant was passed through a 5 ml nickel chromatography column (GE Healthcare) which had been pre-equilibrated with buffer NA (20 mM Tris-HCl pH 7.9, 600 mM NaCl, 5% glycerol). Protein was eluted with an increasing gradient of buffer NB (20 mM Tris-HCl pH 7.9, 600 mM NaCl, 5% glycerol, 200 mM Imidazole). Presence of the correct protein was determined by SDS-PAGE and pure fractions were concentrated by amicon centrifugal filters and dialysed overnight in storage buffer (50% glycerol, 200 mM KCl, 40 mM Tris pH 7.9, 1 mM EDTA, 1 mM DTT). Purified fractions shown in **Figure 7**.

2.2.6 SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate proteins according to their molecular weight and determine purity of purified proteins. Proteins were suspended in Laemmli buffer (Stock of 2x, supplied by Sigma) and run on 4-20% Expedeon pre-cast run blue SDS gels. Gels were subsequently stained overnight with the coomassie blue-based Instant Blue stain (Expedeon).

2.2.7 Silver Staining

As described by Shevchenko *et al*. (1996). After having been resolved via SDS-PAGE, protein gels were incubated in fixation solution (ethanol 40%, acetic acid 10%, water 50%) for 30 minutes. This was followed with a 30 minute incubation in sensitising solution (75 ml ethanol, 10 ml sodium thiosulphate (5% W/V), 17g sodium acetate, water to 250 ml) and then three 5 minute wash steps with water. The gel was then incubated for 20 minutes in silver solution (25 ml silver nitrate solution (2.5% W/V), formaldehyde (37% W/V), water to 250 ml) and a following single wash step. Developing solution (6.25 g sodium carbonate, formaldehyde (37% W/V), water to 250 ml) was used to visualise the protein to the required standard and then stopped with incubation in stopping buffer (3.65 g EDTA-Na_2H_2O, water to

250 ml). Silver staining was used to visualise the FLAG-tagged *Sel*RNAP purifications due to low yield.

2.2.8 Mass Spectrometry

Purified native RNAP samples of *Ec*RNAP, *Sel*RNAP, and *Ssp*RNAP were sent for mass spectrometry. Additionally sent were FLAG-tagged *Sel*RNAP samples which had been purified from different phases of the circadian cycle corresponding to differences in gene expression activity (Woelfle *et al*., 2006). This mass spectrometry analysis was carried out by Achim Treumann of Newcastle University.

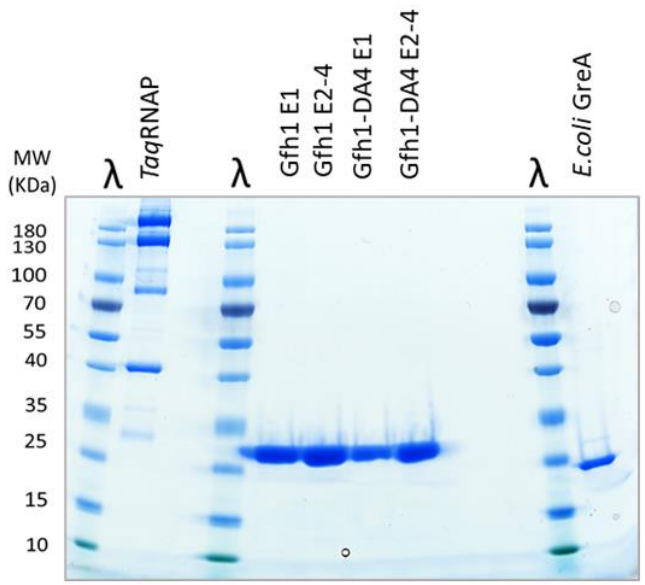


Figure 7. Purified fractions of Additional proteins. Protein gel of purified *Taq*RNAP, *Taq*Gfh1, *Taq*Gfh1-DA4, and *Ec*GreA fractions used in *in vitro* transcription assays.

2.3 *In vitro* Transcription and Analysis

2.3.1 5' RNA [γ-^{32}P]-ATP Labelling

To 5' radiolabel RNA oligonucleotides, a final concentration of 10 pmμl^{-1} RNA was incubated with 5 μl [γ-^{32}P]-ATP (Hartmann Analytic) and 1.5 μl T4 Polynucleotide kinase (PNK) (Thermofisher scientific) in reaction buffer A at 37^{o}C for 45 minutes. The radiolabelled RNA was then cleaned by gel filtration though BioRad Micro Bio-Spin^{TM} 6 chromatography columns equilibrated in milli-Q H_2O.

2.3.2 Artificially Assembled Elongation Complexes

Promoter-independent elongation complexes (**EC**) were assembled as described (Sidorenkov *et al*., 1998; Yuzenkova *et al*., 2010). In a final reaction volume of 10 μl; biotinylated template DNA (10 pmμl^{-1}) and RNA oligonucleotide (20 pmμl^{-1}) were first annealed at 45^{o}C for 10 minutes in Transcription Buffer (20 mM Tris pH 7.9, 40 mM NaCl, 0.1 mM EDTA) and then cooled at 4^{o}C for a minimum of 20 minutes to form the DNA:RNA hybrid. The polymerase was then added at roughly 2 pmol per reaction and the mix incubated at room temperature for 10 minutes. Addition of excess non-template DNA (100 pmμl^{-1}) and incubation at 37^{o}C (40^{o}C for *Taq*RNAP) completed formation of the EC. These complexes were immobilised on GE Healthcare Streptavidin Sepharose^{TM} high performance beads pre-equilibrated in Transcription buffer. If not using 5'-radiolabelled RNA, then an [α-^{32}P]-NTP (Hartmann Analytic) could be incorporated to the RNA 3'-end at this point by incubation with the NTP at 37^{o}C for 1-2 minutes in the presence of 10 mM Mg^{2+}. Additional NTPs were also added if formation of a particular EC further along the template DNA was required. Excess reagents were then removed by washing the complexes twice in transcription buffer containing high salt (1 M NaCl) and then twice in standard transcription buffer. All *Taq*RNAP assays were performed using a transcription buffer of pH 6.8. Transcription assays could then be performed with addition of the necessary NTP, factor, and metal ion as stated below and in the results. Assays were stopped after the time periods indicated in the figures through addition of Stop Buffer (8 M Urea, 20 mM EDTA, 100 μgml^{-1} Heparin, 1 x TBE buffer, diluted in Formamide) and loaded onto denaturing PAGE. All DNA and RNA oligonucleotides were purchased from IDT. A full list of the scaffolds generated by these oligonucleotides and used in this work can be found in **Appendix 5.3.**

2.3.2.1 RNA Hydrolysis

Experimental setup for all transcript cleavage assays. ECs were assembled on the appropriate scaffolds – EC(GA), MEC(A), MEC(U), MEC(modified base) - as described above. All RNA was 5'-radiolabelled. Cyanobacterial RNAP and *Ec*RNAP cleavage was performed at 30ºC, *Taq*RNAP cleavage was performed at 40ºC unless stated otherwise. Reactions were initiated upon the addition of either 10 mM $MgCl_2$ or 10 mM $MnCl_2$ (A range of $MgCl_2$ concentrations were used for determination of K_M for Mg^{2+}) and proceeded for the indicated time periods before addition of stop buffer. To determine cleavage over a pH range, the reactions were performed using a transcription buffer of the required pH. Cleavage activity was tested every 0.5 pH change from pH 5.5 to 11.0. The buffering agents used were: MES (Sigma) for pH 5.5-6.5; Hepes (Sigma) for pH 7-7.5; Tris (Sigma) for pH 7.9-9.0; Ches (Sigma) for pH 9.5; and Caps (Sigma) for pH 10.0-11.0. All buffers were set to the appropriate pH at 30ºC.

2.3.2.2 Transcript Elongation

Experimental setup for all elongation assays on template 1 and template 2. ECs were assembled on the appropriate scaffolds – template 1 or template 2 – as described above. All RNA was 3'-radiolabelled. Cyanobacterial RNAP and *Ec*RNAP elongation was performed at 30ºC, *Taq*RNAP elongation was performed at 40ºC. If required, the ECs were incubated with the stated amounts of additional factor (e.g. Gfh1) for 5 minutes before the assays were initiated. All elongation was carried out with addition of 10 µM NTPs and either 10 mM $MgCl_2$ or 10 mM $MnCl_2$ and proceeded for the indicated time periods before addition of stop buffer.

2.3.2.3 Incorporation of a Single Nucleotide

Experimental setup for all single nucleotide incorporation assays. ECs were assembled on the appropriate scaffolds as described above. All RNA was 3'-radiolabelled. Cyanobacterial RNAP, *Ec*RNAP, and *Taq*RNAP cognate NTP incorporation was performed at 20ºC. Cyanobacterial RNAP and *Ec*RNAP non-cognate NTP incorporation was performed at 30ºC. Titration assays for *Taq*RNAP were performed at 20ºC when titrating NTP and 40ºC when titrating Gfh1. Cognate NTP was added at a concentration of 1 µM and non-cognate at 1 mM. Titrated NTP concentrations ranged from 1, 5, 10, 20, or 100 µM and was always performed in the presence of 10 µM Gfh1. Titrated Gfh1 concentrations ranged from 10, 3, 1, 0.3, 0.1, and 0.03 µM in the presence of 0.5 µM NTP. If required, the ECs were incubated with the stated amounts of additional factor (e.g. Gfh1) for 5 minutes before the assays were initiated. Reactions were initiated upon the addition of NTP with either 10 mM $MgCl_2$ or 10 mM $MnCl_2$ and proceeded for the indicated time periods before addition of stop buffer.

2.3.2.4 Pyrophosphorolysis

Experimental setup for all pyrophosphorolysis assays. ECs were assembled on the appropriate scaffolds – template 1 EC(14/15/16) - as described above. All RNA was 5'-radiolabelled. Cyanobacterial RNAP and *Ec*RNAP pyrophosphorolysis was performed at 20ºC. Reactions were initiated upon the addition of 250 µM pyrophosphate (PPi) and 10 mM $MgCl_2$ and proceeded for the indicated time periods before addition of stop buffer.

2.3.3 Transcript Elongation from the T7A1 Promoter

Experimental setup for all pIA349 elongation assays. Promoter-dependent ECs were assembled from a biotinylated DNA fragment encoding the strong bacteriophage T7A1 promoter. This linear double stranded DNA template was prepared by amplification from pIA349 (plasmid donated by Professor Irina Artsimovitch) (Artsimovich and Landick, 2002). Equal concentrations of DNA and purified RNAP (+ σ factor) were incubated together at 37ºC for 10 minutes at pH 6.8 to form the elongation complex. A 'hot mix' was then added to allow labelled elongation to position EC(G37) by withholding UTP (5 mM $MgCl_2$, 150 µM CAUC, 5 µM ATP, 5 µM CTP, p^{32}-GTP, 1 µM cold GTP) and incubated for 3 minutes at 37ºC. These ECs were captured on GE Healthcare Streptavidin Sepharose™ high performance beads pre-equilibrated in Transcription buffer and washed twice in transcription buffer containing medium salt (200 mM NaCl) and then twice in standard transcription buffer. Reactions were initiated upon the addition of 'start mix' (10 mM $MgCl_2$, 10 mM CTP, 10 mM

ATP, 10 mM UTP, 1 mM GTP) to progress the EC past G37 and proceeded for the indicated time periods before addition of stop buffer. All elongation assays using the pIA349 template were performed at 30°C.

2.3.4 PAGE

All transcription assays were resolved using polyacrylamide gel electrophoresis (PAGE). All assays except pIA349 elongation were resolved on 23% 8 M urea denaturing PAGE. pIA349 elongation was resolved on 10% 8 M urea denaturing PAGE. The radioactive gels were revealed by phosphorimaging using the Typhoon imaging system of GE Healthcare.

2.3.5 Statistical Analyses

Subsequent quantification and analysis of these gels were carried out using ImageQuant software from GE Healthcare, SigmaPlot version 13.0, and visualised graphically using Microsoft Excel 2016. Kinetic data were fit to a single exponent equation using non-linear regression. K_M data for Mg^{2+} affinity and for NTP titration assays were fit to the Michaelis-Menten equation. Gfh1 titration data were first fit to the Michaelis-Menten equation and then fit using exponential decay non-linear regression analysis. Calculation of activation energy required data to be fit to the Arrhenius equation:

$$k = Ae^{-\frac{E_a}{RT}}$$

The Arrhenius plot is created by plotting the logarithm of the rate constant, k, versus the inverse temperature, 1/T (in Kelvin). R is the universal gas constant equal to 8.314 $Jmol^{-K}$. Activation energy, or E_a, can be calculated from the negative slope of this plot using $y = -E_a/R$

3. Results and Discussions

3.1 *Cyanobacterial* RNAP Exhibits More Efficient Intrinsic Transcript Cleavage than *Ec*RNAP Due to Potential Decreased Flexibility of the Active Site Elements

3.1.1 Introduction and Aims

The SCBF Gre, and therefore Gre-mediated transcript cleavage, appears to be absent in the Cyanobacterial lineage. It was hypothesised that an alternate mechanism may exist to carry out the important function performed by Gre. Investigation into this alternate mechanism was initiated by first purifying cyanobacterial RNAP and then testing activity of the polymerase using *in vitro* transcription assays. Two cyanobacterial strains and their purified RNAPs are used throughout this thesis – *Synechococcus elongatus* 7942 (***S.el*/*Sel*RNAP**) and *Synechocystis sp.* 6803 (***S.sp*/*Ssp*RNAP**). These assays allow the artificial assembly of a catalytically active elongation complex (**EC**) from which the rates of each reaction catalysed by RNAP can then be measured (Sidorenkov *et al*., 1998). The rates obtained for each cyanobacterial RNAP were then compared against those for the Gre-encoding model organism *E. coli* RNAP (***Ec*RNAP**) to help highlight any major differences in activity.

In this section, the aim was to compare each purified RNAP in relation to the function of Gre. The proficiency of each RNAP as an endonuclease was assessed by measuring rates of transcript cleavage in complementary and -1bp backtracked assembled complexes. The results of which then were investigated further using purified mutant *Ec*RNAPs which carried specific mutations of the active site elements corresponding to the respective position in *Ssp*RNAP.

3.1.2 Successful Purification of Cyanobacterial RNAP

Native RNAP was purified from both cyanobacterial strains by subjecting cell lysates first to heparin affinity chromatography and then MonoQ ion-exchange chromatography, combining published methods used for the purification of other species of RNAP (Hager *et al*., 1990; Svetlov and Artsimovitch, 2015). The purification of cyanobacterial RNAP included an additional step of overnight 0.35 $mgml^{-1}$ ammonium sulphate precipitation of the intermediate RNAP fraction after the heparin chromatography. Tagged RNAPs underwent further or alternate affinity purification steps. Several types of successful purification were performed in total, leading to purified: native *Sel*RNAP, native *Ssp*RNAP, native *Ec*RNAP, his-tagged *Ec*RNAP, his-tagged *Ssp*RNAP and FLAG-tagged *Sel*RNAP. Native cyanobacterial RNAP was also able to be purified at different phases of the circadian cycle when gene expression is thought to be at the most active and least active levels. Fractions were sent for mass spectrometry to confirm clean and effective purification of cyanobacterial RNAP, and to identify potential light phase-specific RNAP-binding proteins using FLAG-tagged purifications (**Appendix 5.4**).

RNAP subunits are clearly visible on representative SDS-PAGE and compared for each species in **Figure 8B**. There is little observable contaminating protein. RNAP purification by a two-step heparin and ion-exchange chromatography is a known method for the removal of any bound accessory factors, such as Gre, and can fractionate core enzyme from holo-enzyme (RNAP with σ factor). His-tagged *Ec*RNAP was purified from a Δ*GreA*Δ*GreB* strain to ensure no Gre factor would be present, and in all *Ec*RNAP purification there was no evidence of SCBF contamination which would present around 20 kDa. Where possible, fractions were collected with the aim of separating out the primary sigma factor σ^{70}, or SigA/RpoD1 in cyanobacteria, from the core enzyme.

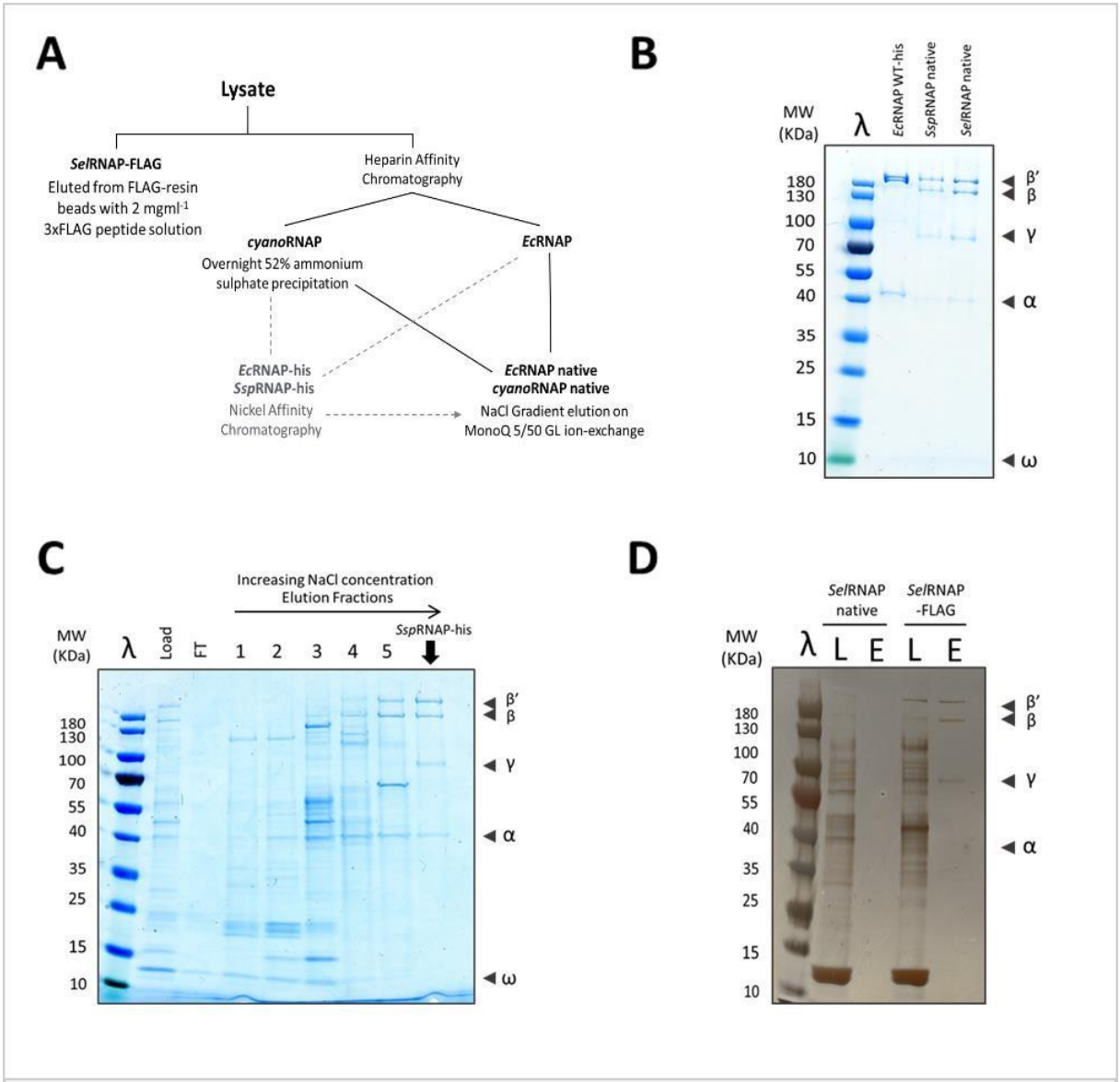


Figure 8. Purification of RNAP. (A) Overview of the methods used to purify native and tagged version of cyanobacterial and *E. coli* RNAP. **(B)** SDS-PAGE of *Ec*RNAP-his and both native cyanobacterial RNAPs after MonoQ chromatography and dialysis into storage buffer to show a comparison of subunit molecular weights and cleanliness of the final product. **(C)** SDS-PAGE showing fractionation of *Ssp*RNAP-his during MonoQ chromatography. Once run, gels were stained in InstantBlue (Expedeon). **(D)** Silver stained SDS-PAGE showing successful elution, E, of *Sel*RNAP-FLAG from FLAG resin beads. *Sel*RNAP native lysates were also tested to determine any non-specific binding to the beads. L, load fraction; FT, Flow-through.

3.1.3 Intrinsic Hydrolytic Activity is much Faster for Cyanobacterial RNAP

A major activity of RNAPs, and perhaps the activity most important regarding the aims of this study, is the ability of the enzyme to perform phosphodiester bond hydrolysis of the nascent RNA transcript. This reaction would usually be assisted through the recruitment of fidelity factor GreA but as the Gre SCBF is naturally absent in Cyanobacteria, assays were initially performed in the absence of any fidelity factor for *Ec*RNAP to allow measurement and comparison of the intrinsic cleavage ability of the RNA polymerase itself.

The first of these is described in **Figure 9,** where the hydrolysis rate of a cognate elongation complex, EC(GA), composed of fully complementary DNA and RNA is assessed. In this case, once the complex is assembled, a pause is simulated through a lack of nucleotides causing the RNAP to first backtrack and then subsequently cleave the 2 nucleotides of the RNA 3' end. Cyanobacterial RNAP was able to accomplish this cleavage approximately 80 times faster than *Ec*RNAP which would suggest very efficient intrinsic hydrolytic activity. To account for the impact of backtracking speed on these rates, the ability of each enzyme to enter a backtracked state was assessed and determined to be similar (Performed by Dr Yuzenkova, data not shown). This was also combated by assaying hydrolysis rates from RNAP complexes assembled to be in a -1 bp backtracked state, thereby eliminating the requirement to move to a backtracked position. These complexes represent an RNAP which has undergone a mis-incorporation event which are major targets of GreA.

Mis-incorporated complexes, MEC(A) and MEC(U), were assembled to give a mismatched and exposed RNA:DNA 3' end. The RNA 3' end of MEC(A) encodes AMP when the cognate would be UMP, and MEC(U) possesses a 3' UMP instead of a cognate CMP. As expected, rates of RNA cleavage from MECs are much faster for all RNAPs (**Figures 10 and 11**). The cyanobacterial RNAP rates of hydrolysis were again greater than those of *Ec*RNAP by 40-fold for MEC(U) and 30-fold for MEC(A). Notably, the rates for both species of cyanobacterial RNAP were comparable for each complex tested. All assays were performed at pH 7.9, 30^{o}C and were initiated with addition of 10 mM $MgCl_2$.

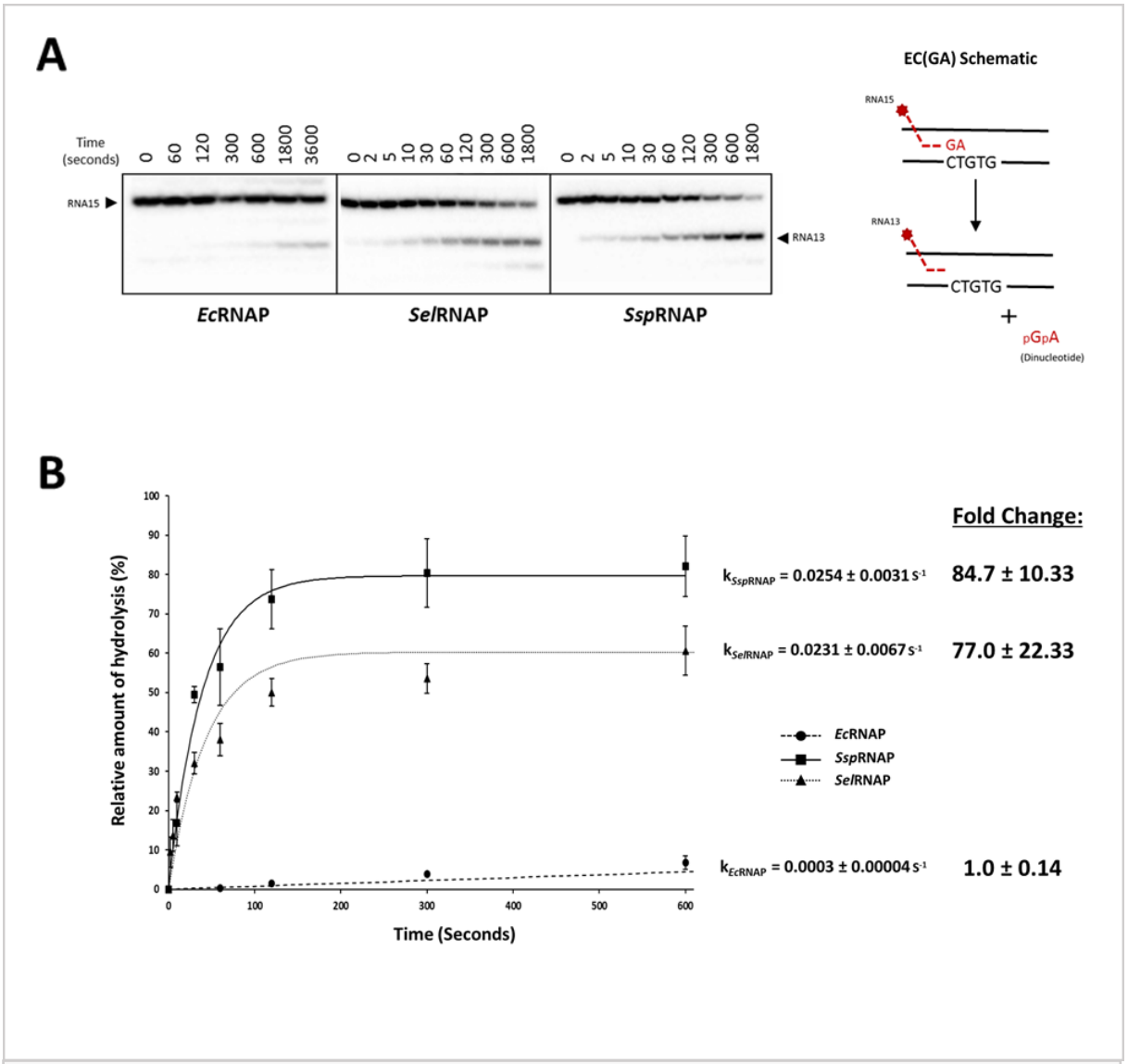


Figure 9. Cyanobacterial RNAP Performs Intrinsic Hydrolysis of a Cognate Elongation Complex Faster than *Ec*RNAP. (A) Representative 23% PAGE of each RNAP showing amount of RNA hydrolysis for the indicated time periods. Scheme of the cognate hydrolysis assay performed *Right:* RNA with 5' radiolabel highlighted red and paired between DNA strands in black. **(B)** Kinetics of phosphodiester bond hydrolysis fit with single exponent regression. Average rate for each RNAP shown to the right. The values for each RNAP are the mean ± standard deviations of at least three independent experiments. *Right:* Comparison of rates via fold change relative to WT *Ec*RNAP.

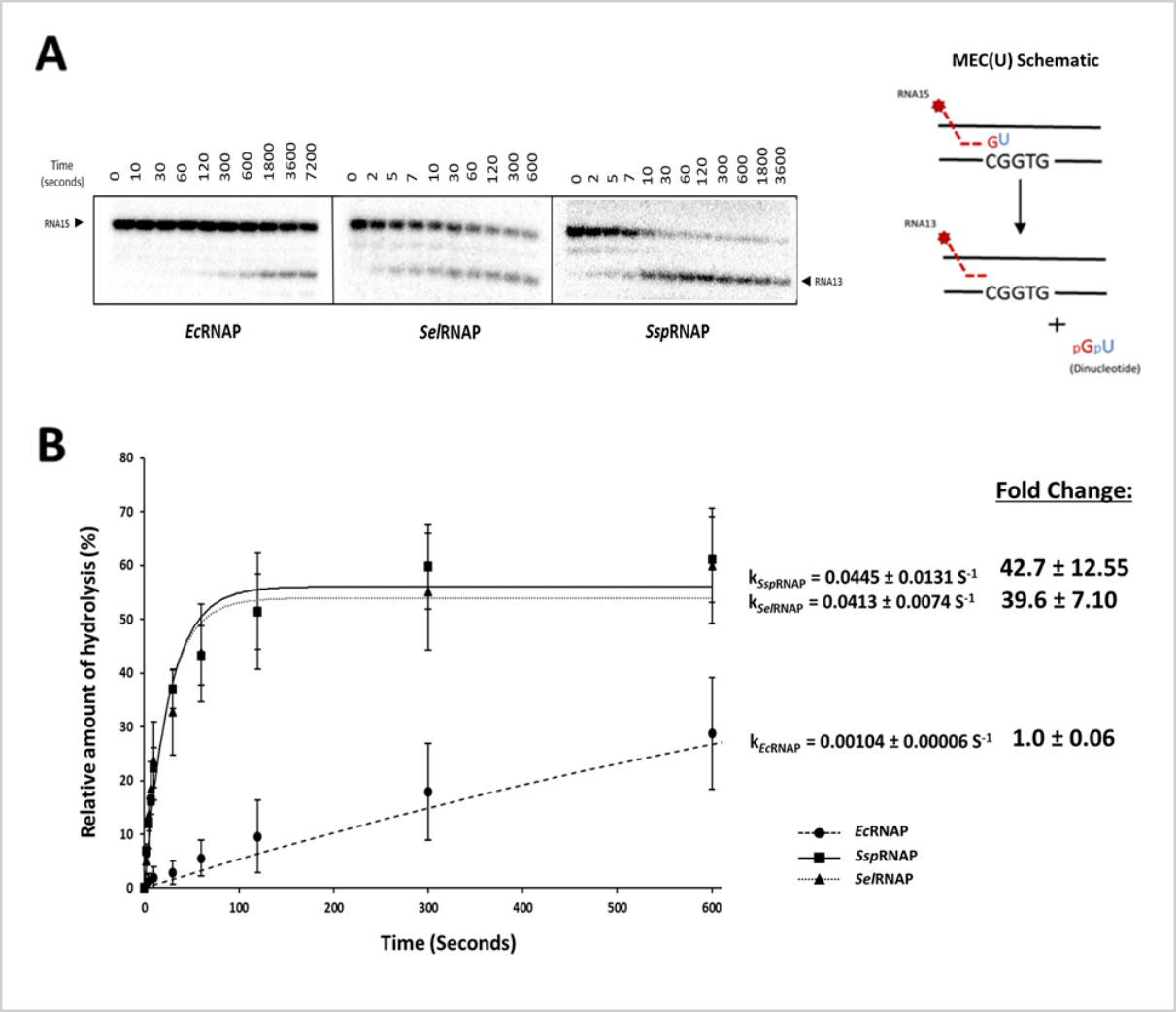


Figure 10. Cyanobacterial RNAP Performs Intrinsic Hydrolysis of a MEC(U) Non-Cognate Elongation Complex Faster than *Ec*RNAP. (A) Representative 23% PAGE of each RNAP showing amount of RNA hydrolysis for the indicated time periods. Scheme of the cognate hydrolysis assay performed *Right:* RNA with 5' radiolabel highlighted red and paired between DNA strands in black, mis-incorporated 3' nucleotide shown in blue. **(B)** Kinetics of phosphodiester bond hydrolysis fit with single exponent regression. Average rate for each RNAP shown to the right. The values for each RNAP are the mean ± standard deviations of at least three independent experiments. *Right:* Comparison of rates via fold change relative to WT *Ec*RNAP.

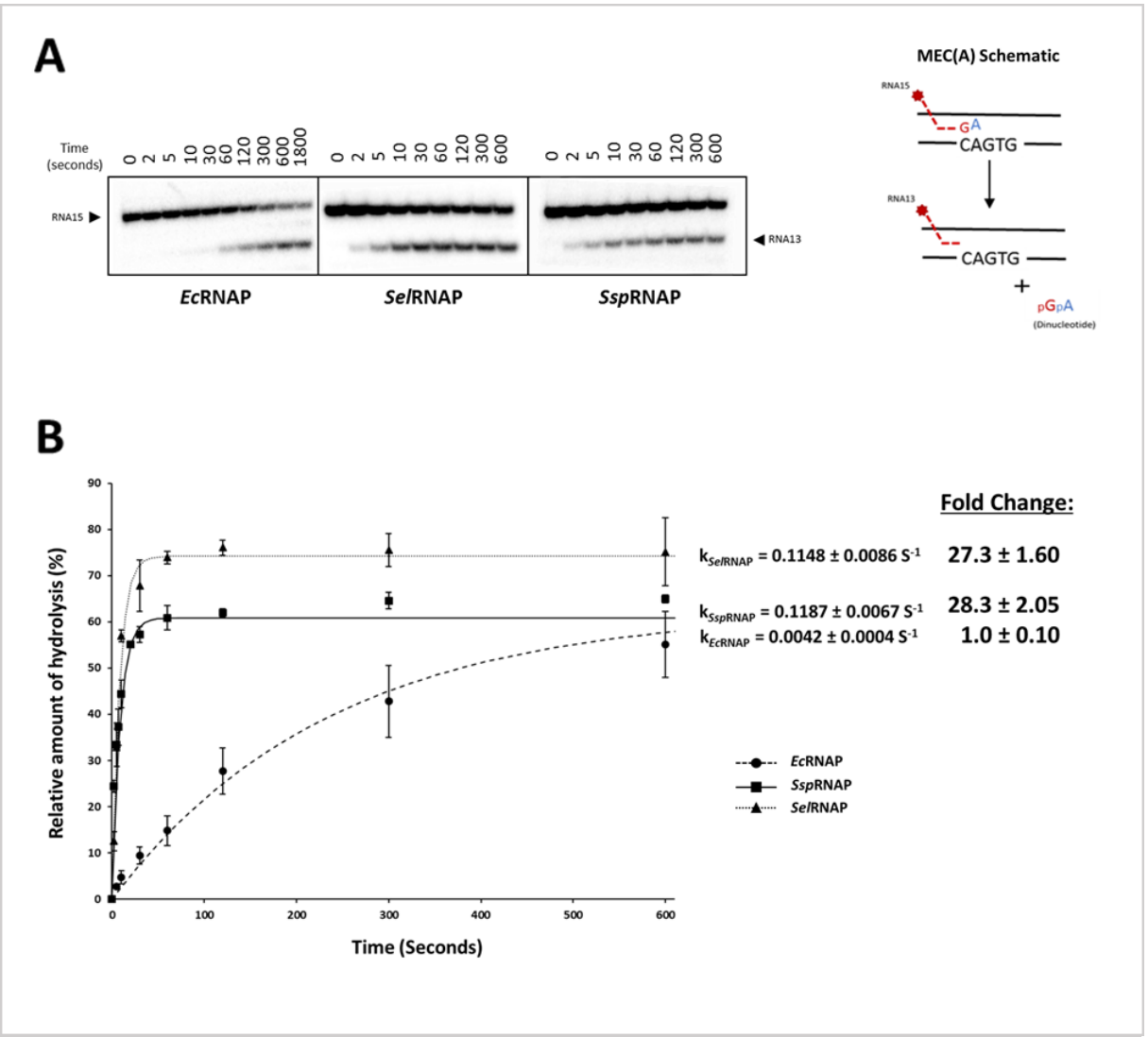


Figure 11. Cyanobacterial RNAP Performs Intrinsic Hydrolysis of a MEC(A) Non-Cognate Elongation Complex Faster than *Ec*RNAP. (A) Representative 23% PAGE of each RNAP showing amount of RNA hydrolysis for the indicated time periods. Scheme of the cognate hydrolysis assay performed *Right:* RNA with 5' radiolabel highlighted red and paired between DNA strands in black, mis-incorporated 3' nucleotide shown in blue. **(B)** Kinetics of phosphodiester bond hydrolysis fit with single exponent regression. Average rate for each RNAP shown to the right. The values for each RNAP are the mean ± standard deviations of at least three independent experiments. *Right:* Comparison of rates via fold change relative to WT *Ec*RNAP.

3.1.4 Mutations of the Trigger loop and Bridge Helix lead to Increased Intrinsic Hydrolytic Activity in EcRNAP

Catalysis occurs in the single active site of RNAP which is composed of the metal ion co-ordinating NADFDGD site, the mobile TL, and the supporting BH. The faster cleavage rates exhibited by cyanobacterial RNAP were investigated further by comparing variation in these active site elements. The ClustalW T-Coffee multiple sequence alignment program was used to compare *Ec*RNAP β' and cyanobacterial RNAP γ and β' protein sequences with particular interest in these regions. 24 single amino acid differences were observed across the three sites, highlighted in **Figure 12**.

These differences were used to identify any residues directly involved in producing the cyanobacterial RNAP increased cleavage ability through the creation of potential gain-of-function mutant *Ec*RNAPs. Site-directed mutagenesis was performed on his-tagged *Ec*RNAP *rpoC* (β') encoding pRL663 plasmid to substitute each of these amino acid differences with the respective cyanobacterial RNAP residue. In most cases both *S.el* and *S.sp* encoded the same amino acid, differing only at positions β' 474, 1135, and 773 (*Ec*RNAP numbering). Any mutations at these positions were generated using the corresponding *Ssp*RNAP amino acid. All single amino acid mutations depicted in **Figure 12** were constructed and purified as described in methods. A mutation of β' K789R the only exception as although a mutant construct of the correct sequence was made, all attempts at purification failed due to excessive production of inclusion bodies.

The purified *Ec*RNAP mutants were subjected to the same mis-incorporated MEC(A) and MEC(U) hydrolysis assays as described previously, with the results reported in **Table 1**. The table is shaded according to the fold change of cleavage activity of mutant *Ec*RNAPs compared to WT *Ec*RNAP, with darker shading indicating a greater increase in cleavage rate. Taking into account the increased activities of some single substitution mutants, particularly β' F773V, G1136Q, and A940V, generation of *Ec*RNAP mutants with multiple substitutions was attempted. Also in **Table 1**, a double mutant of β' G1136Q F773V produced the greatest increase in hydrolytic activity, however, a mutant combining the top three single substitutions of β' G1136Q F773V A940V was attempted but repeatedly failed transformation into the Δ*GreA*Δ*GreB E. coli* strain and so could not be purified. The hydrolysis rates of these top mutant *Ec*RNAPs near 30% of the rates for cyanobacterial RNAP as shown in **Figure 13**. Mutation structural locations are illustrated in **Figure 13B**, represented using *Ec*RNAP PDB 6GH5 (Glyde *et al*., 2018).

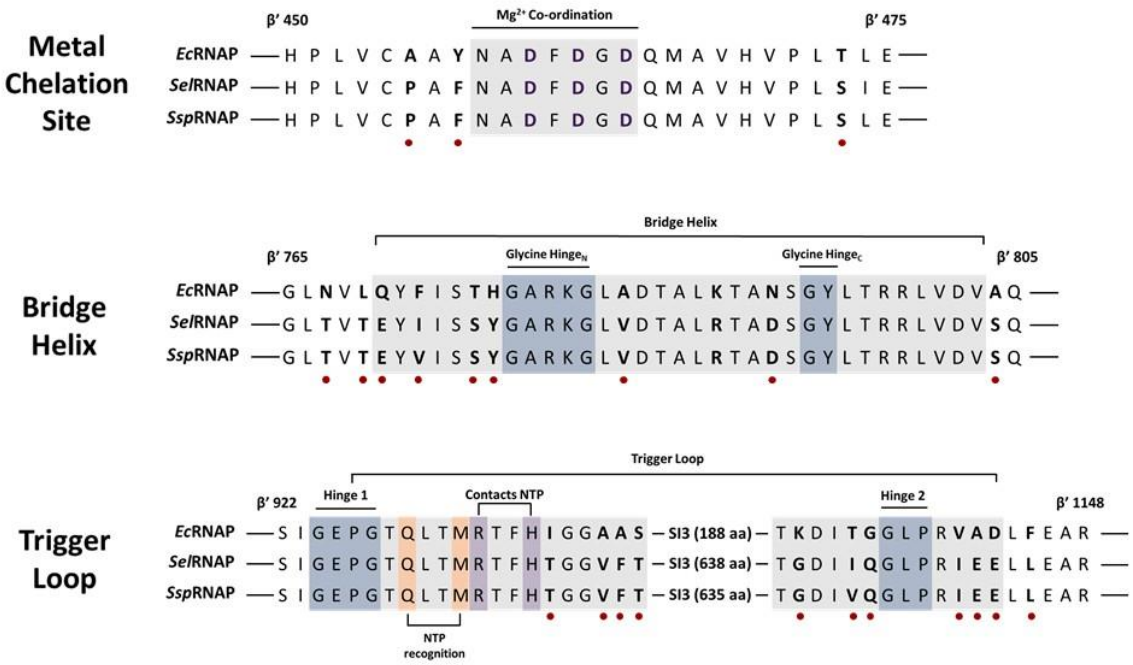


Figure 12. Sequence Alignment of the Active Site Elements. (Top) Aligned sequences for the *E. coli* and cyanobacterial Mg^{2+} I metal chelation site of the β' DPBB and conserved NADFDGD motif. **(Mid)** Aligned sequences of the β' BH with glycine hinge regions highlighted. **(Bot)** Aligned sequences of the TL showing positioning of the SI3 insert, conserved substrate NTP co-ordinating residues R933 and H936, and conserved NTP recognition residues Q929 and M932. Amino acid differences between the species are in bold. Red dots indicate the production and testing of a mutant *Ec*RNAP at this site.

Table 1. Overview of Mutant *Ec*RNAP Cleavage Rates. All mutant *Ec*RNAPs of the trigger loop (TL), bridge helix (BH), and metal chelation site (MCS) arranged in approximate descending order of hydrolytic activity on mis-incorporated complexes MEC(A) and MEC(U). Activity is expressed in fold change relative to WT *Ec*RNAP. Errors are indicative of the standard deviation of at least three repeats. * denotes only two repeats.

Fold Change of Cleavage Activity Relative to WT *Ec*RNAP:

Ec RNAP	Mutation Location	MEC(A) Cleavage	MEC(U) Cleavage
EcRNAP β' WT	-	**1.0 ± 0.10**	**1.0 ± 0.06**
β'G1136Q F773V	TL/BH	7.2 ± 0.38	8.7 ± 0.48
β'F773V	BH	6.5 ± 0.66	8.4 ± 1.51
β'A940V	TL	6.0 ± 0.93	11.6 ± 1.44
β'G1136Q	TL	3.1 ± 0.83	5.3 ± 0.72
β'N792D	BH	3.0 ± 0.46	5.8 ± 0.55
β'D1143E	TL	1.2 ± 0.18	3.4 ± 1.43 *
β'A804S	BH	1.4 ± 0.43	3.0 ± 0.75
β'Y457F	MCS	1.6 ± 0.26	2.8 ± 0.82
β'G1129K	TL	2.0 ± 0.19	1.6 ± 0.56
β'K1132G	TL	1.7 ± 0.16	1.0 ± 0.10
β'T776S	BH	1.7 ± 0.28	1.8 ± 0.34 *
β'A455P	MCS	1.6 ± 0.20	1.0 ± 0.15
β'I937T	TL	1.5 ± 0.08	1.9 ± 0.25
β'T474S	MCS	1.3 ± 0.05	1.6 ± 0.10
β'Q771E	BH	1.1 ± 0.16	1.4 ± 1.02 *
β'N768T	BH	1.0 ± 0.03	1.8
β'L770T	BH	1.4 ± 0.32	-
β'V1141I	TL	1.3 ± 0.17	-
β'A1142E	TL	1.8 ± 0.33	0.6 ± 0.15
β'F1145L	TL	0.9 ± 0.18	1.1 ± 0.45
β'A784V	BH	0.7 ± 0.24	1.0 ± 0.34 *
β'S942T	TL	0.9 ± 0.34	-
β'H777Y	BH	0.4 ± 0.23	-
β'A941F	TL	0.2 ± 0.04	0.5 ± 0.43

Fold Change: 1-2 2-3 3-4 4-5 More Than 5

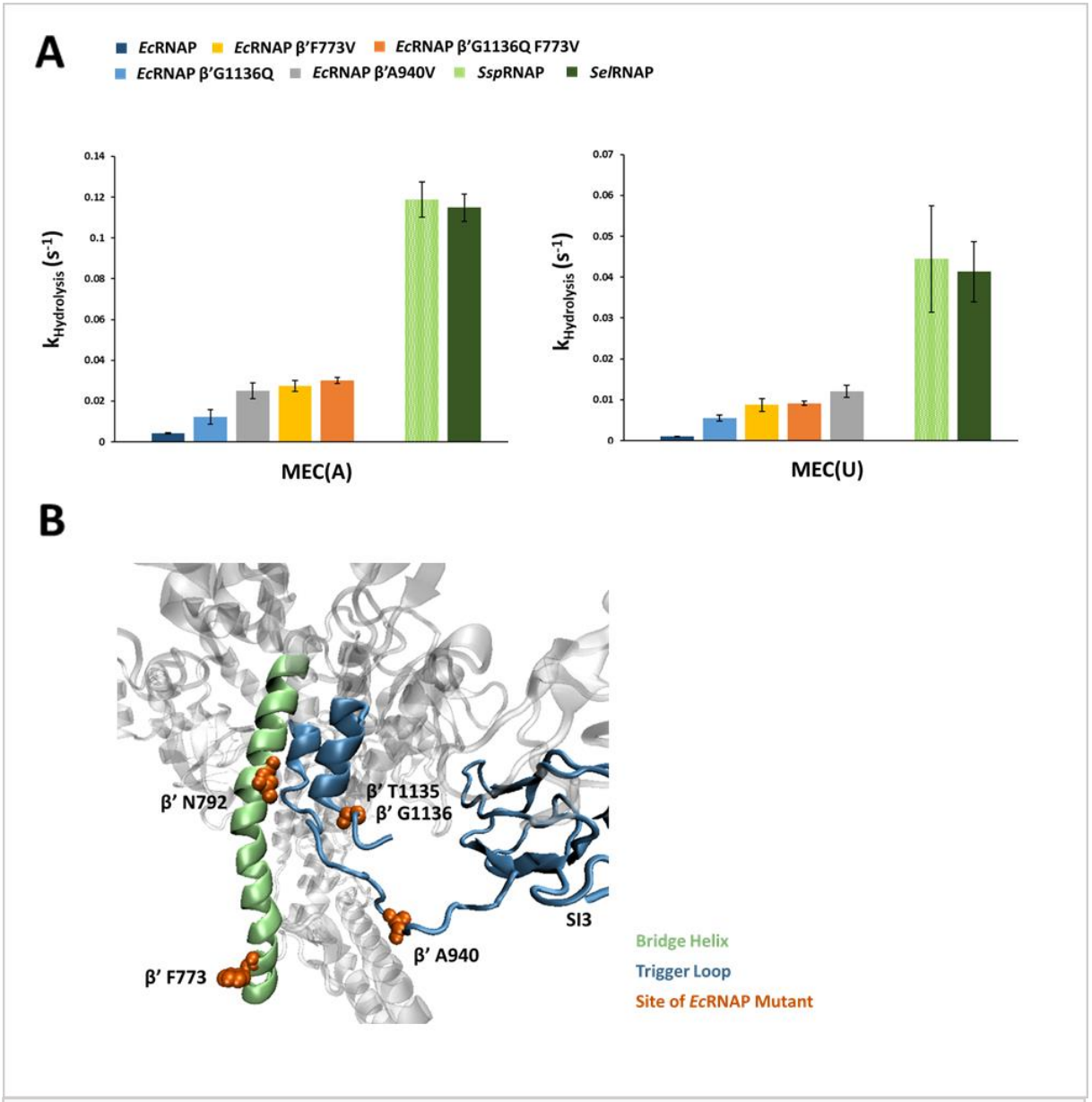


Figure 13. Mutant *Ec*RNAPs near a Third of the Hydrolysis Rate for Cyanobacterial RNAP. **(A)** Cleavage rate comparison for mutant *Ec*RNAPs and *Cyanobacterial* RNAPs on assembled mis- incorporated complexes MEC(A) and MEC(U). The values for each RNAP are the mean ± standard deviations of at least three independent experiments **(B)** Sites within the bridge helix (green) and trigger loop (blue) with mutations giving the highest increase in cleavage rate (orange). Some structure of the trigger loop insert could not be determined due to excessive flexibility. PDB 6GH5 (Glyde *et al*., 2018)

3.1.5 The Trigger Loop SI3 affects GreA-Mediated Cleavage

The SI3 is the lineage-specific insert which divides the trigger loop. It is a surface-exposed region of RNAP which lies close to the entrance of the secondary channel – the entry point for SCBFs like Gre to access the active site. The SI3 has been shown to modulate TL folding into the TH conformation thereby affecting the catalytic and pausing activities of RNAP (Windgassen *et al*., 2014; Artsimovich *et al.*, 2003).

*Ec*RNAP mutants of the SI3 were purified and tested for differences in hydrolytic activity compared to WT *Ec*RNAP. Two mutants were investigated; an *Ec*RNAP lacking the SI3 region β'ΔSI3(943–1130), and an *Ec*RNAP with the SI3 replaced by the much larger SI3 from *Ssp*RNAP, β'*Ssp*SI3 (**Figure 14A**). Plasmid pVS14, with his-tagged *E. coli rpoC*ΔSI3, was used to transform and purify the β'ΔSI3 mutant from the Δ*GreA*Δ*GreB* strain. β'*Ssp*SI3 was constructed and purified by Dr Yuzenkova. These mutant *Ec*RNAPs were first tested for hydrolytic activity on the mis-incorporated MEC(A) and MEC(U) complexes (**Figure 14**). This activity was significantly reduced in both complexes for β'*Ssp*SI3, whereas the β'ΔSI3 mutant demonstrated slightly reduced cleavage on MEC(A) and a complete absence of cleavage activity on MEC(U).

Considering the close location of the SI3 domain to that of the Gre factor binding site and entrance to the secondary channel, the SI3 *Ec*RNAP mutants were then tested for hydrolytic activity in the presence of 5 μM *Ec*GreA on cognate EC(GA) complexes (**Figure 15B**). Both the deletion and expansion of SI3 severely reduced GreA-mediated cleavage. The rate of RNA hydrolysis for WT *Ec*RNAP increases roughly 320-fold upon addition of GreA whereas that of *Ec*RNAP β'*Ssp*SI3 is only increased 13-fold, and worse still *Ec*RNAP β'ΔSI3 by 0.3-fold. Similar intrinsic cleavage rates were found for WT *Ec*RNAP and *Ec*RNAP β'*Ssp*SI3 however no rate could be determined for *Ec*RNAP β'ΔSI3 in the absence of GreA, even with incubation times up to 3 hours.

Addition of *Ec*GreA on cleavage rates for cyanobacterial RNAP were also briefly assessed. A pilot investigation found GreA factor able to impose some stimulatory activity between species (data not shown) however, these were species known to encode and use GreA factor. *Ec*GreA does not appear to affect the hydrolysis rates of cyanobacterial RNAP save for a small decrease (**Figure 15A**). This decrease may be explained by the high glycerol and salt content of the purified GreA storage buffer (**SB**) as addition of an equivalent volume of only SB also results in a decreased rate.

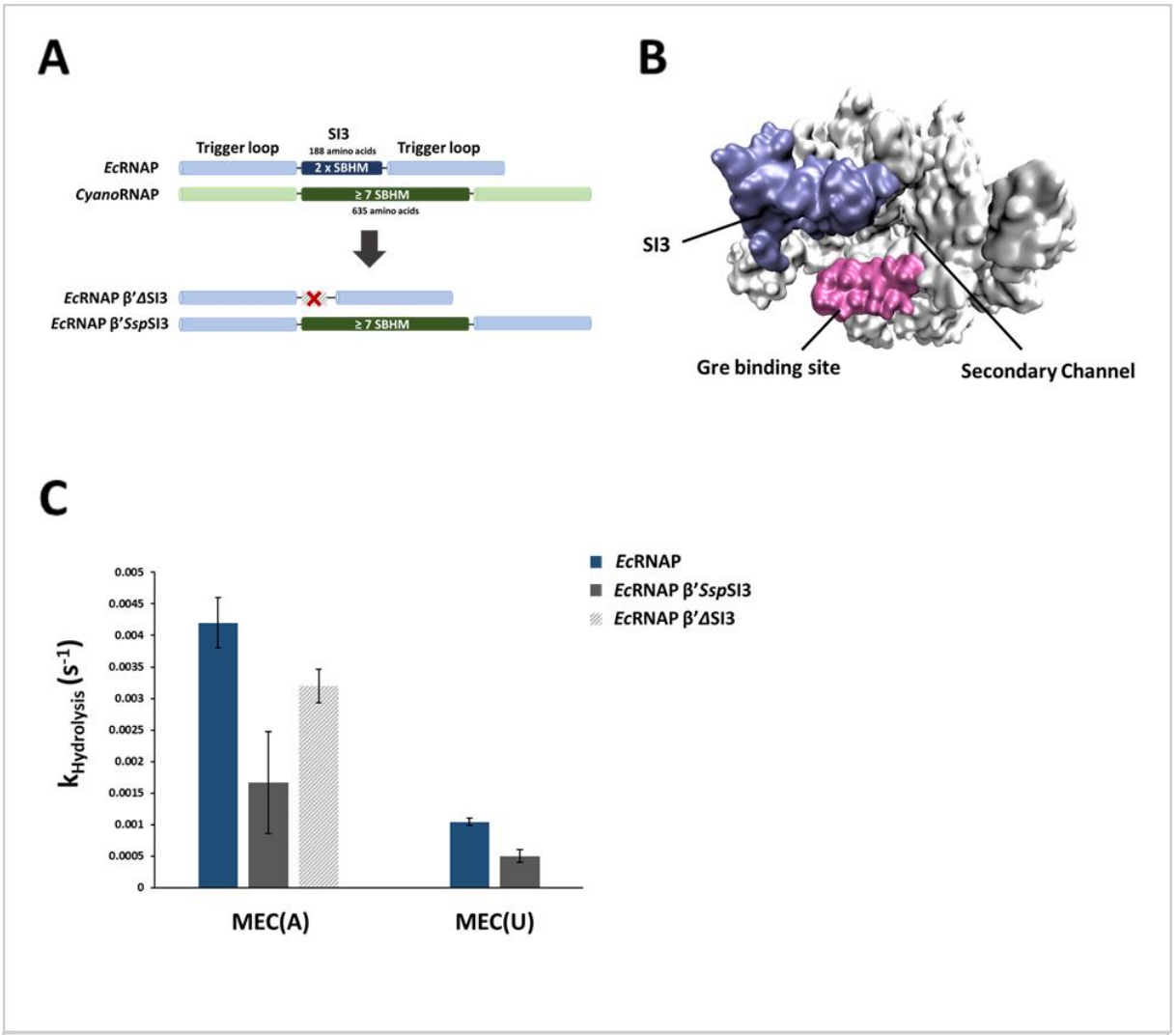


Figure 14. Mutations of the SI3 affect Intrinsic Hydrolytic Activity (A) Schematic of the SI3 *Ec*RNAP mutants. *Ec*RNAP β'ΔSI3 is a deletion of the entire *E. coli* insert whereas in β'*Ssp*SI3, the larger SBHM-encoding SI3 has replaced the shorter native *E. coli* SI3. **(B)** Space fill of *Ec*RNAP showing positions of the SI3 (blue), Gre binding site (pink), and entrance to the secondary channel. PDB 6ALH (Kang *et al*., 2017). **(C)** Hydrolysis rates of SI3 mutants for mis-incorporated MEC(A) and MEC(U) complexes. The values for each RNAP are the mean ± standard deviations of at least three independent experiments.

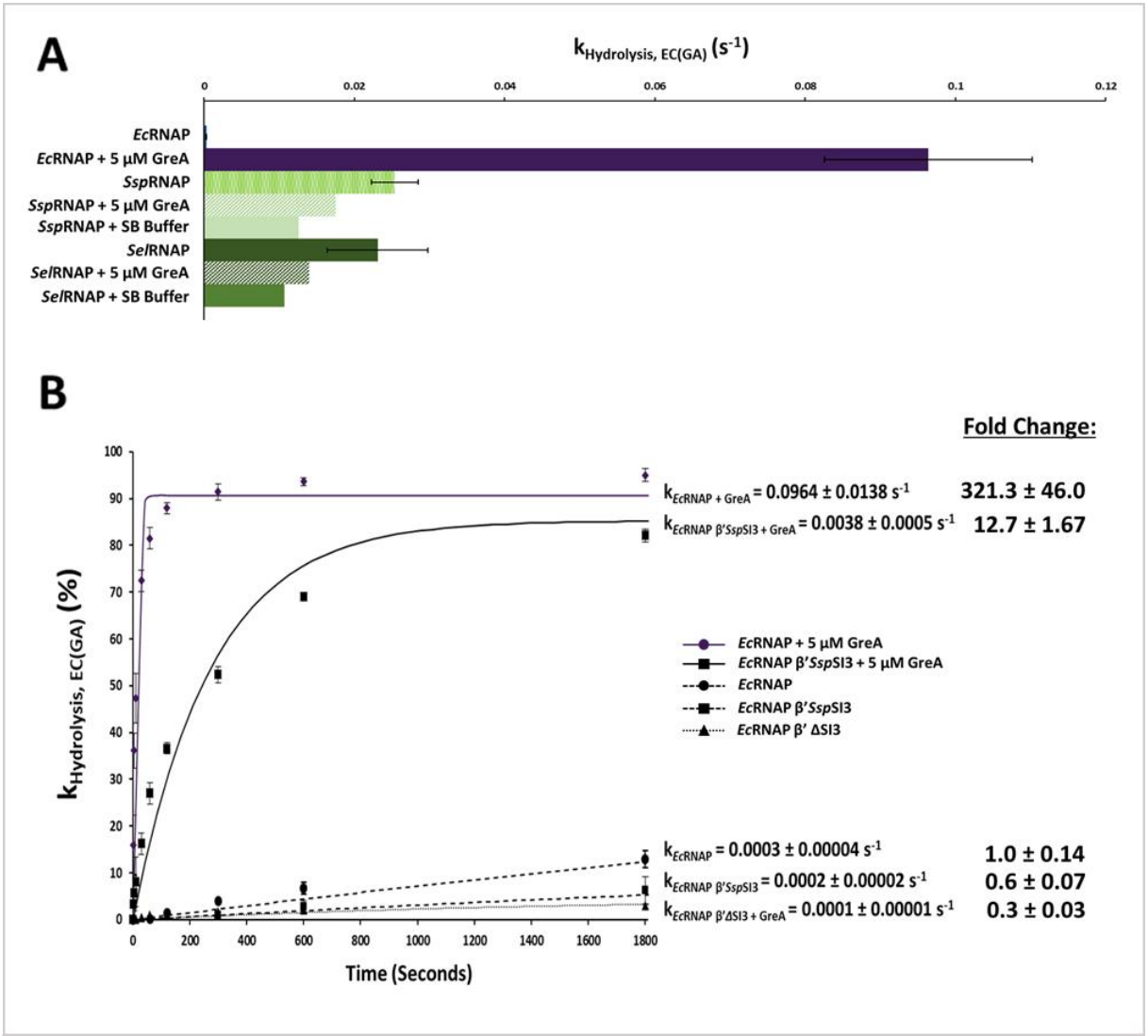


Figure 15A. GreA does not Stimulate *Cyanobacterial* RNAP Hydrolytic Activity. 15B. GreA- Mediated Hydrolysis is Impaired for SI3 *Ec*RNAP Mutants (A) Comparison of hydrolysis rates on cognate complexes EC(GA) with and without addition of 5 µM *Ec*GreA or an equivalent volume of storage buffer (SB). **(B)** Kinetics of phosphodiester bond hydrolysis fit with single exponent regression for hydrolysis rates of SI3 *Ec*RNAP mutants with and without addition of 5 µM *Ec*GreA on cognate EC(GA) complexes. Average rate for each RNAP shown to the right. The values for each RNAP are the mean ± standard deviations of at least three independent experiments. (β'ΔSI3 + GreA only performed once, error bars indicate fit of data). *Right:* Comparison of rates via fold change relative to WT *Ec*RNAP.

3.1.6 Discussion

It was theorised that since GreA and other secondary channel binding factors (**SCBFs**) were absent in cyanobacteria, an alternate mechanism may exist to perform or compensate for the loss. To investigate this, the enzyme responsible for transcription, RNAP, was purified from two cyanobacterial strains and model γ-proteobacterium *E. coli* in order to compare *in vitro* transcription activity. In this section, the hydrolytic activity of each RNAP was assessed on both cognate complementary assembled complexes EC(GA) and non-cognate mis-incorporated complexes (MEC(A)/(U)) which simulate a paused or -1bp backtracked RNAP respectively.

Both cyanobacterial RNAPs were able to perform RNA hydrolysis approximately 80 times faster than *Ec*RNAP for cognate complexes and 30-40 times faster for backtracked complexes (**Figures 9-11**). This suggests that cyanobacterial RNAP is capable of much faster intrinsic cleavage than *Ec*RNAP and is not just faster at entering the necessary backtracked state from which cleavage occurs. An RNAP able to carry out fast and efficient cleavage without the need for any accessory factor could constitute a viable alternate mechanism which compensates for the absence of Gre factors. However, a few considerations need to be taken into account. These include whether there is a minimum rate threshold for cleavage to be effective in the cell, and if the intrinsic cleavage ability of cyanobacterial RNAP meets this rate. Similarly, GreA acts as a fidelity factor to promote excision of the RNA 3' end when a mis-incorporation event has occurred, this may not be necessary if cyanobacterial RNAP has more stringent error prevention mechanisms or if cyanobacteria can tolerate less accurate RNA. The Gre factors are also thought to act upon potentially harmful stalled RNAPs, most often caused by mis-incorporation events, and resolve conflicts with the replication machinery. This may not be required if the cleavage ability demonstrated by cyanobacterial RNAP is able to efficiently resolve stalled complexes, or if mis-incorporation-caused stalled RNAPs are rare in cyanobacteria due to initial competent cleavage of the transcript. Efforts to investigate and address these concerns while attempting to find the source of increased cleavage ability in cyanobacterial RNAP are described below and in following chapters.

Differences in sequence homology of the active site elements between *Ec*RNAP and cyanobacterial RNAP was initially considered as the cause for the increased hydrolysis observed. *Ec*RNAP mutants were constructed with substitutions in the β' subunit to match the relative sites in *Ssp*RNAP in the expectation that inserting a residue important for cleavage activity would then lead to an improved cleavage rate for the *Ec*RNAP mutant. The *Ssp*RNAP sequence was chosen as *Ssp*RNAP gives a slight but consistently higher RNA cleavage rate in

all complexes tested. There was also little variation between the active site sequences of each cyanobacterial RNAP, with differences at three positions β' L474, V1135, V773 for *Ssp*RNAP which all code for isoleucine in *Sel*RNAP (*Ec*RNAP numbering). All of these are branched-chain amino acids, highly hydrophobic and important in determining protein structure (Monirujjaman and Ferdouse, 2014). Although substitutions between these amino acids are generally considered to be conservative, there are examples where single substitutions have resulted in alterations of protein function. While a purpose may exist for these changes, perhaps in the greater reliance of *Sel* on circadian control with photoautotrophic growth, both cyanobacterial RNAPs react similarly in the *in vitro* experiments used in this thesis.

In total, five of the constructed *Ec*RNAP mutants demonstrated cleavage rates that were more than triple the WT *Ec*RNAP rate for both MEC(A) and MEC(U). Three mutants, β' A940V, F773V, and G1136Q F773V gave rates at least six times greater than WT *Ec*RNAP (**Table 1**). The most improved *Ec*RNAP mutants encoded substitutions in the mobile components of the active site, the bridge helix (**BH**) and the trigger loop (**TL**). Previous studies in different species of RNAP have found that mutations in these regions can significantly affect catalytic activity by altering the flexible nature of these elements.

The active site-spanning BH undergoes conformational changes between 'kinked' and 'straight' states during the nucleotide addition cycle. This is largely enabled by two glycine hinges within this long α-helix. Two *Ec*RNAP mutants constructed during this project, β'F773V and β'N792D, increased the cleavage rate of WT *Ec*RNAP at least 3-fold. β'F773V is located in a hydrophobic segment of the BH, termed the YFI motif, which is positioned just before the N-terminal glycine hinge. Nedialkov *et al*. (2013) has shown that amino acid substitutions of this motif considerably alter pausing, termination and fidelity of RNAP. β'F773V is a well-established *Ec*RNAP mutant within the literature after first being identified as able to confer dependence on the antibiotic CBR703 (Artsimovich *et al*., 2003). This mutation is known to suppress RNAP pausing, have stiffer translocation ratchets, and show reduced fidelity (Nedialkov *et al*., 2013; Svetlov *et al*., 2007). In our assays, β'F773V demonstrates a much-improved intrinsic hydrolytic activity of 6.5-fold and 8.4-fold over the WT *Ec*RNAP activity on respective MEC(A) and MEC(U) scaffolds. It is thought that the valine substitution of β'F773V results in a stiffer BH which reduces conformation dynamics. This gives the mutant a 'restrained ratchet', meaning it is less likely to oscillate between translocation states (Nedialkov *et al*., 2012). The YFI motif is suggested to form a brace against which the N-terminal BH hinge can bend, with further contacts between each amino

acid of this motif and other surrounding domains of the active site affecting catalytic activity. β'F773 interacts with the β fork loop, a highly conserved motif thought to maintain the downstream edge of the transcription bubble and assist with translocation following nucleotide addition (Jovanovic *et al*., 2011; Vassylyev *et al*., 2007). Malinen *et al*. (2014) has additionally showed that disrupting BH-β subunit contacts stabilises the folded TL and closed RNAP active site conformation leading to the increased catalysis, slower translocation, and pause-insensitivity seen in β'F773V.

The other mutant of the BH showing significantly increased cleavage activity was β'N792D. This β'N792 has also been proposed to affect translocation as it has been observed in two conformations, within or flipped out of the BH (Temiakov *et al*., 2005). Crystal structures of *Thermus thermophilus* and *Thermus aquaticus* have placed the orthologous residue of β'D1090 in a flipped-out configuration where it assists in stabilising the kinked BH conformation through hydrogen bonding with nearby residue β'R1096 (Zhang *et al*., 1999; Vassylyev *et al*., 2002). Aspartate-arginine side-on interactions within side chains are able to strongly stabilise intramolecular interactions (Mitchell *et al*., 1992). Tan *et al*. (2008) conducted exhaustive mutagenesis of the BH of *Methanocaldococcus jannaschii* RNAP. In their paper, a mutant of *mj*A'Q823D (orthologous to β'N792) increased catalysis by 135% over WT *Methanocaldococcus jannaschii* RNAP, which they suggest is caused by mimicking the aspartate-arginine stabilisation responsible for the kinked BH conformation. The side chain of *mj*A'Q823 was shown to point away from the RNAP catalytic center leading Tan *et al*. (2008) to conclude that increased RNAP activity is caused by changes in BH conformations and subsequent interactions with the TL. It is possible that the *Ec*RNAP β'N792D mutant we present here also forces kinked BH conformation by making these D-R hydrogen bonds and resulting in a stabilised catalytically active conformation, similar to β'F773V. Indeed, *Ec*RNAP β'N792 is positioned close to the base helices of the TL and could affect TL folding stability (**Figure 13B**).

The mobile TL is a positional catalyst of the active site required for efficient catalysis. It also undergoes conformational change by folding into a two α-helical hairpin-like formation, known as the trigger helices (**TH**) (Vassylyev *et al*., 2007). These TH represent the 'closed/active' conformation of RNAP, where two residues in particular β'H936 and R933 assist in co-ordinating nucleophilic attack and enhance rate of catalysis (Mishanina *et al*., 2017; Yuzenkova *et al*., 2010). Several studies have shown that mutations in this region can greatly affect all aspects of RNAP activity, with changes in elongation rate often being accompanied with altered pause behaviour (Bar-Nahum *et al*., 2005; Toulokhonov *et al*.,

2007). In this thesis, two single substitution mutants of the TL displayed increased intrinsic cleavage ability, β'G1136Q and β'A940V. These mutants showed, on average, a respective 4 and 9-fold faster hydrolysis rates on MEC scaffolds than WT *Ec*RNAP. β'A940V maps to the N-terminal region of the TL before the SI3 insert, whereas β'G1136Q is located after the SI3 and just before hinge 2 of the TL. A double mutant of β'G1136Q F773V was also tested and demonstrated faster cleavage activity than either single mutant. This activity does not appear to be additive, however, as the rate is only slightly higher than that for β'F773V, suggesting that these substitutions may increase activity through the same mechanism or pathway which has reached maximal effect. β'G1136 is the site of a known 'fast' RNAP mutant, β'G1136S (Bar-Nahum *et al*., 2005). β'G1136S increases RNAP elongation rate with less discrimination of substrate NTP, decreases responses to pausing, and displays inefficient termination (Bar-Nahum *et al*., 2005). Similar to the BH mutants presented above, β'G1136S is thought to modulate TL-TH oscillation by stabilising the three-helix bundle (**THB**) formed from TH contacts with the BH to promote a closed and catalytically active RNAP conformation (Mejia *et al*., 2015; Zhang *et al*., 2010). There is currently no documentation in the literature for any substitutions at the site of β'A940V having an effect on RNAP activity. This site is in between the NTP-contacting reside of β'H936 and the large SI3 which bisects the TH, introduction of a bulky hydrophobic residue here may restrict flexibility in this region and affect TL-TH oscillation.

The mode of action of these mutants in increasing catalysis by stabilising a closed active site conformation has also been shown in eukaryotic RNAP. A TL rpb1 subunit E1103G substitution in RNAP II of *Saccharomyces cerevisiae,* which increases rates of elongation and mis-incorporation, is thought to do so by promoting a closed active site conformation with a folded TL (Malagon *et al*., 2006; Kireeva *et al*., 2008). We propose that as the substitutions described (β'F773V, β'N792D, β'G1136Q, β'A940V) occur naturally in cyanobacterial RNAP that this enzyme exists in an inherently more closed and catalytically active conformation. Turtola *et al*. (2018) has recently shown that β'F773V and β'G1136S stabilise and decrease recovery from the backtracked state. As this is the necessary state from which RNA hydrolysis occurs, it complements our data showing increased hydrolytic activity for these mutants.

In vivo, hydrolysis of paused and backtracked complexes would usually be stimulated by the SCBF GreA in *E. coli*. To assess this stimulation, the hydrolysis rates of EC(GA) were also measured in the presence of 5 μM *E. coli* GreA (**Figure 15A**). As expected, the rate of RNA hydrolysis increased greatly for *Ec*RNAP at roughly 320-fold faster than the intrinsic

cleavage rate for *Ec*RNAP. A small decrease in rate was observed if *E. coli* GreA was added to the cyanobacterial RNAP complexes. This decrease persisted when only storage buffer was added and so is thought to be due to the high glycerol (50%) and salt (0.8 M NaCl) concentration of the buffer. However, this assay showed that the intrinsic cleavage rates of cyanobacterial RNAPs only reach approximately 30% of the cleavage rate for *Ec*RNAP when GreA is present. It is possible that this is an adequate level of cleavage activity in cyanobacteria considering the slower growth rate and much longer doubling times of 6-7 hours for *S.e*l and 12 hours for *S.sp* (Vermass *et al*., 1988; Mori *et al*., 1996). These species are also polyploid, which presents additional options for the organisation of replication and transcription. Chen *et al*. (2012) showed that chromosome replication in *S.el* is not only asynchronous but that there is only one actively replicating chromosome at a time in most cells. It is unknown whether this same chromosome is used for transcription. The ability to co-ordinate replication on one chromosome and transcription on another would remove the consequences of replication machinery collisions with any stalled RNAPs and would mean that cyanobacteria may only have to resolve stalled 'traffic jams' of RNAP. In this case, a slower intrinsic cleavage rate than that of factor-assisted in *Ec*RNAP may be sufficient to carry out all necessary functions of cleavage, while allowing the enzyme factor-independence.

SI3 is a surface-exposed domain situated close to the entrance of the secondary channel and divides the two TH of the folded TL. The integration within the TL allows SI3 to modulate TL folding and affect the catalytic and pausing activities of RNAP (Windgassen *et al*., 2014; Artsimovitch *et al.*, 2003). Cyanobacterial RNAP is unusual in that the lineage-specific SI3 of the TL is much larger than that found in other species. In order to test if this larger SI3 conferred altered activity behaviours, two mutants of *Ec*RNAP were constructed and tested for hydrolytic activity (**Figures 14 and 15B**). The 188 amino acid SI3 of *Ec*RNAP was removed for the β'ΔSI3 mutant whereas for β'*Ssp*SI3, the SI3 of *Ec*RNAP was replaced with the larger 635 amino acid SI3 of *Ssp*RNAP. β'*Ssp*SI3 showed diminished cleavage activity on both MEC(A/U) scaffolds and on cognate complex EC(GA), while β'ΔSI3 demonstrated completely absent cleavage activity on MEC(U) and EC(GA) complexes. Hydrolytic activity on MEC(A) was only slightly reduced for β'ΔSI3 which was surprising and is thought to be due to the nature of the scaffold with perhaps a greater ability to co-ordinate transcript-assisted cleavage. Both mutants also showed drastically reduced activity to WT *Ec*RNAP when in the presence of *E. coli* GreA. Activity for β'ΔSI3 + GreA was barely observable, whereas the addition of GreA to β'*Ssp*SI3 only increased the rate 12.7-fold when GreA is able to stimulate 320-fold activity for WT *Ec*RNAP. These results indicate that the consequences of removing SI3 in *Ec*RNAP are more deleterious for RNA cleavage than expansion of the

insert, suggesting that it may have some positional effect in allowing the TL to fold properly in intrinsic hydrolysis. Transcript cleavage has previously been shown to be reduced in partial β'SI3 deletions in *Ec*RNAP in work which was particularly focussed on identifying mutants defective in interactions with transcript cleavage factors (Zakharova *et al*., 1998). Zhang *et al*. (2010) was later able to determine that SI3 is essential for the action of cleavage factor GreB in *E. coli* but was unable to conclude whether SI3 assisted in binding or function of GreB. Our data suggest that SI3 may carry out a similar role for the action of GreA.

In this section, we have demonstrated that cyanobacterial RNAP has a fast intrinsic hydrolytic activity which may be caused by specific amino acids of the flexible BH and TL active site elements. The distinct amino acids identified are either known, or are at positions known, to reduce the flexibility of these elements in *Ec*RNAP, resulting in a more closed and catalytically competent active site. This conformation of the active site leads to increased rates of nucleotide addition and elongation with an insensitivity to pause signals. Several studies have also identified β'SI3 mutants of *Ec*RNAP to reduce pausing during elongation, and particularly to increase the rate of escape from hairpin-dependent pause signals (Artsimovich *et al.*, 2003; Ray-Soni *et al*., 2017). Characterisation of cyanobacterial RNAP and the *Ec*RNAP mutants in elongation is presented in future results sections. First, we investigated further differences in the active site which may alter catalysis, including the contribution of intrinsic transcript-assisted hydrolysis.

3.2 The *Cyano*RNAP Active Site may be Configured for Faster Catalysis

3.2.1 Introduction and Aims

Cyanobacterial RNAP demonstrated a much higher capacity for intrinsic RNA hydrolysis than *Ec*RNAP, possibly due to specific residue changes in the active site elements which result in a more closed and catalytically competent enzyme. The active site mechanics and RNAP enzymology were then investigated further in this section to determine if there were any differences in the basic reaction mechanism of each RNAP. In further cleavage assays using MEC(A/U), the binding affinity for the transiently sequestered Mg^{2+}II ion was tested for each RNAP and the activation energy barrier for cleavage was assessed. The possible involvement of a general base in activating the attacking water molecule and therefore increasing activity in cyanobacterial RNAP was examined. The contribution to the intrinsic cleavage rate of co-ordination via transcript-assisted cleavage was also measured.

Binding Affinity for the Mg^{2+} ion is Similar for EcRNAP and SspRNAP

RNAP requires two divalent metal ions, usually Mg^{2+}, for catalysis via the two-metal-ion mechanism. Mg^{2+}I is persistently bound in the active site by the aspartate triad of the β' subunit, while Mg^{2+}II is only transiently sequestered. During hydrolysis, Mg^{2+}II is responsible for activating the attacking group. Metal binding affinity for the active site can impact catalytic activity and so was proposed to be a cause of the increased cleavage activity seen for cyanobacterial RNAP.

The affinity of a metal ion for a protein is described by the dissociation constant, K_d. A smaller K_d would represent a more tightly bound metal ion. In this assay, we assume K_d to be approximately equal to K_m and fit our data using Michaelis-Menten non-linear regression to determine the V_{max} and K_m for each enzyme. The rate of transcript hydrolysis was measured for both *Ec*RNAP and *Ssp*RNAP on MEC(A) and MEC(U) complexes using a range of activating Mg^{2+} concentrations.

V_{max} for both MEC(A) and MEC(U) was much faster for SspRNAP as expected (**Figure 16**). However, the K_m for each enzyme was very similar at roughly 21 mM for MEC(U) and 4 mM for MEC(A). This reflects the faster hydrolysis rates seen when using the MEC(A) assembled complexes from the previous results section. The data for MEC(A) were collected by Dr Yuzenkova and presented here with permission. The similar K_m values obtained for *Ec*RNAP and *Ssp*RNAP means that they are considered to have an equal affinity for the Mg^{2+} ion and that the increased cleavage rates of *Ssp*RNAP are from a different cause.

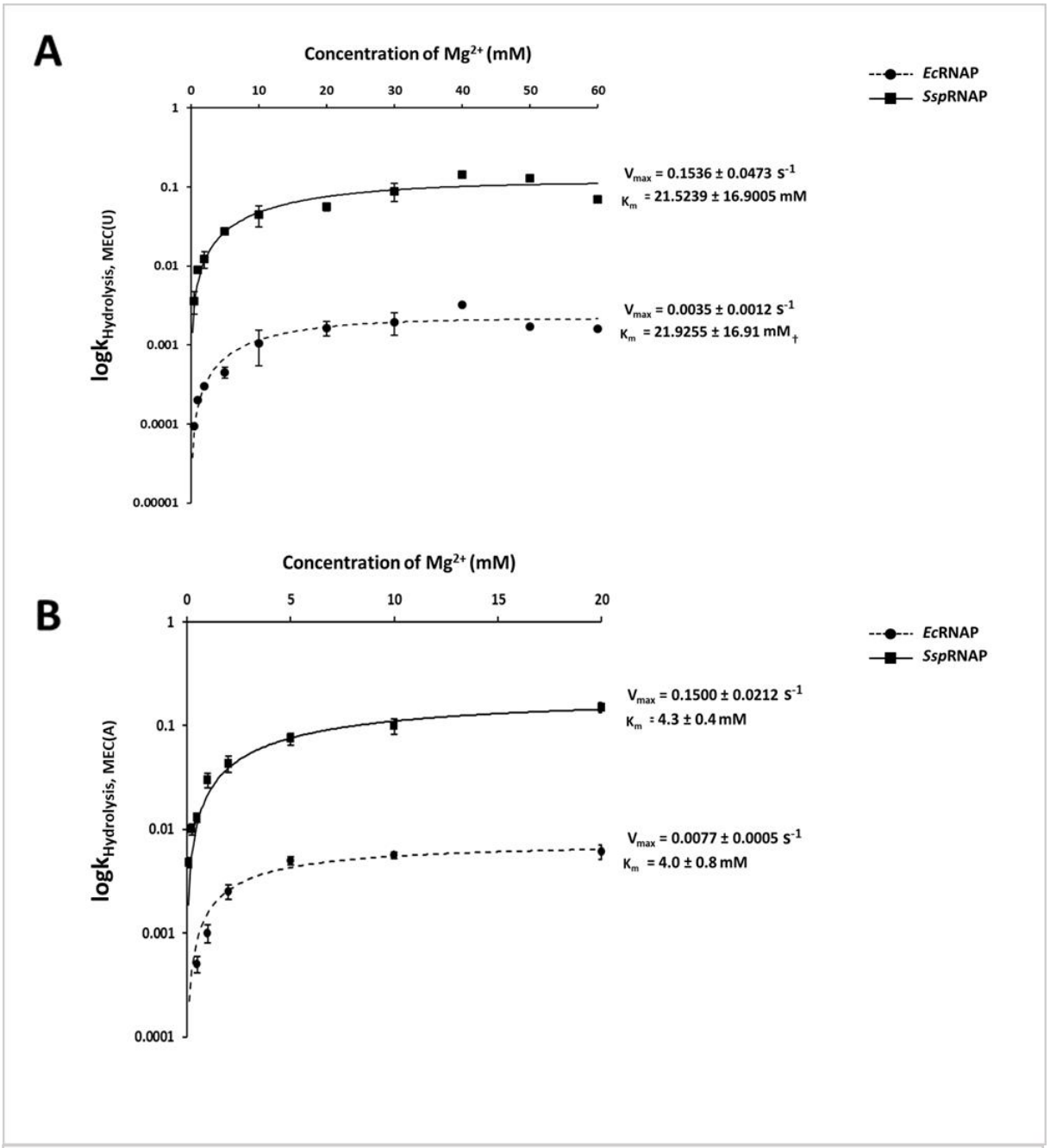


Figure 16. Binding Affinity for Mg^{2+} II is Similar for *Ec*RNAP and *Ssp*RNAP. (A) Rates of RNA hydrolysis on MEC(U) against buffer Mg^{2+} concentration fit using Michaelis-Menten non-linear regression to obtain Vmax and Km for *Ec*RNAP and *Ssp*RNAP. † Vmax and Km for *Ec*RNAP were calculated using data only up to 50 mM Mg^{2+} to combat suspected substrate inhibition at 60 mM. **(B)** Rates of RNA hydrolysis on MEC(A) against buffer Mg^{2+} concentration fit using Michaelis- Menten non-linear regression to obtain Vmax and Km for *Ec*RNAP and *Ssp*RNAP. The values for each RNAP are the mean ± standard deviations of at least three independent experiments.

3.2.2 Less Activation Energy is Required for Hydrolysis by Cyanobacterial RNAP

The Activation Energy, $E_{a,}$ is the minimum energy which must be overcome before a reaction can occur. In order to ascertain whether the E_a required for hydrolysis by *Ec*RNAP and *Ssp*RNAP were similar, rates of transcript cleavage on MEC(U) were measured over a range of reaction temperatures. These data were fit to an Arrhenius plot as described in the methods. E_a can be calculated from the negative slope of this plot using $y = -E_a/R$, where R is the gas constant equal to 8.314 $Jmol^{-K}$. **Figure 17A** shows the Arrhenius plots and E_a calculations for both RNAPs. The energy required for RNA hydrolysis was much lower for *Ssp*RNAP, at about half the amount (61.5 $KJmol^{-1}$) required for *Ec*RNAP (123.5 $KJmol^{-1}$). This lower E_a requirement could contribute to the fast and efficient intrinsic hydrolysis demonstrated by cyanobacterial RNAP.

3.2.3 Deprotonation may be Assisted by a General Base in Cyanobacterial RNAP

The two-metal-ion mechanism of RNAP requires deprotonation during hydrolysis where Mg^{2+}II activates the attacking water molecule. The existence of a general base for deprotonation has been identified as the TL β'H936 reside in some bacterial species but appears to differ greatly in contribution between species (Svetlov and Nudler, 2013; Yuzenkova *et al*., 2010). Elevation of buffer pH is expected to stimulate hydrolysis as water deprotonation would be increased, however, the reaction rate should plateau at a pH equal to the water pKa. pH has also been shown to affect Mg^{2+}II retention, however as the K_m for Mg^{2+}II was determined to be similar in the previous section, all RNAPs are assumed to react comparably (Sosunov *et al*., 2003). Hydrolysis rates for MEC(A) were measured across a buffer pH range to determine deprotonation effects on activity for *Ec*RNAP, *Sel*RNAP, and *Ec*RNAP + 0.2 µM GreA (**Figure 17B**). pH elevation from 6.5 to 7.5 causes a 30-fold increase in cleavage rate for *Sel*RNAP and a 50-fold increase for *Ec*RNAP. However, while the rate for *Ec*RNAP continues to rise with pH, the *Sel*RNAP hydrolysis rate plateaus shortly after. This would suggest the presence of a general base for *Sel*RNAP as water is activated much more readily. This appears to be a similar but weaker deprotonation mechanism to that of Gre-dependent hydrolysis which reduces water pKa substantially.

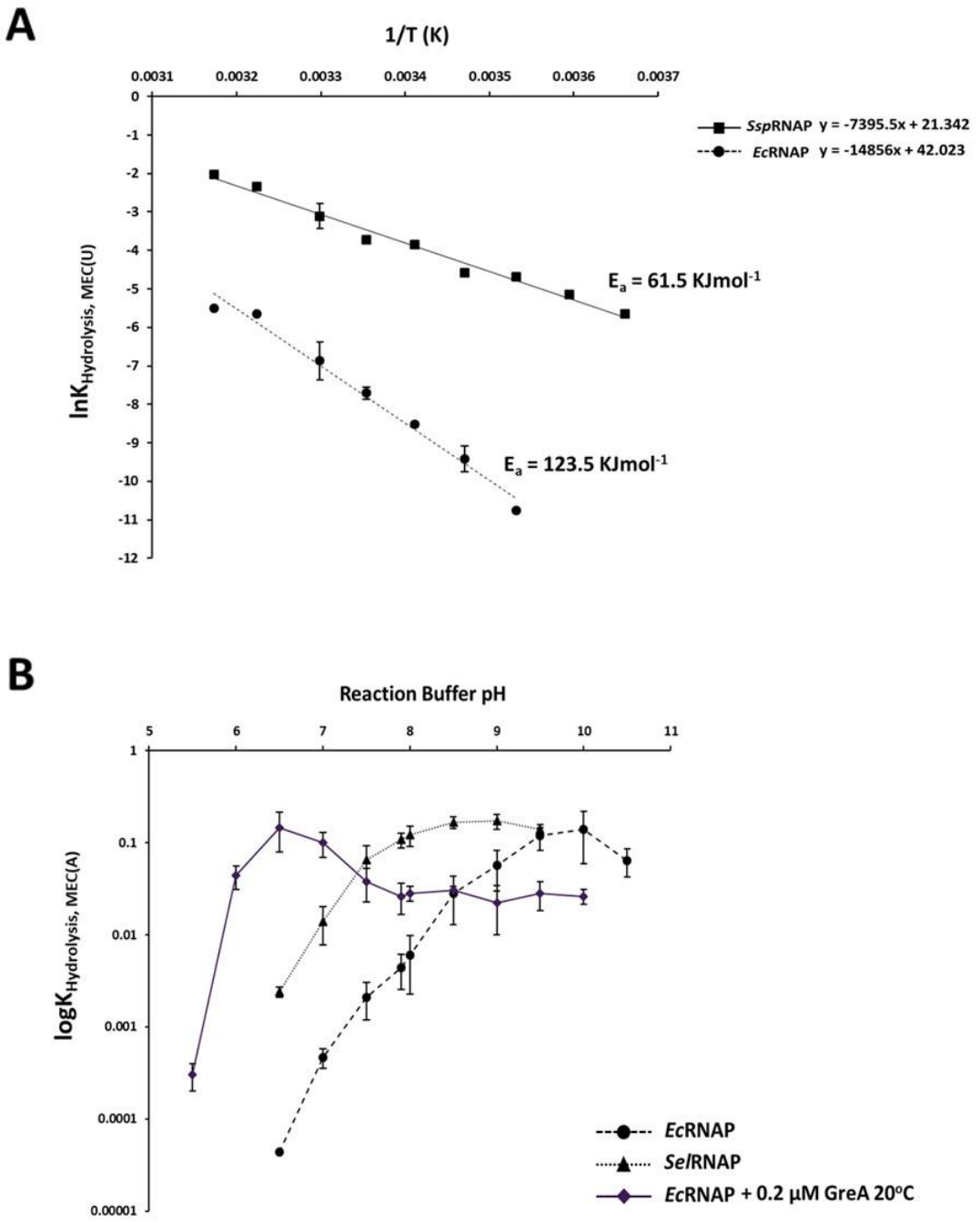


Figure 17. Cyanobacteria*l* RNAP Requires Less Energy for Hydrolysis and May be Aided by a General Base. (A) Arrhenius plot showing rate of hydrolysis of MEC(U) by *Ec*RNAP and *Ssp*RNAP against reaction temperature. Calculated value for Ea is placed next to the appropriate slope. **(B)** Rate of MEC(A) hydrolysis against reaction buffer pH titration for *Ec*RNAP, *Sel*RNAP, and *Ec*RNAP+0.2 µM GreA (purple). The values for each RNAP are the mean ± standard deviations of at least three independent experiments.

3.2.4 Contribution of Transcript-assisted Cleavage is Greater for Cyanobacterial RNAP

In addition to intrinsic hydrolytic activity mediated by either the TL or Gre, RNAPs also possess a ribozyme-like proof-reading activity where the 3' end of the transcript assists in excision (Zenkin *et al*., 2006). Upon mis-incorporating an incorrect NTP, the RNAP enters a thermodynamically favourable 1 bp backtracked state with the penultimate phosphodiester bond of the transcript placed in the active site. The erroneous RNA 3' end is flipped out and co-ordinates hydrolysis of this penultimate bond (**Figure 18A**).

In the same paper, Zenkin *et al*. (2006) assembled mis-incorporated complexes using modified bases and tested the rates of hydrolysis in order to determine the strength of the ribozyme-like mechanism. It was thought that a greater reduction in rate would correlate to a stronger original mechanism, as the modified base would disrupt the bonds required to co-ordinate the reacting groups for catalysis. These experiments were reproduced in this thesis using the MEC(A) complex so as to determine the contribution of transcript-assisted cleavage in cyanobacterial RNAP. The RNA 3' NMP of these complexes was either adenine or a modified base of purine, pyridine-2-one, 7-deaza-A, or 7-deaza-G, whereas the cognate base would be UMP. These complexes were therefore already in the 1 bp backtracked state when assembled. Hydrolysis rates for *Ec*RNAP, *Ssp*RNAP, and *Sel*RNAP on the modified complexes were measured against rates for MEC(A) and percentage residual hydrolytic activity is reported in **Figure 18B**. Rates were consistently lower for both cyanobacterial RNAPs, particularly with 7-deaza-A. This would indicate that flipped out 3' end of the nascent RNA transcript may be held in a different configuration or may be better able to co-ordinate penultimate bond cleavage in a stronger ribozyme-like mechanism. A greater contribution of transcript-assisted cleavage could help account for the overall increased intrinsic cleavage demonstrated by cyanobacterial RNAPs.

The mutant *Ec*RNAPs showing the most increased cleavage ability were also tested (**Figure 18C**). These mutants displayed a much greater reduction in hydrolytic rate, even more so than cyanobacterial RNAP for some bases. Transcript cleavage was consistently most reduced for β' F773V, and could not even be determined for 7-deaza-G. The decreased flexibility of the active site proposed for these mutant RNAPs, and thus for cyanobacterial RNAP, appears to allow better co-ordination of the backtracked RNA 3' in order to perform faster transcript catalysis.

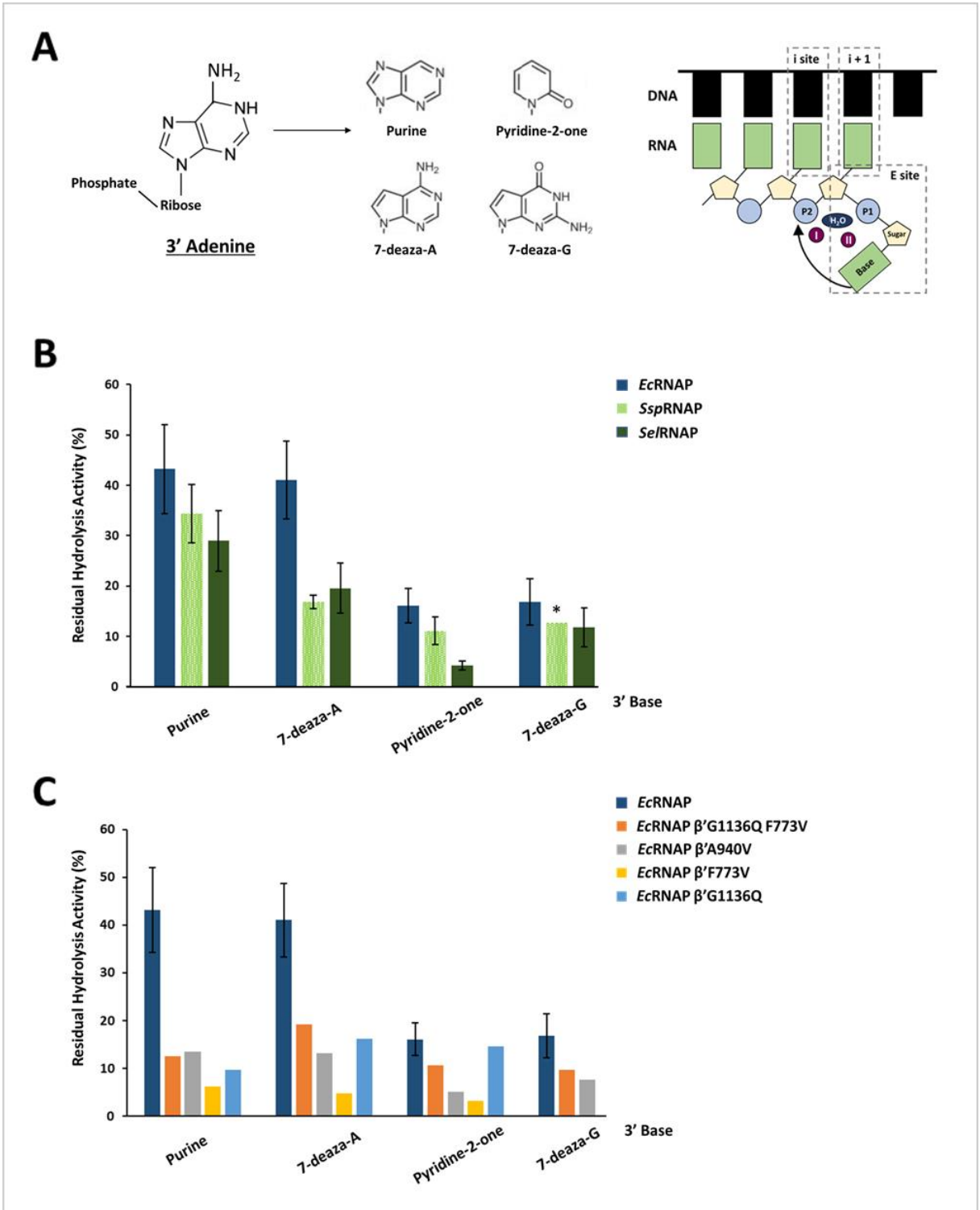


Figure 18. Contribution of Transcript-assisted Cleavage. A) The modified 3'-terminal nucleotides used in the assays. *Right,* schematic representation of the flipped-out RNA 3' end which co- ordinated hydrolysis of the penultimate phosphodiester bond, P2. Mg^{2+} ions shown in purple. Adapted from Zenkin *et al*. (2006). **B)** Residual hydrolytic activity of the modified bases against the rates for hydrolysis of MEC(A) for *Ec*RNAP, *Ssp*RNAP, and *Sel*RNAP. **C)** Residual hydrolytic activity of the modified bases against the rates for hydrolysis of MEC(A) for mutant *Ec*RNAPs. The values for each RNAP are the mean ± standard deviations of at least three independent experiments. * The rate for *Ssp*RNAP on 7-deaza-G was only calculated once.

3.2.5 Discussion

Differences between cyanobacterial and *Ec*RNAP in metal ion co-ordination and acid/base contributions to catalysis were assessed. In RNA hydrolysis, it is the Mg^{2+}II ion which activates the attacking water molecule in order to cleave the phosphodiester bond. The binding affinity for this ion in the active site can therefore impact catalytic activity. Rates of RNA hydrolysis for *Ec*RNAP and *Ssp*RNAP were measured for MEC(A) and MEC(U) in the presence of different buffer concentrations of Mg^{2+} in the form of $MgCl_2$. These data were used to determine the K_m for Mg^{2+}II of both RNAPs (**Figure 16**). A similar K_m was obtained for each RNAP on both scaffolds, approximately 21 mM for MEC(U) and 4 mM for MEC(A). The lower K_m value for MEC(A) reflects the faster rates seen for this scaffold previously and indicates that this complex is better able to co-ordinate Mg^{2+}II for hydrolysis. As the K_m values were similar for cyanobacterial and *Ec*RNAP, they are considered to have the a roughly equal binding affinity for Mg^{2+}II and so the differences observed in cleavage rates are not due to metal ion binding.

The MEC(U) scaffold was also used to determine the activation energy required for hydrolysis for each RNAP by measuring the hydrolysis rate over a range of reaction temperatures (**Figure 17**). The activation energy, E_a, is the energy barrier which must be overcome before the reaction can take place. E_a can be measured using the Arrhenius equation which observes that at higher temperatures, the probability that two molecules will collide is higher and results in a higher kinetic energy, therefore lowering the activation energy requirement. Enzymes like RNAP can catalyse reactions by lowering the activation energy required. One way this is done is by positioning and orientating the reactants close together to facilitate catalysis. The E_a required for RNA hydrolysis was much lower for *Ssp*RNAP than for *Ec*RNAP, at 61.5 $KJmol^{-1}$ and 123.5 $KJmol^{-1}$ respectively. The lower E_a for *Ssp*RNAP could contribute to the faster cleavage activity demonstrated by this enzyme. It could also mean, in combination with the results from the previous section, that the possible more closed conformation of the active site is able to work as a better catalyst by positioning the reactants for increased catalysis through lowering the required reaction activation energy.

Assisted deprotonation of the attacking water molecule in hydrolysis by a general base in the active site of RNAP would contribute to increased cleavage rate. Yuzenkova and Zenkin (2010) have suggested TL residue H1242 of *Thermus aquaticus* (orthologous to H936 of *Ec*RNAP) to perform this role. The existence of a general base for cyanobacteria was assessed by measuring cleavage rate over a buffer pH range (**Figure 17**). High pH stimulates hydrolysis by increasing water ionisation, however this stimulation is expected to plateau as

the pH reaches the water pK_a value (~14). The presence of a general base would activate water much more readily and this plateau would occur at a lower pH. Elevated pH has also been shown to enhance Mg^{2+}II retention by breaking salt bridges between neighbouring β subunit residues and allowing these residues to assist in Mg^{2+}II co-ordination (Sosunov *et al*., 2003). As the K_m for Mg^{2+}II was determined above to be similar for cyanobacterial and *Ec*RNAP on this same MEC(A) complex, we assume that this increased Mg^{2+}II retention will affect all RNAPs equally and can be discounted when comparing hydrolysis at increased buffer pH. The rate curve observed for *Sel*RNAP does indicate the action of a general base. Cleavage rate for *Sel*RNAP plateaus at roughly pH 7.5 which is in the pK_a range for histidine, and similar to the results of Yuzenkova and Zenkin (2010) implying that the conserved H1242 (*Thermus* RNAP numbering) site may also participate as a general base in cyanobacteria. This appears to be in contrast to *Ec*RNAP which does not show any deprotonation assistance by a general base, thus this acid/base mechanism of catalysis may differ in contribution between species. In Gre-dependent cleavage, reside E44 is thought to simultaneously assist with Mg^{2+}II co-ordination and water activation thus facilitating fast catalysis (Sosunova *et al*., 2013). Gre-dependent water activation is thought to be two-fold; the general base of Gre factors is thought to both stabilise the attacking water via hydrogen bonding with the water hydrogen atom, and to accelerate water ionisation by lowering the water pK_a (Sosunova *et al*., 2013). The substantial effect this mechanism has on cleavage rate can also be seen in **Figure 17**.

As part of the intrinsic cleavage mechanism, the frayed RNA 3' end of backtracked complexes is able to support Mg^{2+}II co-ordination and contribute to hydrolysis in a ribozyme-like fashion, known as transcript-assisted cleavage (Zenkin *et al*., 2006). This hydrolysis was tested in **Figure 18**, using the base MEC(A) complex with modified 3' NMP bases. Rates of residual cleavage activity with these modified bases, as opposed to with MEC(A), was assessed for cyanobacterial, mutant *Ec*RNAPs, and WT E*c*RNAP. Both species of cyanobacterial RNAP gave reduced residual cleavage, which was also mimicked by the fast cleaving mutant *Ec*RNAPs. This is particularly true for rates with 3'7-deaza-A for the cyanobacterial enzymes, and the β'F773V *Ec*RNAP mutant which consistently gave the lowest residual activity rates. It is thought that a greater reduction in rate correlates to a stronger original ribozyme-like mechanism, as the modified bases would disrupt orientation of the reacting groups for catalysis. Therefore, the contribution of transcript-assisted cleavage through co-ordination of these groups is greater for cyanobacterial RNAP and the *Ec*RNAP mutants. This increased co-ordination could be due to the limited flexibility of the previously proposed stabilised closed active site conformation for cyanobacterial and mutant *Ec*RNAPs.

3.3 Cyanobacterial RNAP may Respond to Pause signals Less Strongly and have a Similar Tolerance for RNA Mis-incorporation

3.3.1 Introduction and Aims

Changes in elongation rate for mutant RNAPs are often accompanied with altered pause behaviour (Bar-Nahum *et al*., 2005; Toulokhonov *et al*., 2007). This has been shown in the literature *Ec*RNAP mutants of β'F773V, β'N792D, and β'G1136S described previously. These substitution mutants have demonstrated increases RNAP elongation rate with less discrimination of substrate NTP, slower translocation, pause-insensitivity, and inefficient termination (Bar-Nahum *et al*., 2005; Zhang *et al*., 2010; Malinen *et al*. 2014). As these substitutions are naturally occurring in cyanobacteria, it was thought that cyanobacterial RNAP may also display similar elongation traits. In this section, the elongation characteristics of cyanobacterial, *Ec*RNAP and mutant *Ec*RNAPs were analysed on several DNA templates encoding various types of pause signals in order to test pause-sensitivity. Termination efficiency and translocation response to specific DNA sequences were assessed. Amounts of mis-incorporation were tested *in vitro* by measuring the incorporation rates of non-cognate nucleotides and ratio of extension after a mis-incorporation event. Lastly, to examine any loss in proof-reading or transcript fidelity, RNAseq analysis was performed on extracted total RNA to determine nucleotide mis-incorporation rates *in vivo*.

3.3.2 Cyanobacterial RNAP Exhibits Reduced Pausing Intensity and Frequency

The elongation characteristics of all cyanobacterial and *Ec*RNAPs were first tested on promoter-independent 'template 1' and 'template 2' linear DNA templates. These templates have been described by Yuzenkova *et al*. (2013) to affect elongation in a sequence-dependent manner by promoting the formation of particular RNAP translocation states at specific nucleotide positions. Comparison of the pausing profiles for cyanobacterial and *Ec*RNAP are shown in **Figure 19**. On template 1, positions EC15 and EC20 are sites thought to induce pausing by encouraging the pre-translocated state, where the RNA 3' end occupies the i+1 site and blocks the incoming NTP, whereas position EC16 is thought to induce the -1 bp backtracked state (Yuzenkova *et al*. 2013). Both cyanobacterial RNAPs show slightly reduced pausing across the whole of this template when compared to *Ec*RNAP, especially at position EC23 which is an incredibly strong pause site for *Ec*RNAP. Exceptions are found at pre-translocated position EC15 at which *Sel*RNAP pauses the most, and at position EC22 which induces a pause for *Ssp*RNAP. Unfortunately, the translocation states promoted at sites EC22 and EC23 are unknown. **Figure 19B** shows comparisons on template 2 which encodes a

known pre-translocated inducing site at position EC16 and a post-translocated inducing site at position EC15 (Yuzenkova *et al*. 2013). From the band intensity profiles on the right, it can be seen that both cyanobacterial enzymes have considerably reduced pausing at all template positions. The translocation states induced by the sequences of these templates were determined using *Thermus aquaticus* RNAP but are assumed to be the same for cyanobacterial and *Ec*RNAP. These assumptions are tested later in this section by measuring rates of single nucleotide addition at these sites. The cyanobacterial RNAPs also appear to reach the end of the template faster than *Ec*RNAP which could be due to a faster incorporation rate or to reduced pausing and is also assessed in later sections.

The elongation characteristics of all mutant *Ec*RNAPs were then determined using template 1. Elongation of the most improved cleavage rate *Ec*RNAP mutants of β'F773V, β'A940V, β'G1136Q, and β'G1136Q F773V are shown in **Figure 20**. The band profiles at 5 seconds are displayed beneath the gel images and are relative to the WT *Ec*RNAP profile. Thus, data below this point indicate a decreased pause intensity for the mutants relative to WT at that position whereas data above indicates increased signal intensity at that position for the mutant *Ec*RNAP. As can be seen by this graph, these mutant RNAPs also displayed reduced pausing in addition to increased cleavage. Signal intensity was reduced at all known translocation state positions – 15, 16, 20 – and also at the very strong position EC23 pause. The higher signal intensities demonstrated after position EC23 are thought to be because WT *Ec*RNAP is usually paused significantly before this site and so not as much enzyme makes it to the end of the template as it does for the mutant *Ec*RNAPs. β'F773V is the only mutant to show a stronger pause than WT *Ec*RNAP, and this is at position EC22 which also causes a strong pause for *Ssp*RNAP **(Figure 20).**

This is followed by several figures showing representative 23% PAGE elongation profiles of all other tested *Ec*RNAP mutants. In **Figure 21A**, other mutant RNAPs which are considered pause-resistant are displayed. While in **Figure 21B**, mutants demonstrating increased pausing are shown. All *Ec*RNAP mutants judged to be roughly similar to WT *Ec*RNAP in elongation and pausing are shown in **Figure 22**. Elongation characteristics of the two mutants of the SI3, β'ΔSI3 and β'*Ssp*SI3, are depicted in **Figure 22B**. Deletion of the SI3 appears to reduce pausing whereas replacement of the *Ec*RNAP SI3 with that of *Ssp*RNAP significantly increases pausing.

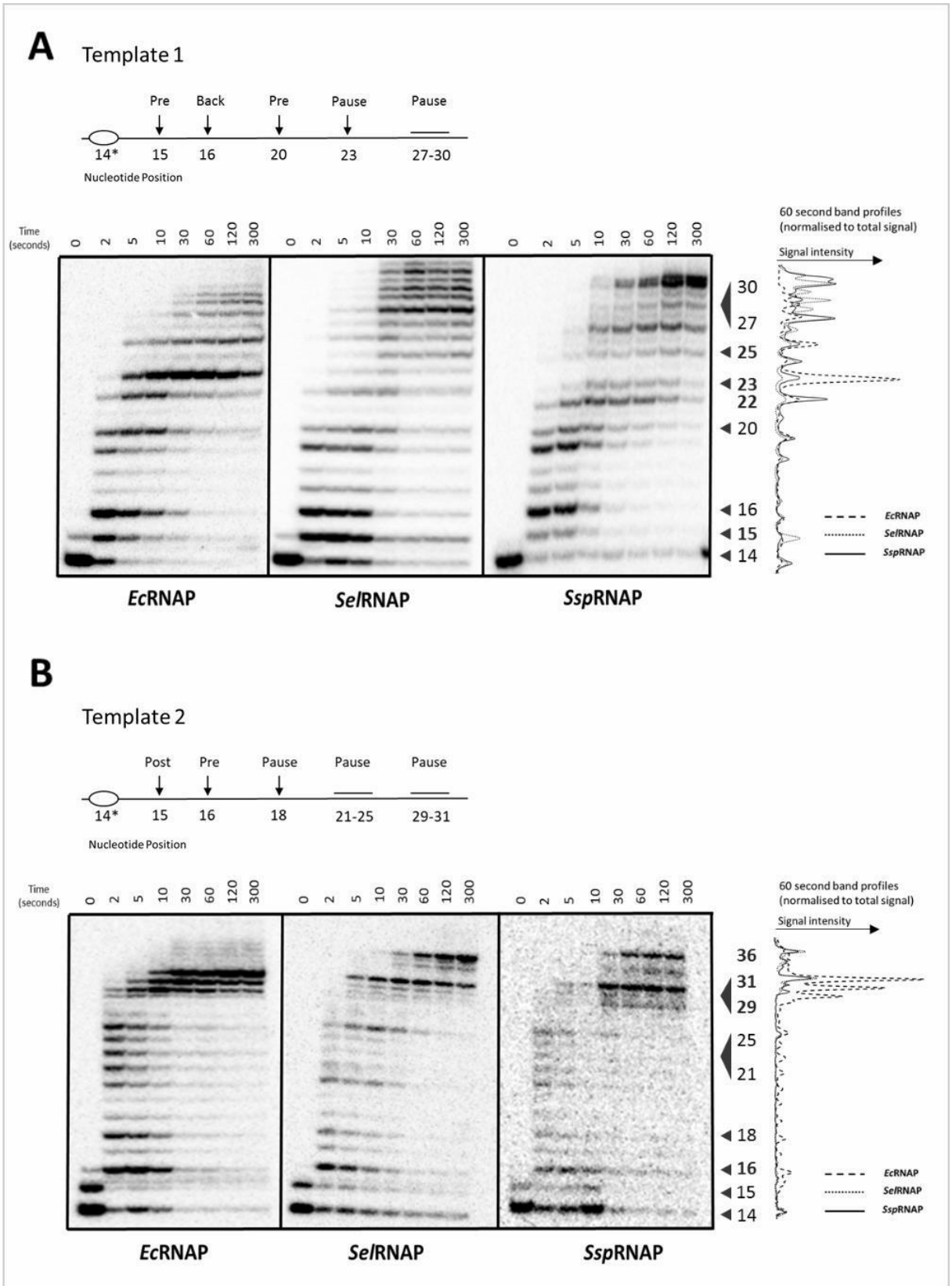


Figure 19. Cyanobacterial RNAP Exhibits a Different Pausing profile and Reduced Pause Intensity. (A) Scheme of short linear 'template 1' DNA template with pause sites indicated. Representative 23% PAGE of each RNAP elongating this template for the indicated time periods. Comparison of normalised 60 second band profiles for each RNAP showing differences in pause intensity at each position (*right*). (**B)** Scheme of short linear 'template 2' DNA template with pause sites indicated. Representative 23% PAGE of each RNAP elongating this template for the indicated time periods. Comparison of normalised 60 second band profiles for each RNAP showing differences in pause intensity at each position (*right*).

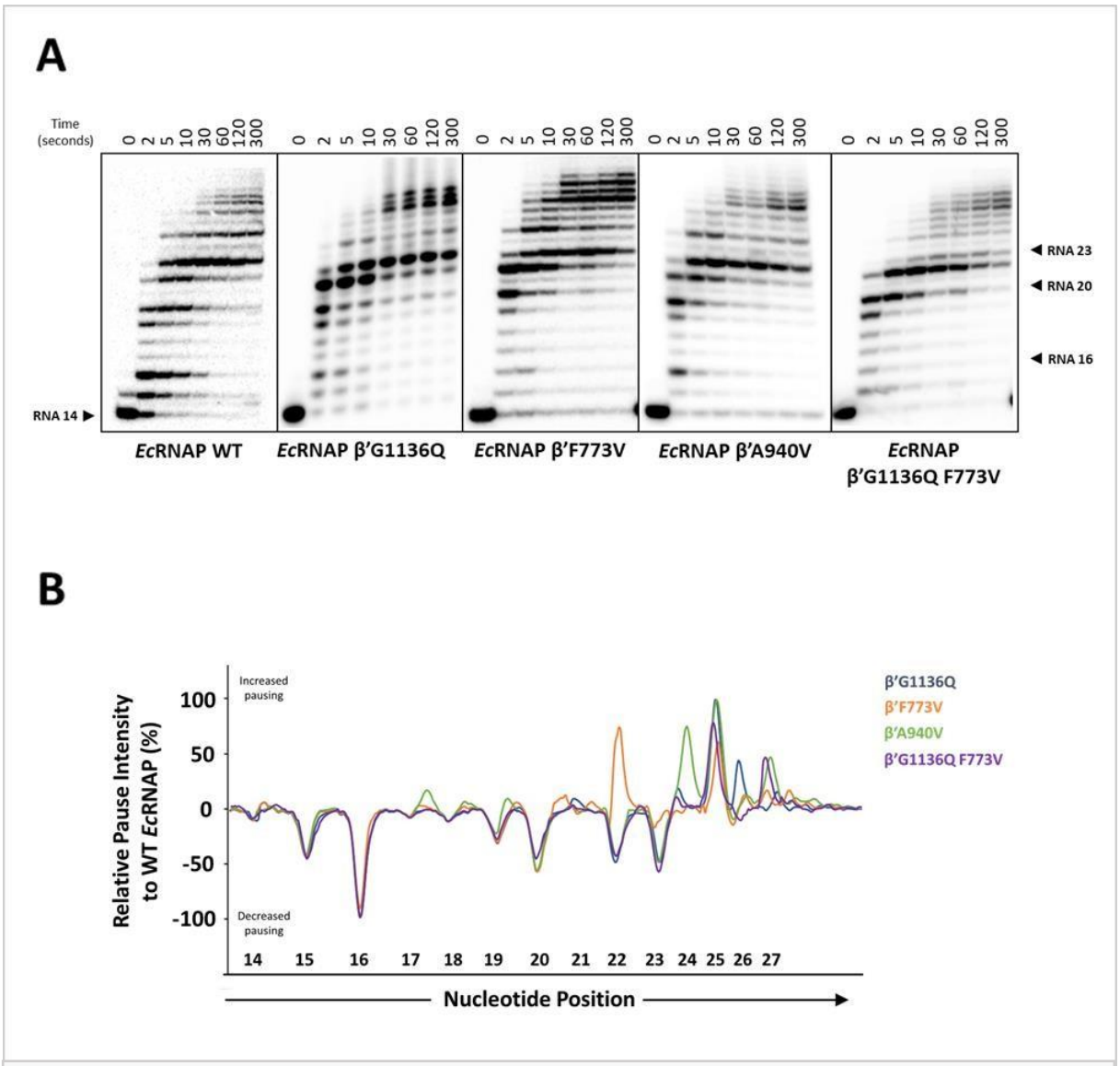


Figure 20. *Ec*RNAP Mutants with Increased Cleavage Also Show Reduced Pausing. (A) Representative 23% PAGE of each RNAP elongating template 1 for the indicated time periods. (**B)** Relative pause intensity across the template of each *Ec*RNAP mutant at 5 seconds compared against pause intensity for WT *Ec*RNAP.

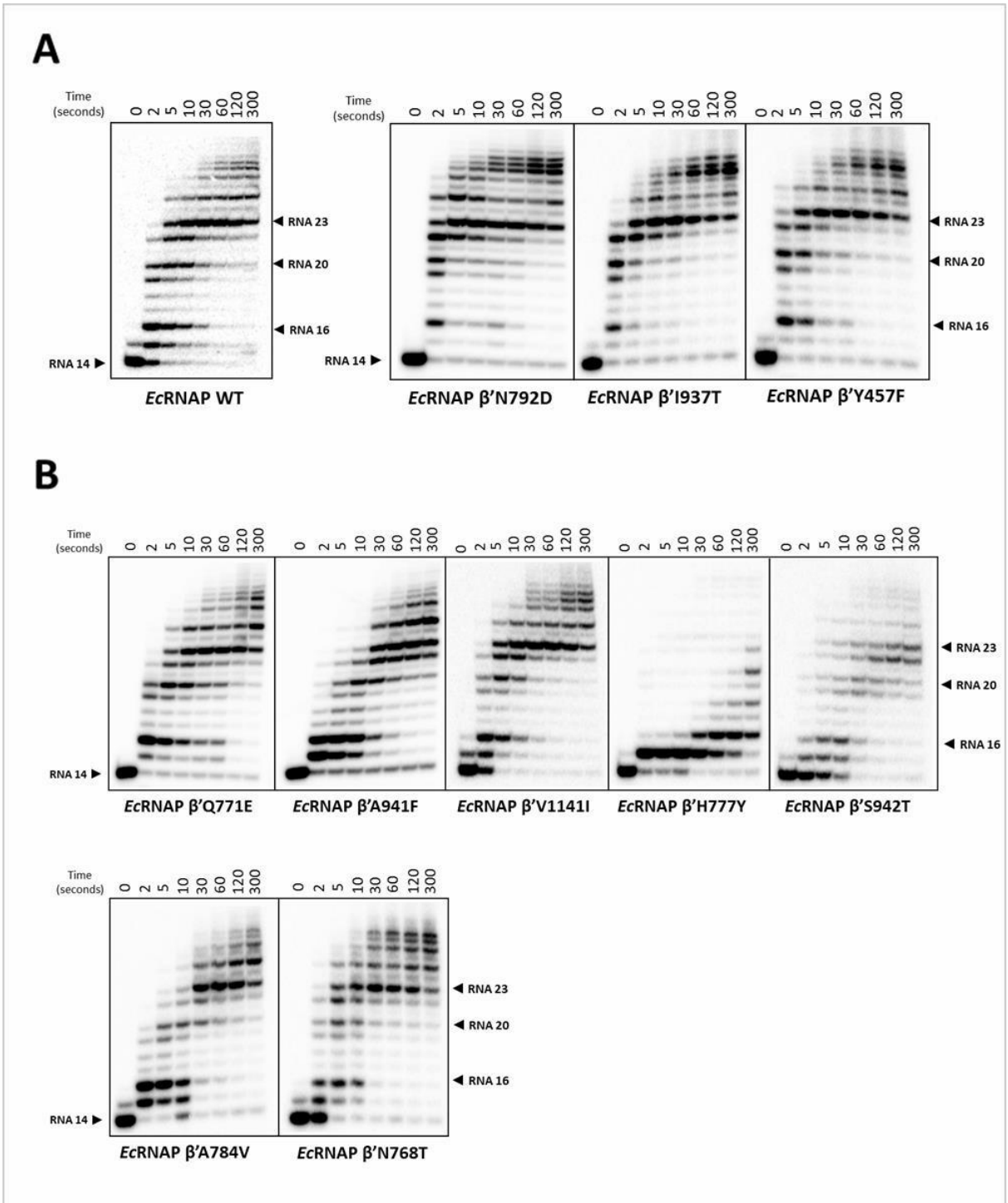


Figure 21. Elongation Profiles of *Ec*RNAP Mutants 1. (A) Representative 23% PAGE of mutant *Ec*RNAPs considered to have reduced pausing, elongating template 1 for the indicated time periods. (**B)** Representative 23% PAGE of mutant *Ec*RNAPs considered to have increased pausing, elongating template 1 for the indicated time periods.

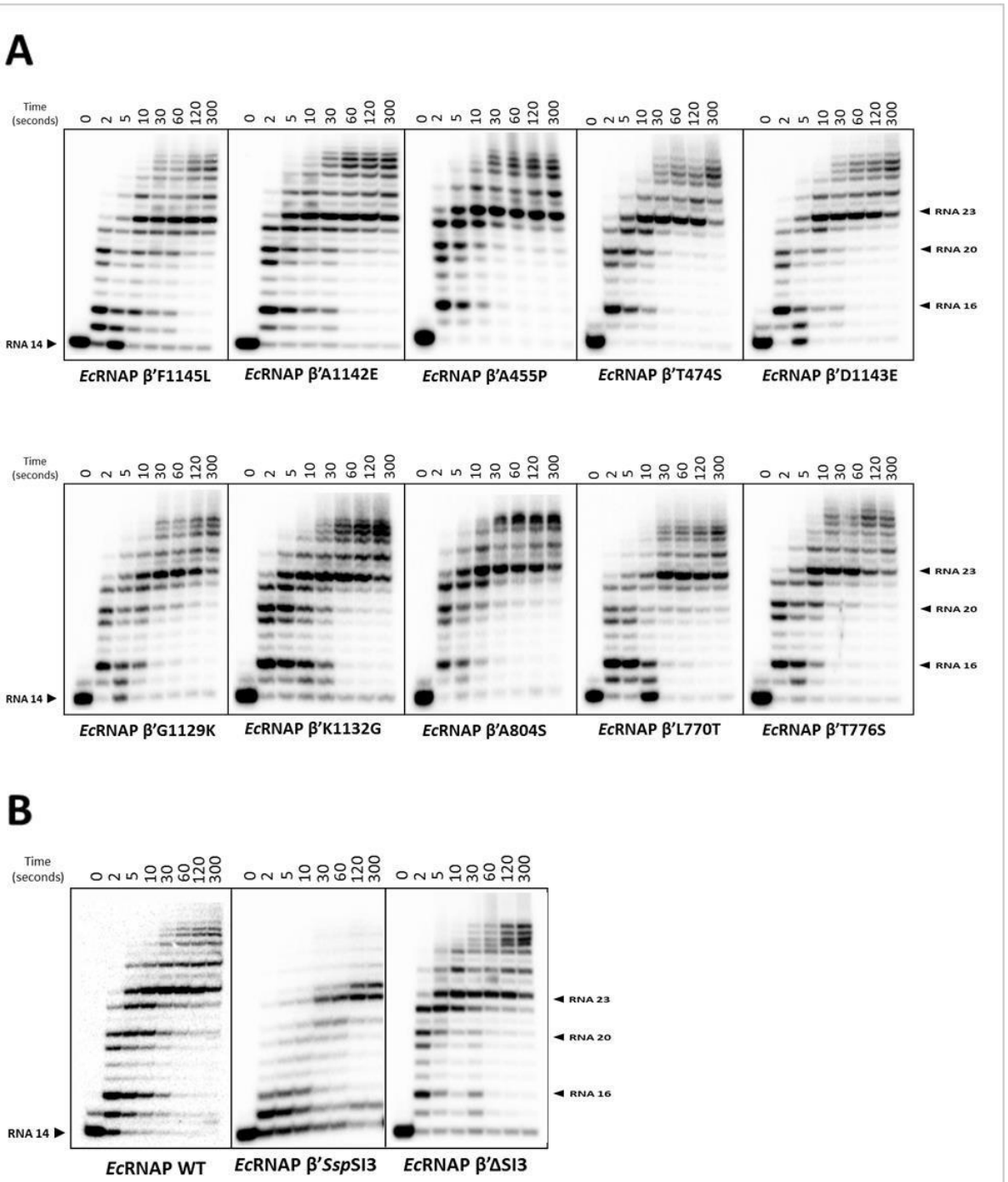


Figure 22. Elongation Profiles of *Ec*RNAP Mutants 2. (A) Representative 23% PAGE of mutant *Ec*RNAPs considered to have similar pausing to WT *Ec*RNAP, elongating template 1 for the indicated time periods. (**B)** Representative 23% PAGE of mutant SI3 *Ec*RNAPs elongating template 1 for the indicated time periods.

*Ssp*RNAP and *Ec*RNAP were also tested on the much longer promoter-dependent pIA349 linear DNA template (Artsimovich and Landick, 2002). This template requires RNAP to initiate at the T7A1 promoter and then elongate along a roughly 290 bp fragment in which a series of strong pause signals are encoded. This template allows the formation of a halted complex at position 37, before the pause sites, by withholding UTP from the reaction mixture. The pause signals encountered on this template include: the operon polarity suppressor (**ops**), sequence pause P1, and a hairpin-dependent *his* pause. The template ends with a Rho-independent hairpin forming *his* termination sequence.

Elongation of this template is shown in **Figure 23**. The band intensities at 60 seconds were normalised and compared and are to the right of each 10% PAGE. *Ssp*RNAP is again much less prone to pausing and appears to almost completely ignore the P1 and *his* pause sites. It does, however, pause strongly at the *ops* site. *Ec*RNAP mutants β'G1136Q and β'F773V were also tested and gave a similar reduced pausing profile to *Ssp*RNAP. This is particularly true for β'F773V which shows the reduced pausing and inefficient termination reported in the literature (Nedialkov *et al*., 2013). In **Figure 23C**, the sum total signal at the *ops* site, position or elongation complex (**EC**) 43-45, and the sum total signal for run-off at EC288 for both *Ssp*RNAP and *Ec*RNAP are plotted over time. In contrast to the other pause sites, both RNAPs show a roughly equal escape from *ops*, with perhaps only a slight increase in rate for *Ssp*RNAP. Elongation profiles for each RNAP on this template can be seen from the total signal curves for run-off. *Ssp*RNAP reaches maximum run-off value after about 50 seconds indicating that this is the amount of time for most *Ssp*RNAP complexes to have transcribed to this point. This curve is much shallower for *Ec*RNAP, demonstrating the increased time spent at each pause site and the increased terminator efficiency before the run-off position is reached.

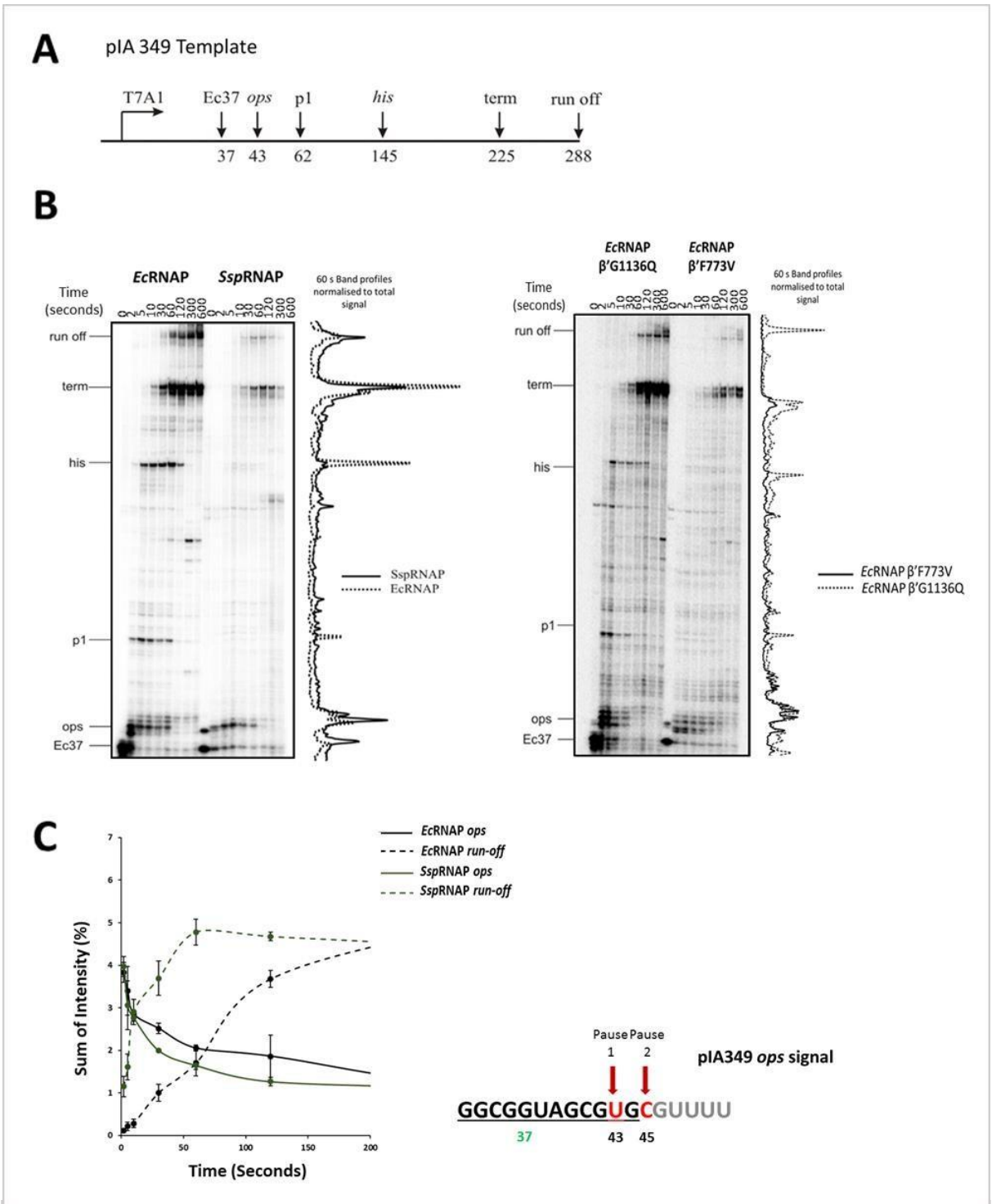


Figure 23. Elongation Profiles on the pIA349 Template (A) Scheme of long linear 'pIA349' DNA template with pause sites indicated. (**B)** Representative 10% PAGE of labelled RNAPs, elongating this template for the indicated time periods. 60 second band profiles are shown to the right. (**C)** Total signal at the *ops* and run-off sites for *Ssp*RNAP and *Ec*RNAP over time. Scheme of the *ops* pause is shown to the right. The values for each RNAP are the mean ± standard deviations of at least two independent experiments.

3.3.3 Cyanobacterial and EcRNAP Demonstrate a Similar Translocation Equilibrium

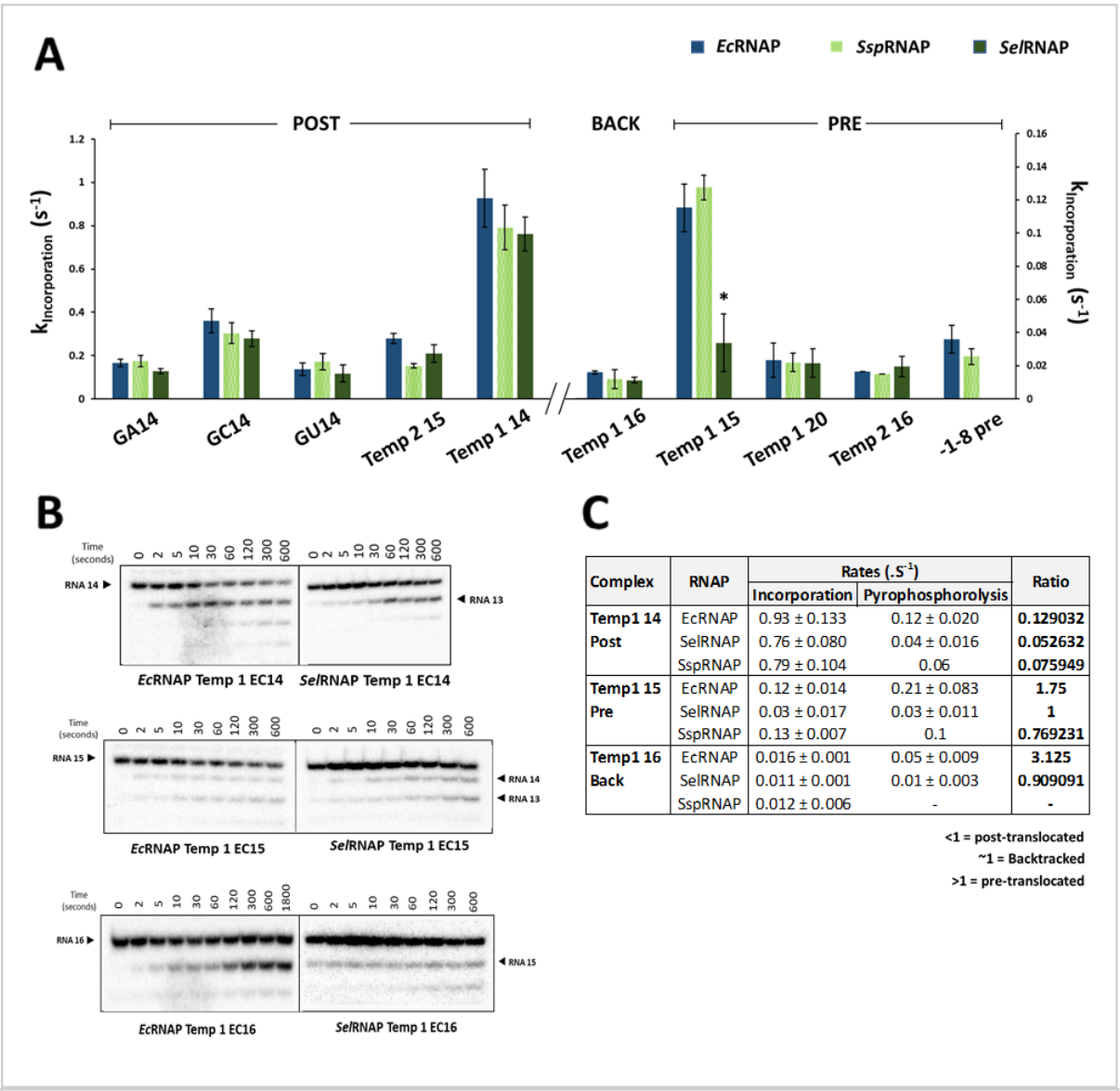
Translocation is thought to occur via a Brownian molecular ratchet mechanism where the RNAP oscillates between the pre-, post-, and backtracked translocation states. The equilibrium of these states is affected greatly by the sequence being transcribed. Yuzenkova *et al*. (2013) determined the sequence-dependent translocation states for template 1 and template 2 DNA templates using *Thermus aquaticus* RNAP. Translocation state at a specific site is resolved by measuring the rates of nucleotide incorporation, pyrophosphorolysis, and phosphodiester bond hydrolysis to determine the relative stabilization of elongation complex in each particular translocation state (Bochkareva *et al*., 2012). Incorporation, pyrophosphorolysis, and hydrolysis occur from the post-, pre-, and backtracked states respectively, therefore the rates of these activities greatly reflect the stability of the elongation complex in a particular state. Due to time constraints, only the rates nucleotide incorporation and pyrophosphorolysis could be measured to give an indication of translocation state.

Figure 24A displays the rates of single cognate nucleotide incorporation for both cyanobacterial RNAPs and *Ec*RNAP for several ECs of known translocation state. These complexes include the known pause positions on template 1 and template 2 described earlier, and five other complexes known to stabilise a particular translocation state (Bochkareva *et al*., 2012; Yuzenkova *et al*., 2013). These are -1-8 which induces a pre-translocated state, GA14, GC14, GU14, and the initial position EC14 on template 1 which all display the post-translocated state. Cyanobacterial and *Ec*RNAP appear to behave similarly and give comparable incorporation rates from each translocation state. Importantly, the rates from the post-translocated state are highest, and are lowest for incorporation from the backtracked state. One anomaly is the rate for *Sel*RNAP from template 1 EC15, as it is much lower than for the other two enzymes. However, this correlates with the elongation profile from **Figure 19**, where *Sel*RNAP demonstrates a pause at EC15 on template 1.

To measure pyrophosphorolysis, 250 µM pyrophosphate (**PP_i**) was added to the reaction mixture of complexes stabilised on template 1 at ECs 14, 15, or 16. The ratio of pyrophosphorolysis rate divided by incorporation rates gives an indication of translocation state (**Figure 24C**). If a RNAP is in the post-translocated state then incorporation is likely to greatly exceed pyrophosphorolysis and give a ration value <1, whereas if the RNAP is in a pre-translocated state where the i+1 site is blocked for incorporation then pyrophosphorolysis is much more likely and the ration value should be >1. The backtracked state is taken as being roughly equal for each reaction, although in actuality the pre-translocated state is reached

before the post- when an RNAP leaves the backtracked state so might give a rate advantage to pyrophosphorolysis. Only one of the ratios calculated is suspect, *Ec*RNAP Long 16 with a value of 3.125 where a value of roughly 1 would be expected. The pyrophosphorolysis assays have only two repeats where an error sign is present so further repeats to exclude potential manual error would be beneficial, as would further analyses of these complexes by the third type of reaction, RNA hydrolysis.

Overall these data indicate that cyanobacterial RNAP, *Thermus aquaticus* RNAP, and *Ec*RNAP are stabilised in the same sequence-dependent translocation states, showing an equal translocation equilibrium.



Complex	RNAP	Rates ($.s^{-1}$)		Ratio
		Incorporation	Pyrophosphorolysis	
Temp1 14	EcRNAP	0.93 ± 0.133	0.12 ± 0.020	**0.129032**
Post	SelRNAP	0.76 ± 0.080	0.04 ± 0.016	**0.052632**
	SspRNAP	0.79 ± 0.104	0.06	**0.075949**
Temp1 15	EcRNAP	0.12 ± 0.014	0.21 ± 0.083	**1.75**
Pre	SelRNAP	0.03 ± 0.017	0.03 ± 0.011	**1**
	SspRNAP	0.13 ± 0.007	0.1	**0.769231**
Temp1 16	EcRNAP	0.016 ± 0.001	0.05 ± 0.009	**3.125**
Back	SelRNAP	0.011 ± 0.001	0.01 ± 0.003	**0.909091**
	SspRNAP	0.012 ± 0.006	-	**-**

<1 = post-translocated
~1 = Backtracked
>1 = pre-translocated

Figure 24. Cyanobacterial and *Ec*RNAP are stabilised in the same sequence-dependent translocation states (A) Single cognate nucleotide incorporation rates for each RNAP from complexes known to stabilise particular translocation states. (**B)** Representative 23% PAGE of pyrophosphorolysis at three positions of template 1 for *Ec*RNAP and *Sel*RNAP. **(C)** Table of calculated rates for incorporation and pyrophosphorolysis at three positions of template 1 for all RNAPs. Ratio indicating translocation state is calculated from the pyrophosphorolysis rate divided by the incorporation rate. The values for each RNAP are the mean ± standard deviations of at least two independent experiments.

3.3.4 Cyanobacterial and EcRNAP React Similarly to Non-cognate NTP

Elongation Complexes were assembled to challenge each RNAP to perform a specific type of mis-incorporation. These complexes were incubated with 1 mM non-cognate NTP and the rates of initial complex presence were measured over time (**Figure 25A**). This rate includes both addition and hydrolysis. Measurement of this aspect of the assay was chosen as it allowed comparison across all RNAPs tested because, particularly for the cyanobacterial RNAPs, cleavage was much too fast and incorporation could not be measured. Therefore, it is not a true test of mis-incorporation ability but rather measures reaction when presented with only non-cognate NTP. Overall, these reactions appear broadly similar for both species but a better experimental design is required to determine actual mis-incorporation rates.

3.3.5 Cyanobacteria show largely similar RNA Mis-Incorporation in vivo

To examine any loss in proof-reading or transcript fidelity, RNAseq analysis was performed on extracted total RNA from exponential cultures of *E. coli* and *S.sp* to determine nucleotide mis-incorporation rates *in vivo*. These RNA transcripts were compared to the appropriate reference genomes to identify any mis-matched nucleotides. Specific error rates for that type of mis-match were calculated as a percentage of total complexes encoding a mis-matched nucleotide. This analysis was carried out by Dr. Katherine James and the results are displayed in **Figure 25B**. Types of mis-incorporation are labelled in the way of C>G which denotes a GMP has been incorporated where a CMP should be. The transcripts of *S.sp* appear to have slightly higher rates than those of *E. coli* for the mis-incorporations of C>A, C>U, A>C, A>G, and A>U. This difference is quite minimal, and we would like to repeat the assay using the other cyanobacterial species as a control. As the rates of the other types of mis-incorporation seem to be similar for both species we believe that rates of mis-incorporation errors *in vivo* are generally similar.

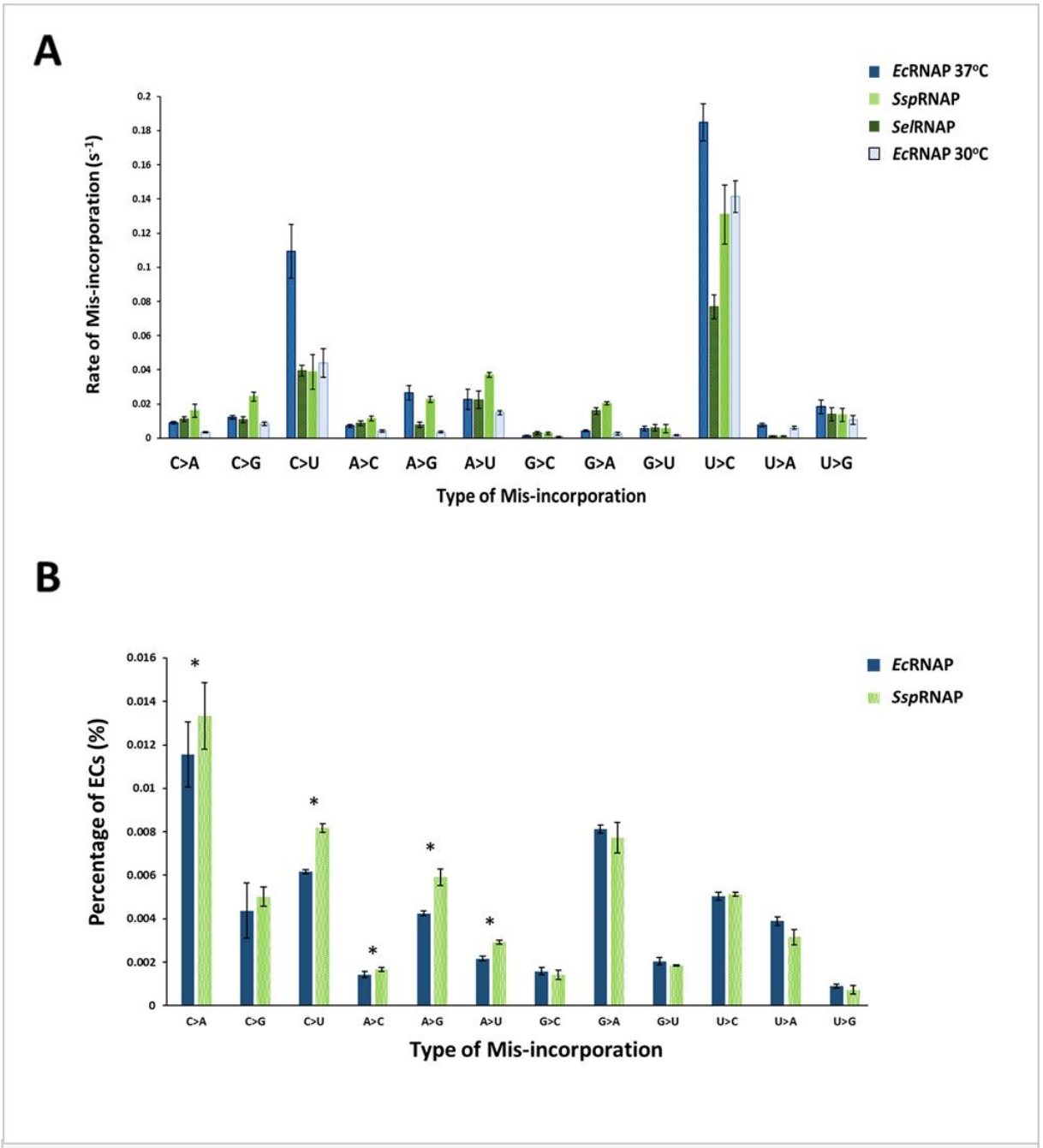


Figure 25. A. Rates of Mis-incorporation *in vitro* are Similar B. Percentage of Mis-incorporated ECs from Extracted Total RNA (A) Rates of initial complex disappearance when exposed to an incorrect nucleotide. The values for each RNAP are the mean ± standard deviations of at least three independent experiments. **(B)** RNAseq analysis to show specific error rates for mis-matches were calculated as a percentage of transcripts encoding a mis-matched nucleotide. Types of mis- incorporation are labelled in the way of C>G which denotes a GMP has been incorporated where a CMP should be. * denotes a higher error rate for that type of mis-incorporation in *S.sp.*

3.3.6 Discussion

Transcription elongation is a discontinuous process of repeating cycles of nucleotide addition, interspersed with sequence-induced entry into a paused state (Herbert *et al*., 2006). Pauses can be divided into three main classes: 1) Elemental pauses caused by the DNA:RNA sequence and involve RNA conformational changes of the BH, TL, and clamp to inhibit nucleotide addition without backtracking (Kireeva and Kashlev, 2009); 2) Backtracked-stabilised pauses where the RNAP reverses on a template and disengages the RNA 3'-OH from the active site and into the secondary channel (Nudler, 2009); 3) Hairpin-stabilised pauses where the nascent RNA transcript forms a hairpin secondary structure which blocks the RNA exit channel (Artsimovitch and Landick, 2000). The last two types of pausing are stabilised into longer- lived pauses from the initial elemental pause. Pausing performs several regulatory roles in co- ordinating co-transcriptional processes including RNA folding, coupled translation, and allows the recruitment of regulatory factors (Zhang and Landick, 2016).

Altered pause behaviour has often been seen in RNAPs encoding mutations of the mobile TL and BH domains (Bar-Nahum *et al*., 2005; Toulokhonov *et al*., 2007). Previous studies of *Ec*RNAP mutants β'F773V, β'N792D, and β'G1136S have demonstrated an increased elongation rate, resistance to pause signals, and inefficient termination (Bar-Nahum *et al*., 2005; Zhang *et al*., 2010; Malinen *et al*. 2014). As these substitutions naturally occur in cyanobacteria, it was thought that cyanobacterial RNAP may also display similar elongation behaviour in addition to the increased cleavage rates determined earlier. This was assessed first on two templates (**Figure 19**) which encoded sequence-dependent pauses to induce the pre-, post-, or backtracked translocation state, and would have to be overcome by the polymerase in order to elongate the whole template. Both cyanobacterial RNAPs demonstrate reduced pausing on each template at sites causing strong pauses for *Ec*RNAP. However, *Sel*RNAP had a stronger pause at EC15, a pre-translocated site, on template 1 which was consistently seen in repeated elongation assays. These templates were originally described for *Thermus aquaticus* RNAP (Yuzenkova *et al*., 2013). To assess whether the same DNA sequences promote the formation of a particular translocation state in both cyanobacterial and *Ec*RNAP, single nucleotide incorporations were performed at each of the proposed pause sites. These data are shown in **Figure 24**, where similar rates were obtained for each RNAP species indicating that they respond comparably to sequence-dependent pauses. The chart displays the high rates expected when incorporating from a post-translocated state, and the lower rates expected when incorporating a nucleotide from the pre- and backtracked states (as translocation to the post-translocated state would have to first occur before the addition could be performed). Again, *Sel*RNAP was an exception with a lower rate for incorporating from

EC15 on template 1. These data suggest that this sequence position might denote either a very strong pre-translocated state or a backtracked state for *Sel*RNAP.

Elongation of the fastest cleaving *Ec*RNAP mutants of β'F773V, β'A940V, β'G1136Q, and β'G1136Q F773V on template 1 are shown in **Figure 20**. The band profiles at 5 seconds were assessed to determine the effect of the substitution on the initial pauses at the beginning of the template and are relative to WT *Ec*RNAP. Data below 0% indicate a decreased intensity for the mutants relative to WT at that position whereas data above indicates increased signal intensity at that position for the mutant *Ec*RNAP. Signal intensity was reduced at all known translocation state positions. The higher signal intensities demonstrated after position 23 are thought to be because WT *Ec*RNAP is usually paused significantly at this site and so not as much enzyme makes it to the end of the template as it does for the mutant *Ec*RNAPs. Overall, the reduced pausing of the mutants agrees with previous studies. However, a stronger pause than WT *Ec*RNAP occurs at position 22 for β'F773V which has an untested translocation state. As this position also causes a strong pause for *Ssp*RNAP, it may be specific for this β'F773V substitution as *Sel*RNAP encodes isoleucine at this site and does not demonstrate this pause. Substitutions of β'N792D, I937T, and Y457F were subjectively determined to also show less pausing, seven mutants had increased pausing, and the rest were similar to WT *Ec*RNAP. β'A941F, β'H777Y, and β'S942T were substantially more pause-sensitive. β'H777 is a BH residue adjacent to the first glycine hinge and is proposed to contact another active site element, the F loop, which can affect catalytic activities by modulating TL conformations (Miropolskaya *et al*., 2014). This site may also be responsible for BH conformational changes as proline substitution of this specific residue has been described to increase BH kinking and result in enhanced catalytic activity (Weinzierl, 2010; Tan *et al*., 2008). The two SI3 mutants, β'ΔSI3 and β'*Ssp*SI3, were also tested on this template as β'SI3 mutants of *Ec*RNAP have been shown to reduce pausing during elongation, and particularly to increase the rate of escape from hairpin-dependent pause signals (Artsimovitch *et al*., 2003; Ray-Soni *et al*., 2017). β'ΔSI3 did demonstrate perhaps slightly reduced pausing whereas β'*Ssp*SI3, with the expanded SI3 of *Ssp*RNAP, increased pause sensitivity for *Ec*RNAP (**Figure 22B**). The slower elongation and decreased cleavage seen in the previous results section would suggest that the larger SI3 reduces or inhibits TL folding, resulting in decreased activity. Although, it cannot be said that this would be the case in cyanobacteria where the large SI3 is naturally found. The corresponding mutations should be made in *Ssp*RNAP in order to fully investigate SI3 effects on pausing.

Elongation of cyanobacterial and *Ec*RNAP was also tested on the longer pIA349 template, highlighted in **Figure 23** (Artsimovitch and Landick, 2002). This template encodes an *ops* pause site which is a member of the class II family of pause signals described by (Artsimovitch and Landick, 2000). In *E. coli*, *ops* signals induce a pre-translocated state and mediate recruitment of elongation factor RfaH which is a pausing and termination suppressor (Artsimovitch and Landick, 2000). Escape from *ops* appeared to be similar between *Ssp*RNAP and *Ec*RNAP which was somewhat unexpected, but as this pause is pre-translocated, the increased cleavage ability of *Ssp*RNAP would not be of any assistance in pause escape until the backtracked state is reached. *Bacillus subtilis* RNAP has also been shown to recognise the *ops* pause site, but with a lower half-life and efficiency which is thought to be due to having a different regulatory function as RfaH is absent in these species (Artsimovitch *et al*., 2000). RfaH also appears to be absent in *S.el* and *S.sp*, meaning that the pausing at this site seen in these species may also be for a different purpose. It has recently been added that the non-template DNA within the paused RNAP at these sites may form a hairpin structure for this RfaH recruitment (Zuber *et al*., 2018). Hairpin formation of the non-template DNA could be the reason for pausing demonstrated by the cyanobacterial RNAPs as escape from this would not be as dependent on cleavage ability.

This was followed by a hairpin-dependent *his* pause site which causes an RNA hairpin to form from the nascent RNA transcript and stabilises the elongation complex in a paused state. *Ec*RNAP strongly recognised these types of pauses whereas recognition by *Bacillus subtilis* RNAP and eukaryotic RNAPII is non-existent (Artsimovitch *et al*., 2000). These differences in response are thought to be due to the diverged regulatory needs of each species (Zhang and Landick, 2016). *Ssp*RNAP also ignores this *his* pause and displays reduced termination at the subsequent Rho-independent hairpin termination sequence. The increased effect of the terminator is likely due to destabilisation by the poly-U tract which follows the hairpin. The *Ec*RNAP β'F773V and β'G1136Q, which show strong *his* pause-resistance, were also able to be tested on this template and agreed with previous literature by showing reduced pausing and termination (Toulokhonov *et al*., 2007). Elongation with the *Ec*RNAP β'F773V β'G1136Q double substitution was attempted but failed to form the initial EC37, indicating that initiation from the promoter for this mutant may be deficient.

3.4 *T.aq* Gfh1 Preferentially Inhibits Activity from the Pre-translocated State

3.4.1 Introduction and Aims

An additional minor investigation was also carried out to analyse the inhibitory activity of *T.aq* Gfh1. Gfh1 is a member of the Gre family SCBFs but exhibits contrasting activities to GreA. Gfh1 of *Thermus thermophilus* has been shown to inhibit transcript cleavage and NTP addition whereas Gfh1 of *Deinococcus radiodurans* does not inhibit cleavage but does increase RNAP pausing and termination (Hogan *et al*., 2002; Symersky *et al*., 2006; Esyunina *et al*., 2016). *Deinococcus radiodurans* Gfh1 has been shown to act on elemental, hairpin-dependent, and backtracked pauses (Agapov *et al*., 2017). The difference in activity to that of GreA is thought to be due to four acidic residues on the tip of the Gfh1 coiled-coil which co-ordinate Mg^{2+}II such that a resultant stabilised active site of RNAP is in a catalytically inactive conformation (Laptenko *et al*., 2006). In this section, the templates used previously for elongation of cyanobacterial and *Ec*RNAP are used to determine if Gfh1 of *T.aq* exerts stronger inhibition on particular RNAP translocation states. These assays were performed at 40^{o}C, pH6.8, and predominantly in the presence of $MnCl_2$ as described in the methods. An inactive Gfh1 mutant was created to be used as a control. This mutant, Gfh1-DA4, had all four aspartates of the coiled-coil tip substituted with alanines and has been observed to be severely compromised (Laptenko *et al.*, 2006).

3.4.2 Gfh1 Inhibits RNAP Elongation

The elongation profiles of *T.aq*RNAP on Template 1 (**Figure 26**) and Template 2 (**Figure 27**) were determined when in the presence of 10 µM Gfh1, 10 µM Gfh1-DA4 (**DA4**), or an equivalent volume of storage buffer (**SB**). On both templates, Gfh1 induces strong pauses at multiple EC positions. The band profiles in sections C compare the signal intensities of each assay at the 60 second time point. The strongest inhibitory activity is seen at EC20 and EC22 on template 1, with additional small pauses at EC14 and EC18. On template 2, pausing is strongest at EC16-18 and again at EC22-23. Of these, EC20 on template 1 and EC16 on template 2 are known sites where RNAP is in a pre-translocated state (Yuzenkova *et al*., 2013). Profiles for *Taq*RNAP and *Taq*RNAP with DA4 or SB were mostly similar as expected, with only EC14 on template 1 producing a strong pause for DA4.

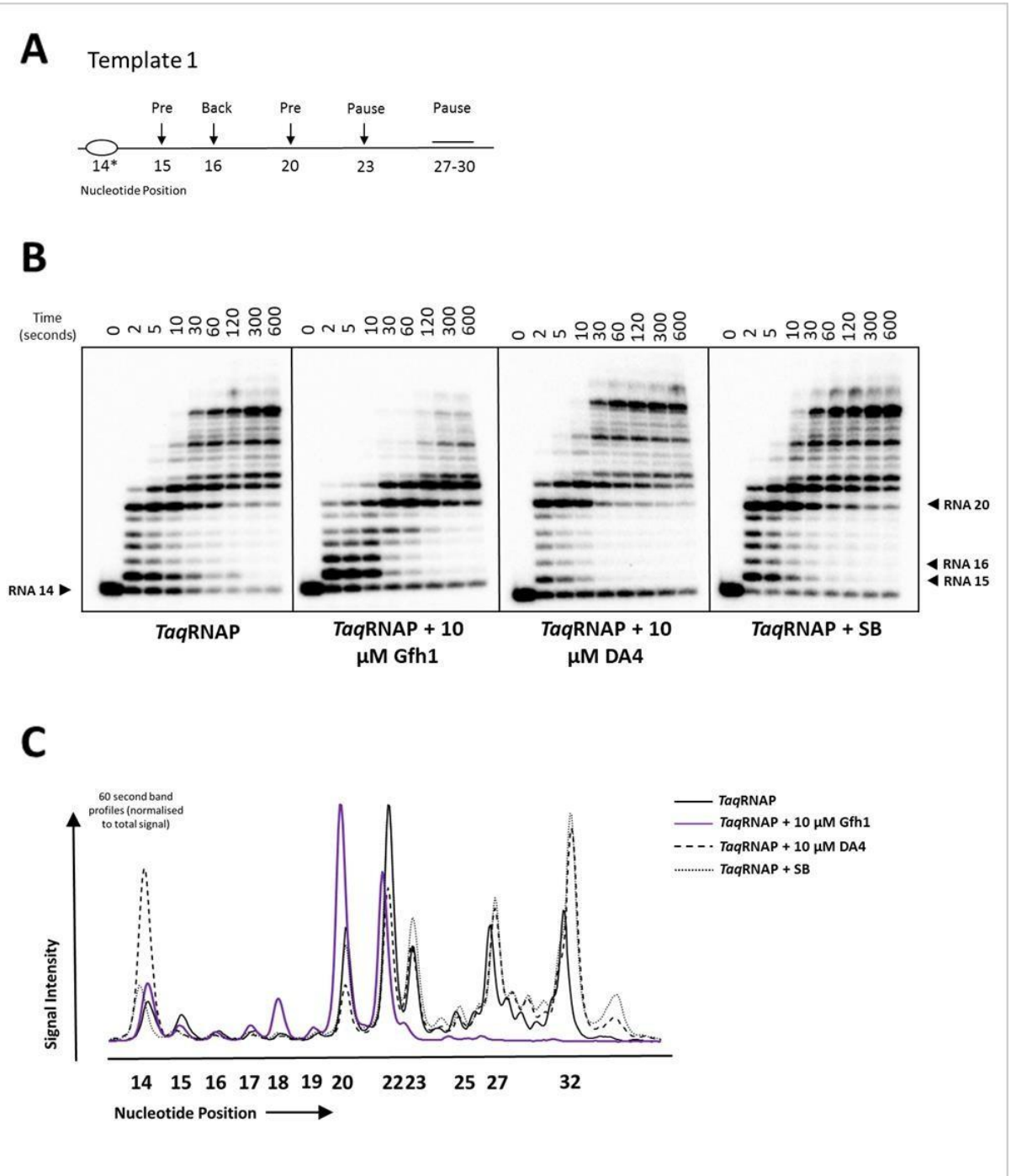


Figure 26. Elongation of Template 1 with *Taq*RNAP and Gfh1. (A) Scheme of short linear Template 1 with pause sites indicated. (**B)** Representative 23% PAGE of each RNAP elongating this template for the indicated time periods. (**C)** Comparison of normalised 60 second band profiles for each RNAP showing differences in pause intensity at each position.

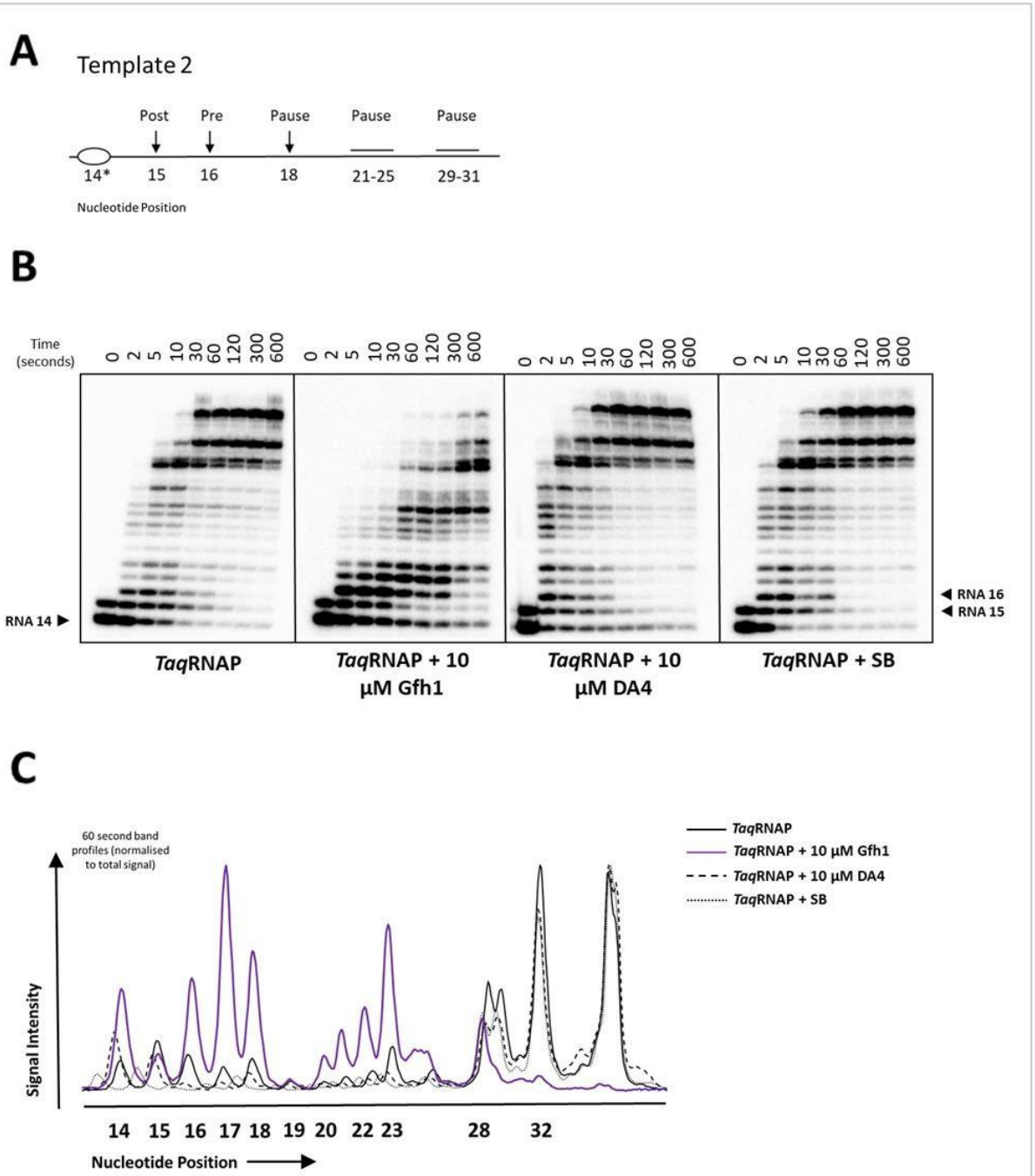


Figure 27. Elongation of Template 2 with *Taq*RNAP and Gfh1. (A) Scheme of short linear Template 2 with pause sites indicated. (**B)** Representative 23% PAGE of each RNAP elongating this template for the indicated time periods. (**C)** Comparison of normalised 60 second band profiles for each RNAP showing differences in pause intensity at each position.

3.4.3 Gfh1 Inhibits the Overall Rate of RNA Hydrolysis but Stimulates Exonucleolytic Activity

The inhibition of Gfh1 on transcript cleavage activity was measured using scaffolds EC(GA), MEC(A), and MEC(U) described previously (**Figure 28**). In the presence of Mg^{2+}, not much inhibition is seen on any of the complexes. However, in the presence of Mn^{2+}, Gfh1 significantly reduces the cleavage ability of *Taq*RNAP to roughly 25% of the WT *Taq*RNAP rate. This is similar to what has been reported for the Gfh1 activity of *Thermus thermophilus* and differs with the lack of inhibition seen in *Deinococcus radiodurans* (Hogan *et al*., 2002; Esyunina *et al*., 2016). An interesting effect of Gfh1 was observed in cleavage of the EC(GA) cognate complex (**Figure 29**). Gfh1 induces cleavage by 1 nucleotide in apparent stimulation of the exonuclease activity of RNAP where the 3' terminal RNA residue is cleaved. This is unusual as exonucleolytic activity is only performed when RNAP is in the pre-translocated state and these complexes would be assembled first in the post-translocated state, and then they would reverse to a backtracked state from a lack of nucleotides. The percentage of RNA14 product is also less than RNA13 in all time points, suggesting that this -1 nucleotide cleavage occurs either very quickly so as to produce both RNA14 and RNA13 cleavage products, or occurs at the same time as intrinsic endonucleolytic cleavage.

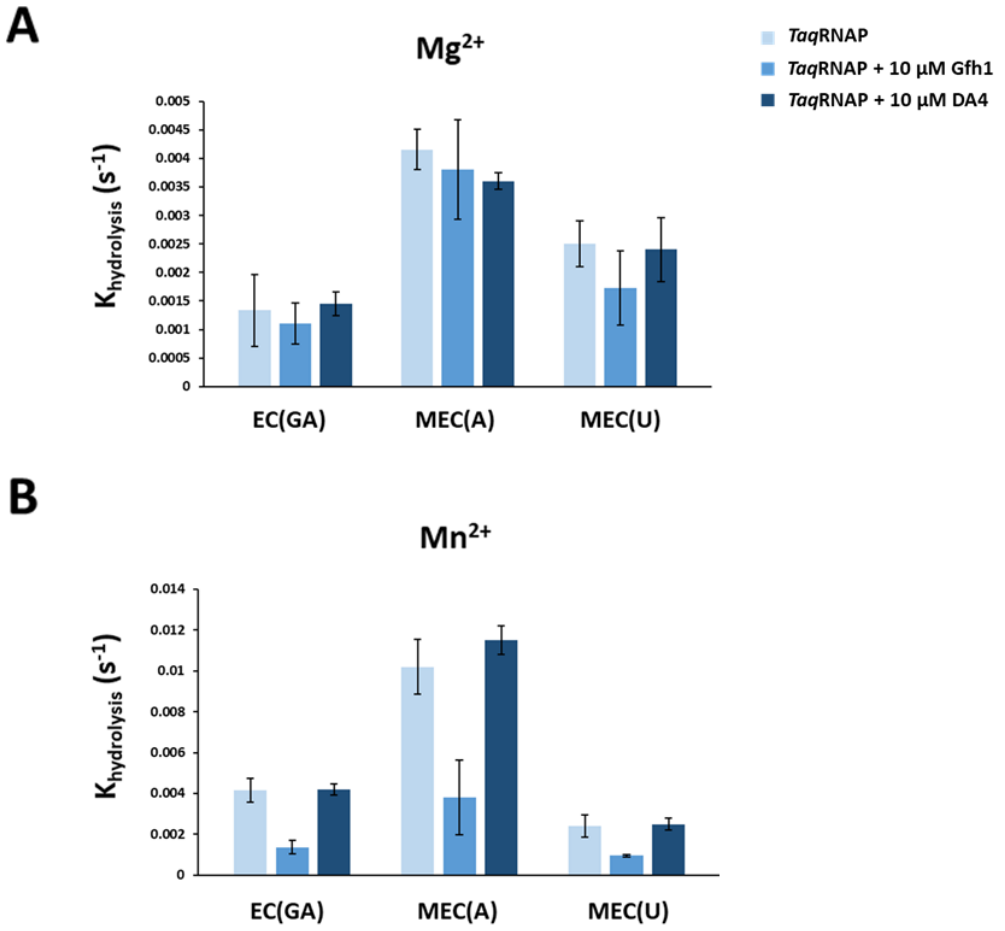


Figure 28. Gfh1 inhibits Transcript Cleavage in the presence of Mn^{2+} (A) Rates of transcript cleavage on EC(GA), MEC(U), and MEC(A) incubated with Mg^{2+} **(B)** Rates of transcript cleavage on EC(GA), MEC(U), and MEC(A) incubated with Mn^{2+}. The values for each RNAP are the mean ± standard deviations of three independent experiments. Assays performed at 40°C, pH 6.8.

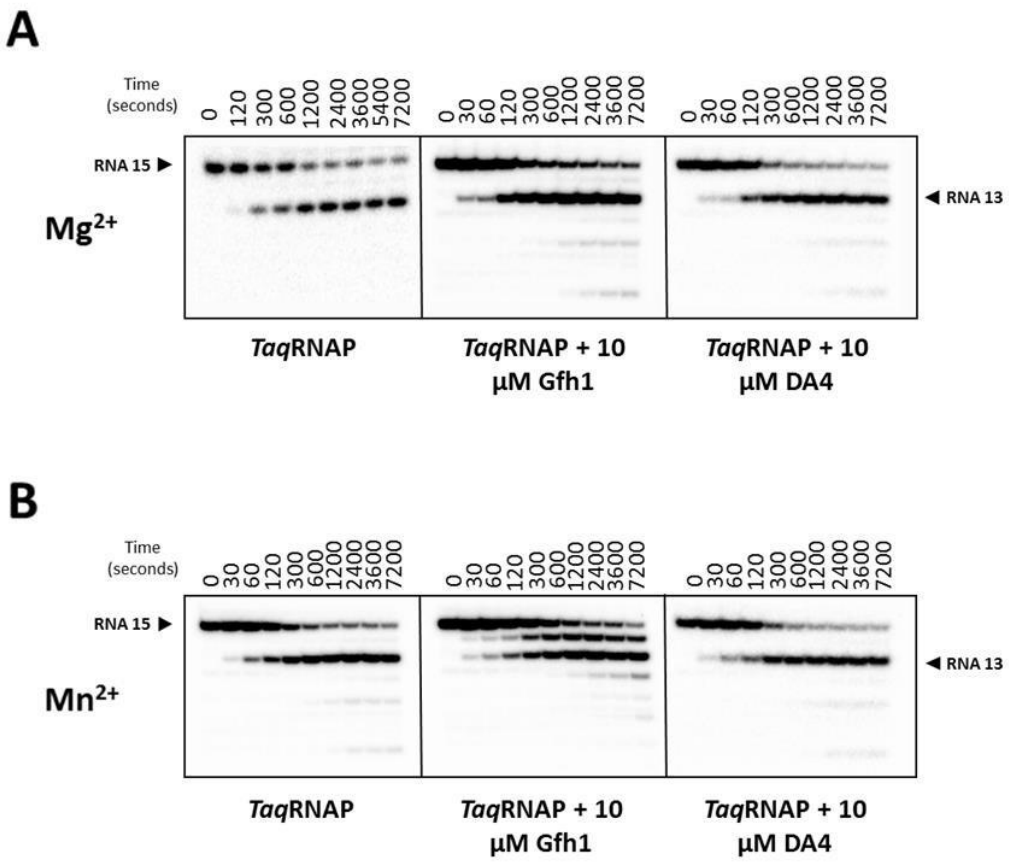


Figure 29. Gfh1 induces cleavage of one nucleotide. (A) Representative 23% PAGE of EC(GA) transcript cleavage in the presence of Mg^{2+} (**B)** Representative 23% PAGE of EC(GA) transcript cleavage in the presence of Mn^{2+}. Assays performed at 40^{o}C, pH 6.8.

3.4.4 Gfh1 Preferentially Inhibits NTP Incorporation from the Pre-translocated State

From the elongation data, the addition of Gfh1 induced strong pauses for *Taq*RNAP at the known pre- and post-translocated sites, but not at the backtracked site. Thus, the pre- and post- states were investigated further by measuring the rate of NTP incorporation, with and without Gfh1, from these states. **Figure 30A** illustrates the assay performed, and the obtained data are shown below. Elongation complexes were assembled at the appropriate pre- or post-translocated site using template 2 and template -1-8 (Bochkareva *et al*., 2012). The complexes were then incubated with 1 µM of the following cognate NTP, +/- 10 µM Gfh1 at 20^{o}C for the indicated time periods. and rates were fit using non-linear regression. The presence of Gfh1 reduced the rates of NTP incorporation for both states, however, the pre-translocated state showed much stronger inhibition, particularly in the presence of manganese. These were followed by titration assays (**Figure 31**) where either the concentration of NTP or Gfh1 was varied, illustrated by **Figure 31A** for a Gfh1 titration. Rates were determined for incorporation under 1, 5, 10, 20, or 100 µM NTP in the presence of either SB or 10 µM Gfh1 at 20^{o}C (**Figure 31B**). In the pre-translocated state, Gfh1 increases the NTP K_m dramatically, indicating competition with the NTP. In **Figure 31C**, Gfh1 is titrated at concentrations of 10, 3, 1, 0.3, 0.1, and 0.03 µM in the presence of 0.5 µM NTP. These data were also fit to the Michaelis-Menten equation and were subsequently fit to an exponential decay, and the rates accompanying the graphs are for the rate of NTP incorporation. The inhibitory activity of Gfh1 is again much more pronounced for the pre-translocated EC states. The inhibition is present but less noticeable for the template 2 EC(15/16) states which matches the similar ratios of activity for these positions in **Figure 30**.

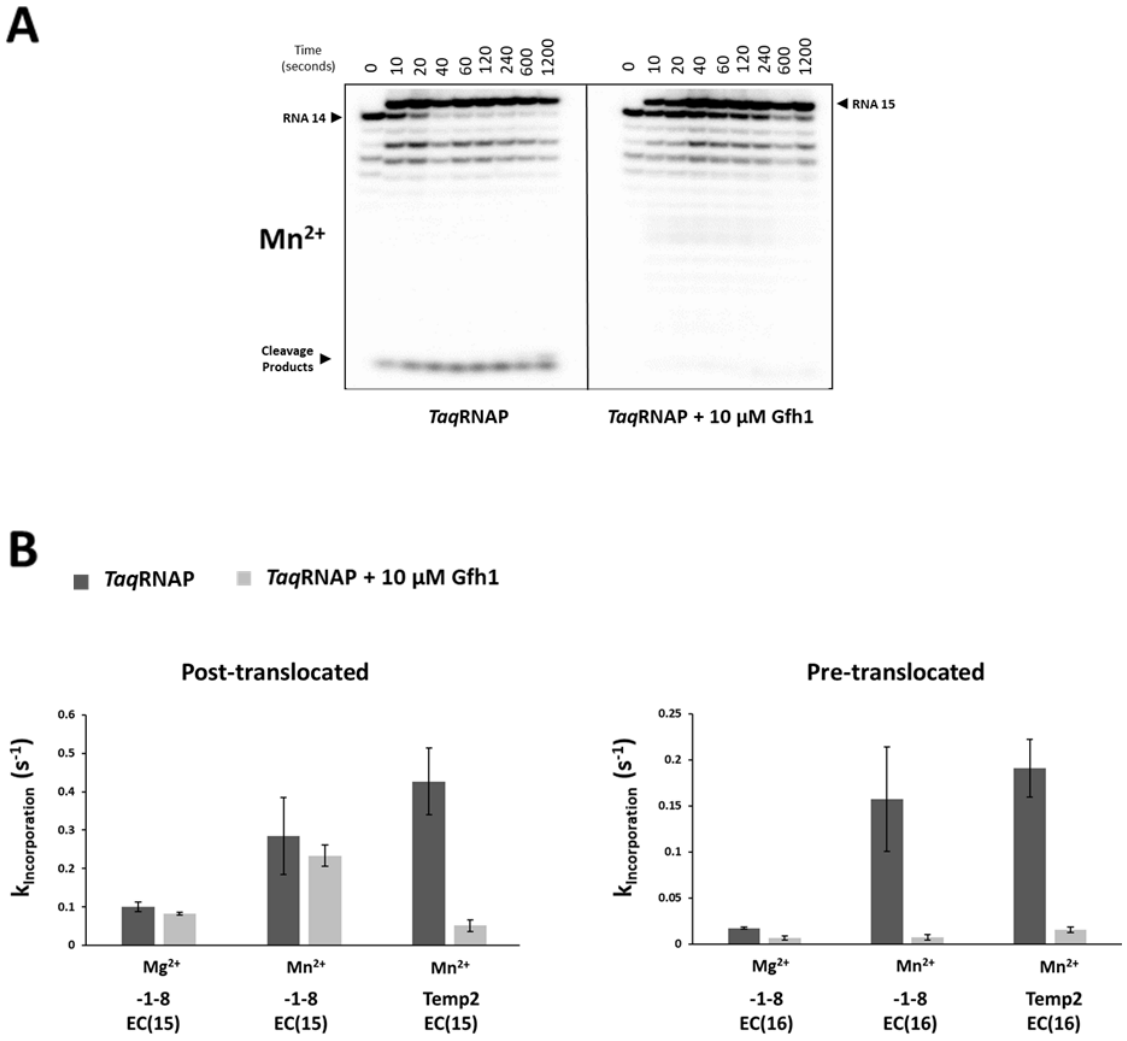


Figure 30. Gfh1 inhibits Nucleotide Incorporation (A) Representative 23% PAGE showing the setup of the NTP incorporation assays. (**B)** Rates of NTP incorporation on the scaffolds indicated incubated with Mn^{2+}/ Mg^{2+}. The values for each RNAP are the mean ± standard deviations of two independent experiments. Assays performed at 20^{o}C, pH 6.8.

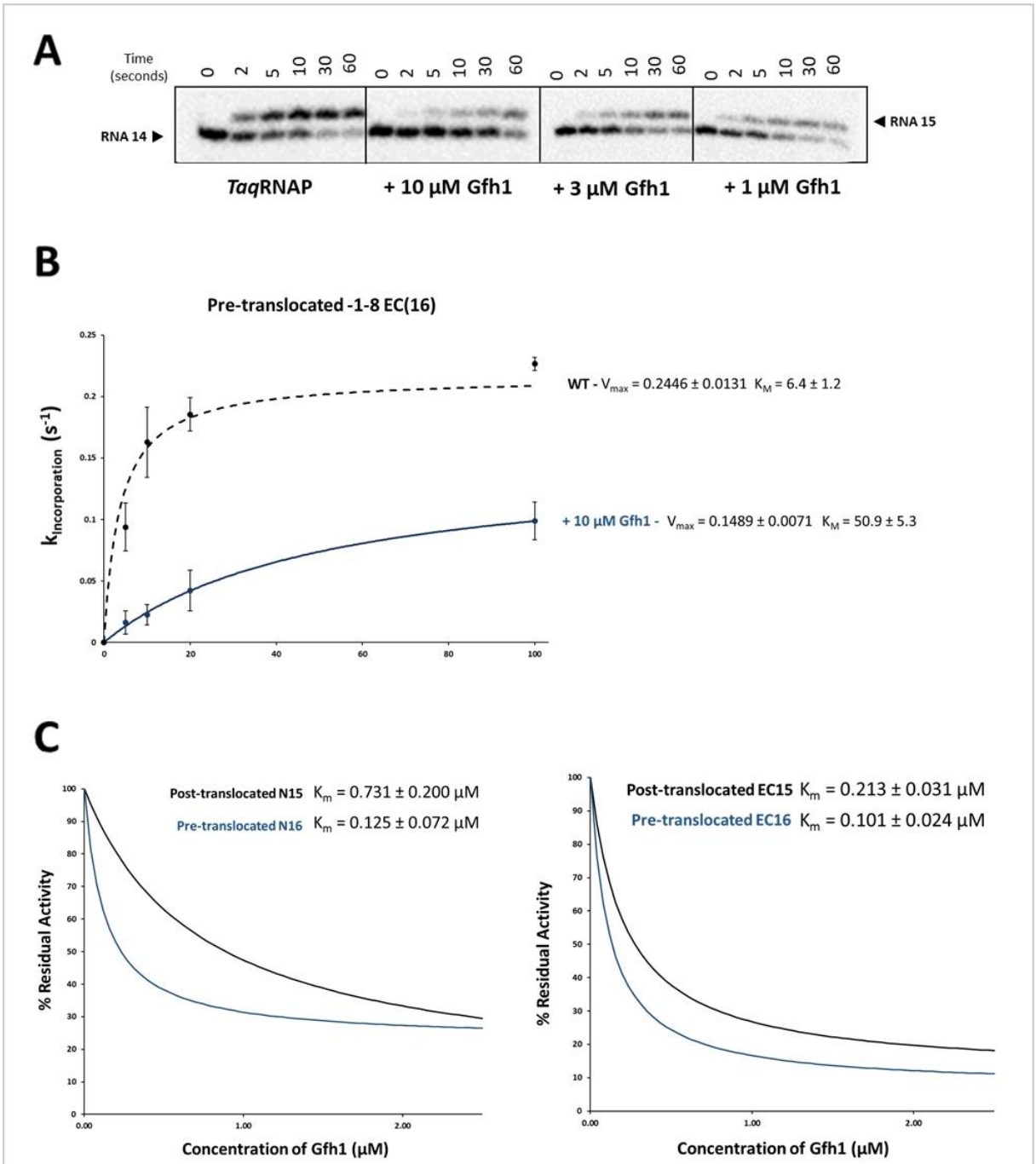


Figure 31. Gfh1 preferentially inhibits the RNAP pretranslocated state. A) Representative 23% PAGE showing the setup of the titration assays. (**B)** KM and Vmax values of NTP incorporation by *Taq*RNAP with SB or Gfh1 on the scaffold indicated, incubated with Mn^{2+}. The values for each RNAP are the mean ± standard deviations of two independent experiments. (**C)** Michaelis-Menten data for Gfh1 inhibition were fit to exponential decay with an assumption of KM = binding affinity.

3.4.5 Discussion

The Gfh1 SCBF had previously been shown to inhibit transcript cleavage and NTP addition in *Thermus thermophilus* and increase RNAP pausing and termination in *Deinococcus radiodurans* (Hogan *et al*., 2002; Symersky *et al*., 2006; Esyunina *et al*., 2016). Gfh1 has two states within the cell and undergoes a pH-induced conformational change into an active state at pH levels below 7 (Laptenko *et al.*, 2006). At a pH above 7, Gfh1 is in an inactive conformation which prevents binding to RNAP and has led to the suggestion that Gfh1 may play a role as a pH sensor (Laptenko *et al.*, 2006). Despite the high sequence similarity to GreA, the differences in activity are thought to be due to the four aspartate DDYDD motif of the flexible Gfh1 coiled-coil tip which stabilises the active site of RNAP is in a catalytically inactive conformation through co-ordination of the second metal ion (Symersky *et al*., 2006). Esyunina *et al*. (2016) showed that use of Mn^{2+} as the second metal ion significantly increases Gfh1 activity, which may be particularly significant in *Deinococcus radiodurans* as this organism accumulates intracellular manganese during the stress response. Using purified *Taq*RNAP, *Taq*Gfh1, and an inactive *Taq*Gfh1-DA4 mutant we present data that corroborates with previous studies and adds a preferential activity for *Taq*Gfh1 on pre-translocated RNAP complexes.

Deinococcus radiodurans Gfh1 has been shown to act on elemental, pre-translocated hairpin-dependent, and backtracked pauses (Agapov *et al*., 2017). We set out to determine if the activity of *Taq*Gfh1 is specific to a particular translocation state of RNAP. *Taq*Gfh1 demonstrated widespread pausing on template 1 and template 2 (**Figures 26**-**27**), but of the sites with known translocated states, pre- and post EC demonstrated the most pausing. Scaffolds with known pre- or post-translocated ECs were used to then measure NTP incorporation rate from these sites when challenged with *Taq*Gfh1. The presence of Gfh1 reduced the rates for both states, however, the pre-translocated state showed much stronger inhibition. This was tested further with titration of NTP and Gfh1 which again exhibited increased inhibitory activity of Gfh1 for the pre-translocated EC states. Unfortunately only one pre-translocated state was tested in NTP titration, to get a better determination of activity more scaffolds will need to be tested. Gfh1 also displayed competition with the NTP substrates, agreeing with the findings of Laptenko *et al*. (2006). To try and determine if Gfh1 was preferentially acting on pre-translocated states through increased binding affinity, the titration data were fit to exponential decay to show the amount of incorporation activity lost as the concentration of Gfh1 increases (**Figure 31C**)**.** Incorporation rate is reduced faster when the pre-translocated state is acted upon however the effects of inhibition also vary depending on the scaffold used.

We propose that while Gfh1 can act on many types of pauses and ECs, it predominantly affects the pre-translocated state of RNAP independent of hairpin formation. Additionally, *Taq*Gfh1 is shown to inhibit transcript cleavage, much like in *Thermus thermophilus*, but it is unknown why this is not the case for *Deinococcus radiodurans (*Esyunina *et al*., 2016). Also unknown is the induction of 1 nucleotide cleavage by *Taq*Gfh1 in cognate EC(GA) complexes, it is possible that Gfh1 stabilises the pre-translocated state of these complexes which may promote phosphorolysis.

4. Summary

4.1 Concluding Discussion

The work presented in this thesis has attempted to characterise the *in vitro* transcription behaviour of cyanobacterial RNAP with a focus on how cyanobacteria might compensate for the absence of proof-reading factor GreA. As GreA stimulates RNA hydrolysis, investigation into the RNA cleavage abilities of cyanobacterial RNAP and *Ec*RNAP was initially performed. Cyanobacterial RNAP was shown to perform intrinsic RNA cleavage at a rate 80-fold faster than *Ec*RNAP on assembled cognate complexes, and 30-40-fold faster on assembled mis-incorporated complexes. To identify the cause of this increased cleavage activity, a series of substitution mutations were constructed for *Ec*RNAP to correspond with the residues of *Ssp*RNAP. These substitutions were localised to the NADFDGD chelation site and the TL and BH mobile domains, which comprise main elements of the active site.

Several of these mutant *Ec*RNAPs demonstrated increased cleavage activity. *Ec*RNAP mutants β'G1136S (Q in this work), β'N792D, and β'F773V had been described in previously studies as possessing increased catalysis, reduced fidelity, and inefficient pausing and termination (Nedialkov *et al*., 2013; Bar-Nahum *et al*., 2005; Toulokhonov *et al*., 2007). These substitutions are thought to reduce BH and TL conformation dynamics to promote a catalytically active closed conformation with a kinked BH and a stabilised THB (Nedialkov *et al*., 2012; Zhang *et al*., 2010). This had led us to propose that as these substitutions occur naturally in cyanobacterial RNAP, this enzyme exists in an inherently more closed and catalytically active conformation.

To see if the cyanobacterial RNAPs also exhibited increased elongation and pause-resistance, the enzymes were made to transcribe templates which encoded known pause signals. Cyanobacterial RNAPs did respond less strongly at these pause signals and were particularly inefficient at recognising the hairpin-forming *his* and termination sites. However, this pause-resistance may only be partially caused by a closed active site conformation. Artsimovitch *et al.* (2000) showed that *Bacillus subtilis* RNAP also ignores *his* pause signals and determined that this was due, in part, to altered contacts between the β and β' subunits and downstream DNA. *Bacillus subtilis* RNAP is missing the β904-950 region of *Ec*RNAP suggested to interact with the hairpin (Wang *et al*., 1997). However, Artsimovitch *et al.* (2000) has shown that replacement of this region in *Ec*RNAP does not hinder *his* pause recognition but likely reduces hairpin stability, thus other elements must be contributing to *his* pause recognition. Interestingly, cyanobacterial RNAP is also lacking this region of the β subunit. As cyanobacteria also have no identified Rho factor, they seem to rely solely on Rho-independent intrinsic transcription which consist of a hairpin formation followed by a poly-U

tract for destabilisation of the complex. A key question is why this is the case when hairpin-dependent termination appears to be inefficient in these organisms, seen in this work and that of others, and whether these signals are used mainly as structures to stabilise the transcript for further RNA processing as has been suggested for chloroplasts (Vijayan *et al*. 2011; Stern and Gruissem, 1987). This is in contrast to the data of Imashimizu *et al*. (2011) who report increased pausing by cyanobacterial RNAP, although this could be due to the use of the cyanobacterial psbA2 gene as the template. In our assays we used templates with known pause signals for *Ec*RNAP. In future work, we would like to test on DNA templates derived from cyanobacteria.

The active site of cyanobacterial RNAP was also found to be more catalytically competent for hydrolysis than *Ec*RNAP. Cyanobacterial RNAP demonstrates a stronger ribozyme-like mechanism of transcript-assisted cleavage where the RNA 3' end of a backtracked complex assists in positioning the reactants. These RNAPs also benefit from the participation of a general base which aids proton transfer and activation of the attacking group during catalysis. Additionally, cyanobacterial RNAP requires a lower activation energy to accomplish RNA hydrolysis. Although Mg^{2+}II binding was similar between the enzymes, cyanobacterial RNAP may be better able to co-ordinate this ion to promote more efficient catalysis through the proposed limited flexibility of the closed active site conformation, and the additional contributions to co-ordination and orientation outlined above.

The mobile TL, supported by the BH, undergoes conformational changes during the NAC to stabilise and co-ordinate the reactants in the active site. The TL is additionally involved in recognition of the correct NTP and couples this substrate selection to catalysis in an induced-fit mechanism where binding of the correct incoming NTP in the i+1 site induces folding of the TL into the TH, while a steric collision is generated by the presence of a non-cognate NTP (Yuzenkova *et al*., 2010; Kaplan *et al*., 2008). The TL of cyanobacteria can therefore act as both a positional, via co-ordinating residues β'H936 and β'R933, and an acid/base, possibly also through β'H936, catalyst to support efficient catalysis. The β'F773V and β'G1136S *Ec*RNAP mutants show reduced fidelity with less discrimination for the substrate NTP (Bar-Nahum *et al*., 2005; Nedialkov *et al*., 2013). This could be due to the more stabilised folded conformation of the TL suggested for these mutants. Steric hindrance by a non-cognate NTP delays catalysis and slows progression of the reaction by obstructing TL folding which allows time for the non-cognate NTP to vacate the active site or results in a slowly performed mis-incorporation event (Yuzenkova *et al*., 2010; Kaplan *et al*., 2008). If the TL had an increased probability via increased stabilisation to be in proximity to the i+1/E site, which hold the

incoming NTP, it could result in faster catalysis (Kaplan *et al*., 2008). RNAseq analysis did report a slightly increased percentage of C>A, C>U, A>C, A>G, and A>U mis-incorporation in cyanobacterial transcripts, indicating that cyanobacterial RNAP may have diminished discrimination for complementary binding with ATP. The rates of other types of mis-incorporation appeared mostly similar for both species. This was only shown for *Ssp*RNAP, it would be interesting to repeat the analysis using total RNA extracted for *Sel*RNAP and see if the same slight differences are present.

Esyunina *et al*. (2016) has performed a similar investigation into the specific activities of *Deinococcus radiodurans* RNAP, which do not encode an SI3 but have faster RNA cleavage than *Ec*RNAP, to show that lineage-specific variations of the TL can account for the differences seen in RNAP activity. They identify a particular residue of β'G1136M (as opposed to Q/S) to be largely responsible for the highly efficient catalysis seen in this organism. *Taq*RNAP also demonstrates increased cleavage and has high sequence similarity to cyanobacterial RNAP, with only 5 residue differences in both the TL and BH compared to the respective 11 and 10 found in *Ec*RNAP. However, the cleavage rates determined for cyanobacterial RNAP are still many times faster despite *Taq*RNAP naturally coding for many of the substitutions which increased cleavage in mutant *Ec*RNAP. The only large difference is the presence of the SI3 within the TL, which is absent in *Taq*RNAP.

In *E. coli*, the SI3 is known to be involved in pause recognition, pause escape and assists with termination (Artsimovitch *et al*., 2003; Ray-Soni *et al*., 2017). The SI3 has also demonstrated modulation of TL folding into TH and affecting rates of nucleotide addition and transcript cleavage (Windgassen *et al*., 2014). It is possible that the SI3 of cyanobacterial RNAP causes the fast cleavage ability through modulation of the cyanobacterial TL. Two *Ec*RNAP mutants were constructed with either an SI3 deletion, β'ΔSI3, or an SI3 replacement, β'*Ssp*SI3. Both showed reduced intrinsic hydrolytic ability and impaired GreA-mediated hydrolysis. *Ec*RNAP β'*Ssp*SI3 additionally displayed increased pausing. These mutants highlight the ability of the SI3 to modulate RNAP activity through the TL but are not an accurate representation of the actual species-specific function. The SI3 of cyanobacteria also encode multiple SBHMs, the purpose of which is yet to be identified. SBHM domains in general are thought to facilitate protein:protein or protein:nucleic acid interactions (Chlenov *et al*., 2005). Most evidence so far has shown interactions of this motif, within the β flap region, with σ or holoenzyme function and so the increased number of SBHMs may serve some role in regulating transcription initiation given the larger numbers of σ factors for cyanobacteria (Kuznedelov *et al*., 2002; Osanai *et al*., 2008).

The SI3 is also essential for the activity of transcription factor GreB, and responsible for the ability of another factor, DksA, to bind the secondary channel. (Zhang *et al*., 2010; Furman *et al*., 2013). In this regard, the large SI3 of cyanobacterial RNAP may occlude the entrance to the secondary channel, lending more evidence for a lack of regulation through this structural domain. Furthermore, the binding of GreA requires an unfolded TL in order to access the active site through the secondary channel (Sekine *et al*., 2015). This would also be in disagreement with the current proposal of a stabilised closed and catalytically competent conformation for cyanobacterial RNAP.

Rates of GreA-dependent cleavage for *Ec*RNAP and the intrinsic cleavage of cyanobacterial RNAP were compared to show that cyanobacterial RNAPs only reach approximately 30% of the cleavage rate for *Ec*RNAP when GreA is present. However, considering the slower growth rate and polyploid nature of these organisms, this level of activity may be sufficient to carry out all necessary functions of cleavage. If cyanobacteria do co-ordinate replication on one chromosome and transcription on another, the risk of detrimental machinery collisions would be significantly reduced. Additionally, as cyanobacteria derive energy from photosynthesis they are particularly vulnerable to the damaging effects of excess light or UV irradiation. Sivaramakrishnan *et al*. (2017) have recently shown that the repair of double strand breaks is promoted by the backtracking of RNAPs upon encountering damage through assisting with the loading of repair protein RecA. They found that repair was repressed in the presence of GreA, which restarts backtracked complexes, suggesting a trade-off exists between the completion of transcription and the preservation of genome stability. This could be a reason for the absence of GreA in cyanobacteria where UV radiation-induced DNA damage may make up a large part of everyday life (Kumar *et al*., 2004; Vass *et al*., 2013). Insertions within active site elements, such as SI3 within the TL, have been proposed to allow quicker accumulation of advantageous adaptive mutations to new environmental conditions (Windgassen *et al*., 2014). It is possible that the large cyanobacterial SI3, in part, modulates faster intrinsic cleavage by the TL as a compromise between the rescue of stalled RNAPs and repair of DNA damage.

Gfh1 of *T.aq* was also briefly investigated in this thesis. *Taq*Gfh1 is shown to inhibit transcript cleavage and nucleotide addition activities. Although more data is required in order to determine the details of this inhibition, we suggest that Gfh1 acts predominantly on the pre-translocated state and may provide some stabilisation of this state.

4.1.1 Final Word

The proof reading GreA factor is absent in cyanobacteria. In this work we describe a potential compensatory mechanism which offsets the need for GreA-dependent cleavage. We propose that a sufficient level of RNA cleavage may be provided by the intrinsic activity of cyanobacterial RNAP without the need for external factors. This activity is realised through a possibly more closed and catalytically competent conformation of the active site and is supported by the action of a general base.

5. Appendix

5.1 List of Strains

Strain	Description	Resistance	Source/Reference
Synechococcus elongatus PCC 7942	WT Lab strain	-	Professor Conrad Mullineaux
Synechocystis sp PCC 6803	WT Lab strain	-	Professor Nigel Robinson
E. coli Δ*greA*Δ*greB*	GreA and GreB deleted from genome	-	Dr Pamela Gamba
E. coli Δ*greA*Δ*greB*Δ*dksA*	GreA, GreB, and DksA deleted from genome	-	Dr Pamela Gamba
E. coli rpoC-FLAG	rpoC or β' subunit has a C-terminal FLAG-tag	-	Dr Pamela Gamba
E. coli DH5α	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	-	New England Biolabs
E. coli XL10 gold	TetrD(mcrA)183 D(mcrCB-hsdSMR-mrr) 173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F′ proAB lacIqZDM15 Tn10(Tetr) Amy Camr]	tetracycline and chloramphenicol	Agilent Technologies
E. coli T7 Express	fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10	-	New England Biolabs
Synechococcus elongatus rpoC2-FLAG	rpoC or β' subunit has a C- terminal FLAG-tag	Spec/strep	Constructed in this project
Synechocystis sp rpoC2- his	rpoC or β' subunit has a C- terminal his-tag	Spec/Strep	Dr Yulia Yuzenkova

5.2 List of Plasmids

Plasmid	Description	Resistance	Source/Reference
pIA349	Used to amplify the pIA349 elongation template	Amp	Professor Irina Artsimovitch
pVS10	Encodes rpoA, rpoB and rpoC-his of *E. coli* RNAP	Amp	Professor Irina Artsimovitch
pVS14	Encodes the β'ΔSI3 *E. coli* mutant	Amp	Professor Irina Artsimovitch
pRL663	Encodes rpoA, rpoB and rpoC-his of *E. coli* RNAP	Amp	Professor Irina Artsimovitch
pET21-GreA	Encodes c-terminal his tagged GreA	Amp	Dr Flint Stevenson-Jones
pTaqABC	Encodes all need subunits of Taq with rpoc-his tag	Kan	Dr Mohammad Roghanian
pTaqGfh1	Encodes Gfh1 of taq, not tagged	Amp	Dr Mohammad Roghanian
pET28	Cloning plasmid	Kan	Dr Yulia Yuzenkova

5.3 List of Scaffolds

RNA Hydrolysis:

Name		Sequence		Type	Used for
GA36T	5′	/5Biosg/TGGCTATTCGCCGTGTCCCTCTCGATGGCTGTAAGT	3′	Template	Used with RNA15A to form EC(GA)
GA36NT	5′	ACCGATAAGCGGCACAGGGAGAGCTACCGACATTCA	3′	Non-template	
GU36T	5′	5Biosg/TGGCTATTCGCCGTGACCCTCTCGATGGCTGTAAGT	3′	Template	Used with RNA15A to form MEC(A)
GU36NT	5′	ACCGATAAGCGGCACTGGGAGAGCTACCGACATTCA	3′	Non-template	
GC36T	5′	5Biosg/TGGCTATTCGCCGTGGCCCTCTCGATGGCTGTAAGT	3′	Template	Used with RNA15U to form MEC(U)
GC36NT	5′	ACCGATAAGCGGCACCGGGAGAGCTACCGACATTCA	3′	Non-template	

RNAs:

Name		Sequence		Type	Used for
RNA 13	5′	AAUAAUCGAGAGG	3′	RNA	EC assembly
RNA15A	5′	AAUAAUCGAGAGGGA	3′	RNA	EC assembly
RNA15U	5′	AAUAAUCGAGAGGGU	3′	RNA	EC assembly
RNA-1-8	5′	AAUAAUCCGAAGU	3′	RNA	EC assembly
RNA nu	5′	GAGUCUGCGGCGAUG	3′	RNA	EC assembly

Elongation templates 1 and 2:

Name		Sequence		Type	Used for
1	5′	/5Biosg/TGCCGATGTTGTGTGCCGTACGATGCCCTCTCGATCCGTGTAAGACCACTTATG	3′	Template	Elongation
1	5′	ACGGCTACAACACAGGCATGCTACGGGAGAGCTAGGCACATTCTGGTGAAATAC	3′	Non-template	Elongation
2	5′	5Biosg/GGTGTCCATTGGGATGGCTATTCGCCGTGTCCCTCTCGATGGCTGTAAGT	3′	Template	Elongation
2	5′	CCACAGGTAACCCTACCGATAAGCGGCACAGGGAGAGCTACCGACATTCA	3′	Non-template	Elongation

NTP Incorporation templates -1-8 and Nu:

Name		Sequence		Type	Used for
-1-8 T	5′	/5Biosg/TGAATGTCATTAGGCTTCATGACTGCCGCTTATCGGT	3′	Template	Elongation
-1-8 NT	5′	ACTTACAGTAATCCGAAGTACTGACGGCGAATAGCCA	3′	Non-template	Elongation
Nu T	5′	5Biosg/GGGTCCTGTCTGAAATCTACATCGCCGC	3′	Template	Elongation
Nu NT	5′	AGATTTCAGACAGGACCC	3′	Non-template	Elongation

5.4 Mass Spectrometry Data

List of proteins eluted with purified *Sel*RNAP from different circadian phases. Proteins with unknown function are highlighted red. These were briefly analysed and show no similarity to the SCBFs. They are currently being cloned in order to test whether they do act on RNAP and to try and determine function. Dx-Dx denotes proteins found in both phases.

D5 (Night)	D5-D11	D11 (Dawn)
Q31L41	Fructose-bisphosphate aldolase	Porphobilinogen synthase
Phosphoribulokinase	Transaldolase	Phycocyanin linker protein 9K
2-phospho-D-glycerate dehydratase (enolase)	**NusG**	CheA signal transduction histidine kinase
dehydratase FabZ	**Bacterial nucleoid protein Hbs**	TPR repeat
phycobilisome rod linker protein	O-acetylhomoserine/O-acetylserine sulfhydrylase	N-acetyl-gamma-glutamyl-phosphate reductase
Glutamine synthetase, type I	Iron deficiency-induced protein A	
S-adenosyl meth synthetase		
ATP synthase subunit beta		
GDP-mannose 4,6-dehydratase		
HAD superfamily hydrolase subfamily iA var 3		
D-fructose-1,6-bisphosphate 1-phosphohydrolase class 2		

D11-L11	L11 (Dusk)	L11-D11
SigA1	**Q31NE2**	**Q31NP0**
Phosphoglycerate kinase	**Q31NG3**	**Q31KJ2**
SSU ribosomal protein S1P	**Q31R92**	
IMP dehydrogenase related 2	**Q31KJ7**	
Ribulose 1,5-bisphosphate carboxylase	**SigF**	
Dihydrolipoyl dehydrogenase	HAD superfamily hydrolase subfamily IIB	
Ketol-acid reductoisomerase	tRNA ligase beta subunit	
NAD kinase	30 kD rod-rod linker	
	Transcriptional regulator, XRE	
	Ribonuclease	
	DNA gyrase	
	Cold shock protein	
	Nitrate transport protein NrtA	
	DNA topoisomerase 1	
	Glycerol-3-phosphate acyltransferase	
	30S ribosomal protein	
	50S ribosomal protein (various	
	ChlA	
	ATP-dependent Clp protease	
	Translation initiation factor (various	

6. References

Agapov, A., Olina, A., Esyunina, D. & Kulbachinskiy, A. (2017) Gfh factors and NusA cooperate to stimulate transcriptional pausing and termination. *FEBS Letters*. 591 (6), 946–953.

Allison, L.A., Moyle, M., Shales, M. & James Ingles, C. (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell*. 42 (2), 599–610.

Artsimovitch, I. (2003) A New Class of Bacterial RNA Polymerase Inhibitor Affects Nucleotide Addition. *Science*. 302 (5645), 650–654.

Artsimovitch, I. & Landick, R. (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proceedings of the National Academy of Sciences of the United States of America*. 97 (13), 7090–7095.

Artsimovitch, I. & Landick, R. (2002) The Transcriptional Regulator RfaH Stimulates RNA Chain Synthesis after Recruitment to Elongation Complexes by the Exposed Nontemplate DNA Strand. *Cell*. 109 (2), 193–203.

Artsimovitch, I., Svetlov, V., Anthony, L., Burgess, R.R. & Landick, R. (2000) RNA Polymerases from *Bacillus subtilis* and *Escherichia coli* Differ in Recognition of Regulatory Signals In Vitro. *Journal of Bacteriology*. 182 (21), 6027–6035.

Artsimovitch, I., Svetlov, V., Murakami, K.S. & Landick, R. (2003) Co-overexpression of *Escherichia coli* RNA Polymerase Subunits Allows Isolation and Analysis of Mutant Enzymes Lacking Lineage-specific Sequence Insertions. *Journal of Biological Chemistry*. 278 (14), 12344–12355.

Artsimovitch, I., Svetlov, V., Nemetski, S.M., Epshtein, V., Cardozo, T. & Nudler, E. (2011) Tagetitoxin Inhibits RNA Polymerase through Trapping of the Trigger Loop. *The Journal of Biological Chemistry*. 286 (46), 40395–40400.

Ball, S., Colleoni, C., Cenci, U., Raj, J.N. & Tirtiaux, C. (2011) The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. *Journal of Experimental Botany*. 62 (6), 1775–1801.

Bar-Nahum, G., Epshtein, V., Ruckenstein, A.E., Rafikov, R., Mustaev, A. & Nudler, E. (2005) A Ratchet Mechanism of Transcription Elongation and Its Control. *Cell*. 120 (2), 183–193.

Bell-Pedersen, D., Cassone, V.M., Earnest, D.J., Golden, S.S., Hardin, P.E., Thomas, T.L. & Zoran, M.J. (2005) Circadian Rhythms from multiple oscillators: Lessons from diverse organisms. *Nature reviews. Genetics*. 6 (7), 544–556.

Blank, A., Gallant, J.A., Burgess, R.R. & Loeb, L.A. (1986) An RNA polymerase mutant with reduced accuracy of chain elongation. *Biochemistry*. 25 (20), 5920–5928.

Bochkareva, A., Yuzenkova, Y., Tadigotla, V.R. & Zenkin, N. (2012a) Factor-independent transcription pausing caused by recognition of the RNA–DNA hybrid sequence. *The EMBO Journal*. 31 (3), 630–639.

Bochkareva, A., Yuzenkova, Y., Tadigotla, V.R. & Zenkin, N. (2012b) Factor-independent transcription pausing caused by recognition of the RNA–DNA hybrid sequence. *The EMBO Journal*. 31 (3), 630–639.

Börner, T., Aleynikova, A.Y., Zubo, Y.O. & Kusnetsov, V.V. (2015) Chloroplast RNA polymerases: Role in chloroplast biogenesis. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 1847 (9), 761–769.

Borukhov, S., Sagitov, V. & Goldfarb, A. (1993) Transcript cleavage factors from *E. coli. Cell*. 72 (3), 459–466.

Cagliero, C., Zhou, Y.N. & Jin, D.J. (2014) Spatial organization of transcription machinery and its segregation from the replisome in fast-growing bacterial cells. *Nucleic Acids Research*. 42 (22), 13696–13705.

Carter, R. & Drouin, G. (2010) The Increase in the Number of Subunits in Eukaryotic RNA Polymerase III Relative to RNA Polymerase II Is due to the Permanent Recruitment of General Transcription Factors. *Molecular Biology and Evolution*. 27 (5), 1035–1043.

Castro, C., Smidansky, E., Maksimchuk, K.R., Arnold, J.J., Korneeva, V.S., Götte, M., Konigsberg, W. & Cameron, C.E. (2007) Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. *Proceedings of the National Academy of Sciences of the United States of America*. 104 (11), 4267–4272.

Chanfreau, G.F. (2013) Zinc'ing down RNA polymerase I. *Transcription*. 4 (5), 217–220.

Chen, A.H., Afonso, B., Silver, P.A. & Savage, D.F. (2012) Spatial and Temporal Organization of Chromosome Duplication and Segregation in the Cyanobacterium *Synechococcus elongatus* PCC 7942 Mark Isalan (ed.). *PLoS ONE*. 7 (10), e47837.

Chen, H., Tang, H. & Ebright, R.H. (2003) Functional interaction between RNA polymerase alpha subunit C-terminal domain and sigma70 in UP-element- and activator-dependent transcription. *Molecular Cell*. 11 (6), 1621–1633.

Cheung, A.C.M. & Cramer, P. (2011) Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*. 471 (7337), 249–253.

Chi, W., He, B., Manavski, N., Mao, J., Ji, D., Lu, C., Rochaix, J.D., Meurer, J. & Zhang, L. (2014) RHON1 Mediates a Rho-Like Activity for Transcription Termination in Plastids of Arabidopsis thaliana. *The Plant Cell*. 26 (12), 4918–4932.

Chlenov, M., Masuda, S., Murakami, K.S., Nikiforov, V., Darst, S.A. & Mustaev, A. (2005) Structure and Function of Lineage-specific Sequence Insertions in the Bacterial RNA Polymerase β′ Subunit. *Journal of Molecular Biology*. 353 (1), 138–154.

Ciampi, M.S. (2006) Rho-dependent terminators and transcription termination. *Microbiology*. 152 (9), 2515–2528.

Cramer, P., Bushnell, D.A. & Kornberg, R.D. (2001) Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. *Science*. 292 (5523), 1863–1876.

Da, L.-T., Wang, D. & Huang, X. (2012) Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II. *Journal of the American Chemical Society*. 134 (4), 2399–2406.

Daines, S.J. & Lenton, T.M. (2016) The effect of widespread early aerobic marine ecosystems on methane cycling and the Great Oxidation. *Earth and Planetary Science Letters*. 43442–51.

Daniell, H., Lin, C.-S., Yu, M. & Chang, W.-J. (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology*. 17.

Dulin, D., Bauer, D.L.V., Malinen, A.M., Bakermans, J.J.W., Kaller, M., Morichaud, Z., Petushkov, I., Depken, M., Brodolin, K., Kulbachinskiy, A. & Kapanidis, A.N. (2018) Pausing controls branching between productive and non-productive pathways during initial transcription in bacteria. *Nature Communications*. 9.

Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M.E. & Nudler, E. (2011) Linking RNA Polymerase Backtracking to Genome Instability. *Cell*. 146 (4), 533–543.

Ebright, R.H. (2000) RNA Polymerase: Structural Similarities Between Bacterial RNA Polymerase and Eukaryotic RNA Polymerase II. *Journal of Molecular Biology*. 304 (5), 687–698.

Epshtein, V., Mustaev, A., Markovtsov, V., Bereshchenko, O., Nikiforov, V. & Goldfarb, A. (2002) Swing-Gate Model of Nucleotide Entry into the RNA Polymerase Active Center. *Molecular Cell*. 10 (3), 623–634.

Esyunina, D., Agapov, A. & Kulbachinskiy, A. (2016) Regulation of transcriptional pausing through the secondary channel of RNA polymerase. *Proceedings of the National Academy of Sciences*. 113 (31), 8699–8704.

Esyunina, D., Turtola, M., Pupov, D., Bass, I., Klimašauskas, S., Belogurov, G. & Kulbachinskiy, A. (2016) Lineage-specific variations in the trigger loop modulate RNA proofreading by bacterial RNA polymerases. *Nucleic Acids Research*. 44 (3), 1298–1308.

Farnham, P.J. & Platt, T. (1981) Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro. *Nucleic Acids Research*. 9 (3), 563–577.

Feklistov, A. & Darst, S.A. (2009) Promoter recognition by bacterial alternative σ factors: the price of high selectivity? *Genes & Development*. 23 (20), 2371.

Fish, R.N. & Kane, C.M. (2002) Promoting elongation with transcript cleavage stimulatory factors. *Biochimica Et Biophysica Acta*. 1577 (2), 287–307.

Freemont, P.S., Friedman, J.M., Beese, L.S., Sanderson, M.R. & Steitz, T.A. (1988) Cocrystal structure of an editing complex of Klenow fragment with DNA. *Proceedings of the National Academy of Sciences*. 85 (23), 8924–8928.

Furman, R., Tsodikov, O.V., Wolf, Y.I. & Artsimovitch, I. (2013) An insertion in the catalytic trigger loop gates the secondary channel of RNA polymerase. *Journal of molecular biology*. 425 (1), 82–93.

Gabizon, R., Lee, A., Vahedian-Movahed, H., Ebright, R.H. & Bustamante, C.J. (2018) Pause sequences facilitate entry into long-lived paused states by reducing RNA polymerase transcription rates. *Nature Communications*. 9.

Ghosh, P., Ishihama, A. & Chatterji, D. (2001) Escherichia coli RNA polymerase subunit ω and its N-terminal domain bind full-length β′ to facilitate incorporation into the α2β subassembly. *European Journal of Biochemistry*. 268 (17), 4621–4627.

Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., Hutchison, C.A., Smith, H.O. & Venter, J.C. (2006) Essential genes of a minimal bacterium. *Proceedings of the National Academy of Sciences of the United States of America*. 103 (2), 425–430.

Glyde, R., Ye, F., Jovanovic, M., Kotta-Loizou, I., Buck, M. & Zhang, X. (2018) Structures of Bacterial RNA Polymerase Complexes Reveal the Mechanism of DNA Loading and Transcription Initiation. *Molecular Cell*. 70 (6), 1111-1120.e3.

Gnatt, A.L. (2001) Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 A Resolution. *Science*. 292 (5523), 1876–1882.

Goto-Seki, A., Shirokane, M., Masuda, S., Tanaka, K. & Takahashi, H. (1999) Specificity crosstalk among group 1 and group 2 sigma factors in the cyanobacterium *Synechococcus sp.* PCC7942: in vitro specificity and a phylogenetic analysis. *Molecular Microbiology*. 34 (3), 473–484.

Gould, S.B., Waller, R.F. & McFadden, G.I. (2008) Plastid Evolution. *Annual Review of Plant Biology*. 59 (1), 491–517.

Gruber, T.M. & Gross, C.A. (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annual Review of Microbiology*. 57441–466.

Gunnelius, L., Hakkila, K., Kurkela, J., Wada, H., Tyystjärvi, E. & Tyystjärvi, T. (2014) The omega subunit of the RNA polymerase core directs transcription efficiency in cyanobacteria. *Nucleic Acids Research*. 42 (7), 4606–4614.

Gunnelius, L., Kurkela, J., Hakkila, K., Koskinen, S., Parikainen, M. & Tyystjärvi, T. (2014) The ω Subunit of RNA Polymerase Is Essential for Thermal Acclimation of the Cyanobacterium *Synechocystis Sp.* PCC 6803. *PLOS ONE*. 9 (11), e112599.

Hager, D.A., Jin, D.J. & Burgess, R.R. (1990) Use of mono Q high resolution ion exchange chromatography to obtain highly pure and active *Escherichia coli* RNA polymerase. *Biochemistry*. 29 (34), 7890–7894.

Harley, C.B. & Reynolds, R.P. (1987) Analysis of *E. coli* promoter sequences. *Nucleic Acids Research*. 15 (5), 2343–2361.

Hausner, W., Lange, U. & Musfeldt, M. (2000) Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *The Journal of Biological Chemistry*. 275 (17), 12393–12399.

Herbert, K.M., La Porta, A., Wong, B.J., Mooney, R.A., Neuman, K.C., Landick, R. & Block, S.M. (2006) Sequence-Resolved Detection of Pausing by Single RNA Polymerase Molecules. *Cell*. 125 (6), 1083–1094

de Hoon, M.J.L., Makita, Y., Nakai, K. & Miyano, S. (2005) Prediction of Transcriptional Terminators in *Bacillus subtilis* and Related Species. *PLoS Computational Biology*. 1 (3), .

Hu, Y., Morichaud, Z., Chen, S., Leonetti, J.-P. & Brodolin, K. (2012) *Mycobacterium tuberculosis* RbpA protein is a new type of transcriptional activator that stabilizes the σ A-containing RNA polymerase holoenzyme. *Nucleic Acids Research*. 40 (14), 6547–6557.

Huang, X., Wang, D., Weiss, D.R., Bushnell, D.A., Kornberg, R.D. & Levitt, M. (2010) RNA polymerase II trigger loop residues stabilize and position the incoming nucleotide triphosphate in transcription. *Proceedings of the National Academy of Sciences of the United States of America*. 107 (36), 15745–15750.

Huckauf, J., Nomura, C., Forchhammer, K. & Hagemann, M. (2000) Stress responses of *Synechocystis sp.* strain PCC 6803 mutants impaired in genes encoding putative alternative sigma factors. *Microbiology*. 146 (11), 2877–2889.

Imamura, S. & Asayama, M. (2009) Sigma Factors for Cyanobacterial Transcription. *Gene Regulation and Systems Biology*. 365–87.

Imashimizu, M., Fujiwara, S., Tanigawa, R., Tanaka, K., Hirokawa, T., Nakajima, Y., Higo, J. & Tsuzuki, M. (2003) Thymine at −5 Is Crucial for cpc Promoter Activity of *Synechocystis sp*. Strain PCC 6714. *Journal of Bacteriology*. 185 (21), 6477–6480.

Imashimizu, M., Oshima, T., Lubkowska, L. & Kashlev, M. (2013) Direct assessment of transcription fidelity by high-resolution RNA sequencing. *Nucleic Acids Research*. 41 (19), 9090–9104.

Ishihama, A. (1981) Subunit of assembly of *Escherichia coli* RNA polymerase. *Advances in Biophysics*. 141–35.

Iyer, L.M., Koonin, E.V. & Aravind, L. (2004) Evolution of bacterial RNA polymerase: implications for large-scale bacterial phylogeny, domain accretion, and horizontal gene transfer. *Gene*. 33573–88.

Iyer, L.M., Koonin, E.V. & Aravind, L. (2003) Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Structural Biology*. 31.

Johnson, C.H. (2007) Bacterial circadian programs. *Cold Spring Harbor Symposia on Quantitative Biology*. 72395–404.

Jokerst, R.S., Weeks, J.R., Zehring, W.A. & Greenleaf, A.L. (1989) Analysis of the gene encoding the largest subunit of RNA polymerase II in Drosophila. *MGG Molecular & General Genetics*. 215 (2), 266–275.

Jovanovic, M., Burrows, P.C., Bose, D., Cámara, B., Wiesler, S., Zhang, X., Wigneshweraraj, S., Weinzierl, R.O.J. & Buck, M. (2011) Activity Map of the *Escherichia coli* RNA Polymerase Bridge Helix. *Journal of Biological Chemistry*. 286 (16), 14469–14479.

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., et al. (1996) Sequence Analysis of the Genome of the Unicellular Cyanobacterium *Synechocystis sp.* Strain PCC6803. II. Sequence Determination of the Entire Genome and Assignment of Potential Protein-coding Regions. *DNA Research*. 3 (3), 109–136.

Kang, J.Y., Mishanina, T.V., Bellecourt, M.J., Mooney, R.A., Darst, S.A. & Landick, R. (2018) RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. *Molecular Cell*. 69 (5), 802-815.e1.

Kapanidis, A.N., Margeat, E., Ho, S.O., Kortkhonjia, E., Weiss, S. & Ebright, R.H. (2006) Initial transcription by RNA polymerase proceeds through a DNA scrunching mechanism. *Science* 314 (5802), 1144–1147.

Kaplan, C.D., Larsson, K.-M. & Kornberg, R.D. (2008) The RNA Polymerase II Trigger Loop Functions in Substrate Selection and Is Directly Targeted by α-Amanitin. *Molecular Cell*. 30 (5), 547–556.

Kireeva, M., Trang, C., Matevosyan, G., Turek-Herman, J., Chasov, V., Lubkowska, L. & Kashlev, M. (2018) RNA–DNA and DNA–DNA base-pairing at the upstream edge of the transcription bubble regulate translocation of RNA polymerase and transcription rate. *Nucleic Acids Research*. 46 (11), 5764–5775.

Kireeva, M.L. & Kashlev, M. (2009) Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*. 106 (22), 8900–8905.

Kireeva, M.L., Nedialkov, Y.A., Cremona, G.H., Purtov, Y.A., Lubkowska, L., Malagon, F., Burton, Z.F., Strathern, J.N. & Kashlev, M. (2008) Transient Reversal of RNA Polymerase II Active Site Closing Controls Fidelity of Transcription Elongation. *Molecular Cell*. 30 (5), 557–566.

Komissarova, N. & Kashlev, M. (1997) Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3′ end of the RNA intact and extruded. *Proceedings of the National Academy of Sciences*. 94 (5), 1755–1760.

Kopp, R.E., Kirschvink, J.L., Hilburn, I.A. & Nash, C.Z. (2005) The Paleoproterozoic snowball Earth: A climate disaster triggered by the evolution of oxygenic photosynthesis. *Proceedings of the National Academy of Sciences*. 102 (32), 11131–11136.

Kulasooriya, Sabharatna. (2011). S. A. Kulasooriya (2011) Cyanobacteria: Pioneers of Planet earth. Ceylon Journal of Science (Bio. Sci.) 40 (2): 71 – 88. Ceylon Journal of Science (Biological Sciences). 40. 71 – 88.

Korkhin, Y., Unligil, U.M., Littlefield, O., Nelson, P.J., Stuart, D.I., Sigler, P.B., Bell, S.D. & Abrescia, N.G.A. (2009) Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure. *PLOS Biology*. 7 (5), e1000102.

Kuznedelov, K., Korzheva, N., Mustaev, A. & Severinov, K. (2002) Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA–DNA hybrid length. *The EMBO Journal*. 21 (6), 1369–1378.

Lamour, V., Hogan, B.P., Erie, D.A. & Darst, S.A. (2006) Crystal structure of *Thermus aquaticus* Gfh1, a Gre-factor paralog that inhibits rather than stimulates transcript cleavage. *Journal of Molecular Biology*. 356 (1), 179–188.

Landick, R. (2006) The regulatory roles and mechanism of transcriptional pausing. *Biochemical Society Transactions*. 34 (6), 1062–1066.

Lane, W.J. & Darst, S.A. (2010a) Molecular Evolution of Multisubunit RNA Polymerases: Sequence Analysis. *Journal of Molecular Biology*. 395 (4), 671–685.

Lane, W.J. & Darst, S.A. (2010b) Molecular Evolution of Multisubunit RNA Polymerases: Structural Analysis. *Journal of Molecular Biology*. 395 (4), 686–704.

Laptenko, O., Kim, S.-S., Lee, J., Starodubtseva, M., Cava, F., Berenguer, J., Kong, X.-P. & Borukhov, S. (2006) pH-dependent conformational switch activates the inhibitor of transcription elongation. *The EMBO Journal*. 25 (10), 2131–2141.

Laptenko, O., Lee, J., Lomakin, I. & Borukhov, S. (2003) Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *The EMBO Journal*. 22 (23), 6322–6334.

Malagon, F., Kireeva, M.L., Shafer, B.K., Lubkowska, L., Kashlev, M. & Strathern, J.N. 2006) Mutations in the *Saccharomyces cerevisiae* RPB1 Gene Conferring Hypersensitivity to 6-Azauracil. *Genetics*. 172 (4), 2201–2209.

Malinen, A.M., NandyMazumdar, M., Turtola, M., Malmi, H., Grocholski, T., Artsimovitch, I. & Belogurov, G.A. (2014) CBR antimicrobials alter coupling between the bridge helix and the β subunit in RNA polymerase. *Nature Communications*. 53408.

Malinen, A.M., Turtola, M., Parthiban, M., Vainonen, L., Johnson, M.S. & Belogurov, G.A. (2012) Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Research*. 40 (15), 7442–7451.

Markson, J.S., Piechura, J.R., Puszynska, A.M. & O'Shea, E.K. (2013) Circadian Control of Global Gene Expression by the Cyanobacterial Master Regulator RpaA. *Cell*. 155 (6), 1396–1408.

Martin, W. & Kowallik, K. (1999) Annotated English translation of Mereschkowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren imPflanzenreiche'. *European Journal of Phycology*. 34 (3), 287–295.

Martinez-Rucobo, F.W. & Cramer, P. (2013) Structural basis of transcription elongation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. 1829 (1), 9–19.

Mathew, R. (2006) Deletion of the rpoZ gene, encoding the subunit of RNA polymerase, results in pleiotropic surface-related phenotypes in *Mycobacterium smegmatis*. *Microbiology*. 152 (6), 1741–1750.

Mazard, S., Penesyan, A., Ostrowski, M., Paulsen, I.T. & Egan, S. (2016) Tiny Microbes with a Big Impact: The Role of Cyanobacteria and Their Metabolites in Shaping Our Future. *Marine Drugs*. 14 (5)

McEwen, J.T., Machado, I.M.P., Connor, M.R. & Atsumi, S. (2013) Engineering *Synechococcus elongatus* PCC 7942 for continuous growth under diurnal conditions. *Applied and Environmental Microbiology*. 79 (5), 1668–1675.

Mejia, Y.X., Nudler, E. & Bustamante, C. (2015) Trigger loop folding determines transcription rate of *Escherichia coli's* RNA polymerase. *Proceedings of the National Academy of Sciences*. 112 (3), 743–748.

Merrick, M.J. (1993) In a class of its own — the RNA polymerase sigma factor σ;54 (σN). *Molecular Microbiology*. 10 (5), 903–909.

Merrikh, H., Machón, C., Grainger, W.H., Grossman, A.D. & Soultanas, P. (2011) Co-directional replication-transcription conflicts lead to replication restart. *Nature*. 470 (7335), 554–557.

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H. & Severinov, K. (2001) Bacterial RNA polymerase subunit ω and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proceedings of the National Academy of Sciences*. 98 (3), 892–897.

Minakhin, L., Nechaev, S., Campbell, E.A. & Severinov, K. (2001) Recombinant *Thermus aquaticus* RNA Polymerase, a New Tool for Structure-Based Analysis of Transcription. *Journal of Bacteriology*. 183 (1), 71–76.

Mirkin, E.V. & Mirkin, S.M. (2007) Replication fork stalling at natural impediments. *Microbiology and molecular biology reviews: MMBR*. 71 (1), 13–35.

Miropolskaya, N., Esyunina, D. & Kulbachinskiy, A. (2017) Conserved functions of the trigger loop and Gre factors in RNA cleavage by bacterial RNA polymerases. *Journal of Biological Chemistry*. 292 (16), 6744–6752.

Miropolskaya, N., Nikiforov, V., Klimašauskas, S., Artsimovitch, I. & Kulbachinskiy, A. (2010) Modulation of RNA polymerase activity through trigger loop folding. *Transcription*. 1 (2), 89–94.

Mishanina, T.V., Palo, M.Z., Nayak, D., Mooney, R.A. & Landick, R. (2017) Trigger loop of RNA polymerase is a positional, not acid–base, catalyst for both transcription and proofreading. *Proceedings of the National Academy of Sciences of the United States of America*. 114 (26), E5103–E5112.

Mitchell, J.B.O., Thornton, J.M., Singh, J. & Price, S.L. (1992) Towards an understanding of the arginine-aspartate interaction. *Journal of Molecular Biology*. 226 (1), 251–262.

Mooney, R.A., Darst, S.A. & Landick, R. (2005) Sigma and RNA Polymerase: An On-Again, Off-Again Relationship? *Molecular Cell*. 20 (3), 335–345.

Mori, T., Binder, B. & Johnson, C.H. (1996) Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours. *Proceedings of the National Academy of Sciences*. 93 (19), 10183–10188.

Nabout, João & da Silva Rocha, Barbbara & Carneiro, Fernanda & C.L, Sant'Anna. (2013). How many species of Cyanobacteria are there? Using a discovery curve to predict the species number. Biodiversity and Conservation. 22. 10.1007/s10531-013-0561-x.

Nair, U., Ditty, J.L., Min, H. & Golden, S.S. (2002) Roles for Sigma Factors in Global Circadian Regulation of the Cyanobacterial Genome. *Journal of Bacteriology*. 184 (13), 3530–3538.

Nedialkov, Y.A., Nudler, E. & Burton, Z.F. (2012) RNA polymerase stalls in a post-translocated register and can hyper-translocate. *Transcription*. 3 (5), 260–269.

Nedialkov, Y.A., Opron, K., Assaf, F., Artsimovitch, I., Kireeva, M.L., Kashlev, M., Cukier, R.I., Nudler, E. & Burton, Z.F. (2013) The RNA polymerase bridge helix YFI motif in catalysis, fidelity and translocation. *Biochimica et biophysica acta*. 1829 (2), 187–198.

Nedialkov, Y.A., Opron, K., Caudill, H.L., Assaf, F., Anderson, A.J., Cukier, R.I., Wei, G. & Burton, Z.F. (2018) Hinge action versus grip in translocation by RNA polymerase. *Transcription*. 9 (1), 1–16.

Nickels, B.E. & Hochschild, A. (2004) Regulation of RNA Polymerase through the Secondary Channel. *Cell*. 118 (3), 281–284.

Nielsen, S. & Zenkin, N. (2013) Transcript assisted phosphodiester bond hydrolysis by eukaryotic RNA polymerase II. *Transcription*. 4 (5), 209–212.

Nowotny, M. & Yang, W. (2006) Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. *The EMBO journal*. 25 (9), 1924–1933.

Nudler, E. (2009) RNA polymerase active center: the molecular engine of transcription. RNA Polymerase Active Center: The Molecular Engine of Transcription. *Annual review of biochemistry, Annual review of biochemistry*. 78, 78335, 335–361.

Och, L.M. & Shields-Zhou, G.A. (2012) The Neoproterozoic oxygenation event: Environmental perturbations and biogeochemical cycling. *Earth-Science Reviews*. 110 (1–4), 26–57.

Olson, J.M. (2006) Photosynthesis in the Archean Era. *Photosynthesis Research*. 88 (2), 109–117.

Opalka, N., Mooney, R.A., Richter, C., Severinov, K., Landick, R. & Darst, S.A. (2000) Direct localization of a β-subunit domain on the three-dimensional structure of Escherichia coli RNA polymerase. *Proceedings of the National Academy of Sciences*. 97 (2), 617–622.

Orlova, M., Newlands, J., Das, A., Goldfarb, A. & Borukhov, S. (1995) Intrinsic transcript cleavage activity of RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*. 92 (10), 4596–4600.

Osanai, T., Ikeuchi, M. & Tanaka, K. (2008) Group 2 sigma factors in cyanobacteria. *Physiologia Plantarum*. 133 (3), 490–506.

Paget, M.S. (2015) Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules*. 5 (3), 1245–1265.

Paget, M.S. & Helmann, J.D. (2003) The σ70 family of sigma factors. *Genome Biology*. 4 (1), 203.

Perederina, A., Svetlov, V., Vassylyeva, M.N., Tahirov, T.H., Yokoyama, S., Artsimovitch, I. & Vassylyev, D.G. (2004) Regulation through the secondary channel--structural framework for ppGpp-DksA synergism during transcription. *Cell*. 118 (3), 297–309.

Peters, J.M., Vangeloff, A.D. & Landick, R. (2011) Bacterial Transcription Terminators: The RNA 3′-End Chronicles. *Journal of molecular biology*. 412 (5), 793–813.

Pollari, M., Gunnelius, L., Tuominen, I., Ruotsalainen, V., Tyystjärvi, E., Salminen, T. & Tyystjärvi, T. (2008) Characterization of Single and Double Inactivation Strains Reveals New Physiological Roles for Group 2 σ Factors in the Cyanobacterium *Synechocystis sp. PCC* 6803. *Plant Physiology*. 147 (4), 1994–2005.

Ponce-Toledo, R.I., Deschamps, P., López-García, P., Zivanovic, Y., Benzerara, K. & Moreira, D. (2017) An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids. *Current Biology*. 27 (3), 386–391.

Poole, A.M. & Logan, D.T. (2005) Modern mRNA Proofreading and Repair: Clues that the Last Universal Common Ancestor Possessed an RNA Genome? *Molecular Biology and Evolution*. 22 (6), 1444–1455.

Ray-Soni, A., Mooney, R.A. & Landick, R. (2017) Trigger loop dynamics can explain stimulation of intrinsic termination by bacterial RNA polymerase without terminator hairpin contact. *Proceedings of the National Academy of Sciences*. 114 (44), E9233–E9242.

Rédei, G.P. (ed.) (2008) 'Transcription Termination in Prokaryotes', in *Encyclopedia of Genetics, Genomics, Proteomics and Informatics*. Dordrecht: Springer Netherlands. pp. 1996– 1997.

Revyakin, A., Liu, C., Ebright, R.H. & Strick, T.R. (2006) Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science (New York, N.Y.)*. 314 (5802), 1139–1143.

Richard Blaustein; The Great Oxidation Event: Evolving understandings of how oxygenic life on Earth began, *BioScience*, Volume 66, Issue 3, 1 March 2016, Pages 189–195

Roghanian, M., Yuzenkova, Y. & Zenkin, N. (2011) Controlled interplay between trigger loop and Gre factor in the RNA polymerase active centre. *Nucleic Acids Research*. 39 (10), 4352–4359.

Saecker, R.M., Record, M.T. & deHaseth, P.L. (2011) Mechanism of Bacterial Transcription Initiation: RNA Polymerase - Promoter Binding, Isomerization to Initiation-Competent Open Complexes, and Initiation of RNA Synthesis. *Journal of molecular biology*. 412 (5), 754–771.

Samanta, S. & Martin, C.T. (2013) Insights into the Mechanism of Initial Transcription in *Escherichia coli* RNA Polymerase. *The Journal of Biological Chemistry*. 288 (44), 31993– 32003.

Sekerci, Y. & Petrovskii, S. (2015) Mathematical Modelling of Plankton–Oxygen Dynamics Under the Climate Change. *Bulletin of Mathematical Biology*. 77 (12), 2325–2353.

Sekine, S., Murayama, Y., Svetlov, V., Nudler, E. & Yokoyama, S. (2015) Ratcheting of RNA polymerase toward structural principles of RNA polymerase operations. *Transcription*. 6 (3), 56–60.

Severinov, K. (2000) RNA polymerase structure–function: insights into points of transcriptional regulation. *Current Opinion in Microbiology*. 3 (2), 118–125.

Shaevitz, J.W., Abbondanzieri, E.A., Landick, R. & Block, S.M. (2003) Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature*. 426 (6967), 684–687.

Shestakov, S.V. & Khyen, N.T. (1970) Evidence for genetic transformation in blue-green alga Anacystis nidulans. *Molecular & general genetics: MGG*. 107 (4), 372–375.

Sidorenkov, I., Komissarova, N. & Kashlev, M. (1998) Crucial Role of the RNA:DNA Hybrid in the Processivity of Transcription. *Molecular Cell*. 2 (1), 55–64.

Singh, J.S., Kumar, A., Rai, A.N. & Singh, D.P. (2016) Cyanobacteria: A Precious Bio-resource in Agriculture, Ecosystem, and Environmental Sustainability. *Frontiers in Microbiology*. 7.

Smith, A.C. & Purton, S. (2002) The transcriptional apparatus of algal plastids. *European Journal of Phycology*. 37 (3), 301–311.

Smith, R.M. & Williams, S.B. (2006) Circadian rhythms in gene transcription imparted by chromosome compaction in the cyanobacterium Synechococcus elongatus. *Proceedings of the National Academy of Sciences of the United States of America*. 103 (22), 8564–8569.

Sosunov, V., Zorov, S., Sosunova, E., Nikolaev, A., Zakeyeva, I., Bass, I., Goldfarb, A., Nikiforov, V., Severinov, K. & Mustaev, A. (2005) The involvement of the aspartate triad of the active center in all catalytic activities of multisubunit RNA polymerase. *Nucleic Acids Research*. 33 (13), 4202–4211.

Sosunova, E., Sosunov, V., Epshtein, V., Nikiforov, V. & Mustaev, A. (2013a) Control of Transcriptional Fidelity by Active Center Tuning as Derived from RNA Polymerase Endonuclease Reaction. *Journal of Biological Chemistry*. 288 (9), 6688–6703.

Sosunova, E., Sosunov, V., Epshtein, V., Nikiforov, V. & Mustaev, A. (2013b) Control of Transcriptional Fidelity by Active Center Tuning as Derived from RNA Polymerase Endonuclease Reaction. *Journal of Biological Chemistry*. 288 (9), 6688–6703.

Sosunova, E., Sosunov, V., Kozlov, M., Nikiforov, V., Goldfarb, A. & Mustaev, A. (2003) Donation of catalytic residues to RNA polymerase active center by transcription factor Gre. *Proceedings of the National Academy of Sciences of the United States of America*. 100 (26), 15469–15474.

Stanier, R.Y., Kunisawa, R., Mandel, M. & Cohen-Bazire, G. (1971) Purification and Properties of Unicellular Blue-Green Algae (Order Chroococcales). *BACTERIOL. REV.* 3535.

Stebbins, C.E., Borukhov, S., Orlova, M., Polyakov, A., Goldfarb, A. & Darst, S.A. (1995) Crystal structure of the GreA transcript cleavage factor from *Escherichia coli*. *Nature*. 373 (6515), 636–640.

Steitz, T.A. (1998) Structural biology: A mechanism for all polymerases. *Nature*. 391 (6664), 231–232.

Steitz, T.A. & Steitz, J.A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proceedings of the National Academy of Sciences of the United States of America*. 90 (14), 6498–6502.

Stern, D.B. & Gruissem, W. (1987) Control of plastid gene expression: 3′ inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. *Cell*. 51 (6), 1145–1157.

Susan S. Golden, Judy Brusslan, Robert Haselkorn. Genetic engineering of the cyanobacterial chromosome. Methods in Enzymology,Academic Press,Volume 153,1987,Pages 215-231, ISSN 0076-6879, ISBN 9780121820541,

Summerfield, T.C. & Sherman, L.A. (2007) Role of Sigma Factors in Controlling Global Gene Expression in Light/Dark Transitions in the Cyanobacterium *Synechocystis sp.* Strain PCC 6803. *Journal of Bacteriology*. 189 (21), 7829–7840.

Sun, T., Li, S., Song, X., Diao, J., Chen, L. & Zhang, W. (2018) Toolboxes for cyanobacteria: Recent advances and future direction. *Biotechnology Advances*. 36 (4), 1293–1307.

Svetlov, V. & Artsimovitch, I. (2015) Purification of bacterial RNA polymerase: tools and protocols. *Methods in molecular biology (Clifton, N.J.)*. 127613–29.

Svetlov, V. & Nudler, E. (2013) Basic mechanism of transcription by RNA polymerase II. *Biochimica et biophysica acta*. 1829 (1), 20–28.

Svetlov, V., Vassylyev, D.G. & Artsimovitch, I. (2004) Discrimination against Deoxyribonucleotide Substrates by Bacterial RNA Polymerase. *The Journal of biological chemistry*. 279 (37), 38087–38090.

Sweetser, D., Nonet, M. & Young, R.A. (1987) Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proceedings of the National Academy of Sciences*. 84 (5), 1192–1196.

Symersky, J., Perederina, A., Vassylyeva, M.N., Svetlov, V., Artsimovitch, I. & Vassylyev, D.G. (2006) Regulation through the RNA Polymerase Secondary Channel. *The Journal of biological chemistry*. 281 (3), 1309–1312.

Tan, L., Wiesler, S., Trzaska, D., Carney, H.C. & Weinzierl, R.O. (2008) Bridge helix and trigger loop perturbations generate superactive RNA polymerases. *Journal of Biology*. 7 (10), 40.

Temiakov, D., Zenkin, N., Vassylyeva, M.N., Perederina, A., Tahirov, T.H., Kashkina, E., Savkina, M., Zorov, S., Nikiforov, V., Igarashi, N., Matsugaki, N., Wakatsuki, S., Severinov, K. & Vassylyev, D.G. (2005) Structural basis of transcription inhibition by antibiotic streptolydigin. *Molecular Cell*. 19 (5), 655–666.

Thomas, M.J., Platas, A.A. & Hawley, D.K. (1998) Transcriptional Fidelity and Proofreading by RNA Polymerase II. *Cell*. 93 (4), 627–637.

Timmis, J.N., Ayliffe, M.A., Huang, C.Y. & Martin, W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics*. 5 (2), 123–135.

Toulokhonov, I., Zhang, J., Palangat, M. & Landick, R. (2007) A Central Role of the RNA Polymerase Trigger Loop in Active-Site Rearrangement during Transcriptional Pausing. *Molecular Cell*. 27 (3), 406–419.

Troschl, C., Meixner, K. & Drosg, B. (2017) Cyanobacterial PHA Production—Review of Recent Advances and a Summary of Three Years' Working Experience Running a Pilot Plant. *Bioengineering*. 4 (2), .

Van Der Giezen, M. & Lenton, T.M. (2012) The rise of oxygen and complex life. *The Journal of Eukaryotic Microbiology*. 59 (2), 111–113.

Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S. & Yokoyama, S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature*. 417 (6890), 712–719.

Vassylyev, D.G., Vassylyeva, M.N., Perederina, A., Tahirov, T.H. & Artsimovitch, I. (2007a) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*. 448 (7150), 157–162.

Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I. & Landick, R. (2007b) Structural basis for substrate loading in bacterial RNA polymerase. *Nature*. 448 (7150), 163–168.

Vassylyeva, M.N., Svetlov, V., Dearborn, A.D., Klyuyev, S., Artsimovitch, I. & Vassylyev, D.G. (2007) The carboxy-terminal coiled-coil of the RNA polymerase β′-subunit is the main binding site for Gre factors. *EMBO Reports*. 8 (11), 1038–1043.

Vermass, W.F.J., Rutherford, A.W. & Hansson, Ö. (1988) Site-directed mutagenesis in photosystem II of the cyanobacterium *Synechocystis sp*. PCC 6803: Donor D is a tyrosine residue in the D2 protein. *Proceedings of the National Academy of Sciences of the United States of America*. 85 (22), 8477–8481.

Vijayan, V., Jain, I.H. & O'Shea, E.K. (2011) A high resolution map of a cyanobacterial transcriptome. *Genome Biology*. 12 (5), R47.

Vijayan, V., Zuzow, R. & O'Shea, E.K. (2009) Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proceedings of the National Academy of Sciences*. 106 (52), 22564–22568.

Vrentas, C.E., Gaal, T., Ross, W., Ebright, R.H. & Gourse, R.L. (2005) Response of RNA polymerase to ppGpp: requirement for the omega subunit and relief of this requirement by DksA. *Genes & Development*. 19 (19), 2378–2387.

Wang, B., Opron, K., Burton, Z.F., Cukier, R.I. & Feig, M. (2015) Five checkpoints maintaining the fidelity of transcription by RNA polymerases in structural and energetic details. *Nucleic Acids Research*. 43 (2), 1133–1146.

Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D. & Kornberg, R.D. (2006) Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell*. 127 (5), 941–954.

Weinzierl, R.O. (2010) The nucleotide addition cycle of RNA polymerase is controlled by two molecular hinges in the Bridge Helix domain. *BMC Biology*. 8 (1), 134.

Weinzierl, R.O.J. (2012) The Bridge Helix of RNA Polymerase Acts as a Central Nanomechanical Switchboard for Coordinating Catalysis and Substrate Movement. *Archaea*. 2011.

Weixlbaumer, A., Leon, K., Landick, R. & Darst, S.A. (2013) Structural basis of transcriptional pausing in bacteria. *Cell*. 152 (3), 431–441.

Westover, K.D., Bushnell, D.A. & Kornberg, R.D. (2004) Structural Basis of Transcription: Nucleotide Selection by Rotation in the RNA Polymerase II Active Center. *Cell*. 119 (4), 481–489.

Windgassen, T.A., Mooney, R.A., Nayak, D., Palangat, M., Zhang, J. & Landick, R. (2014) Trigger-helix folding pathway and SI3 mediate catalysis and hairpin-stabilized pausing by *Escherichia coli* RNA polymerase. *Nucleic Acids Research*. 42 (20), 12707–12721.

Wong, T.N., Sosnick, T.R. & Pan, T. (2007) Folding of noncoding RNAs during transcription facilitated by pausing-induced nonnative structures. *Proceedings of the National Academy of Sciences of the United States of America*. 104 (46), 17995–18000.

work(s):, J.W.S.R. (1993) Microfossils of the Early Archean Apex Chert: New Evidence of the Antiquity of Life. *Science, New Series*. 260 (5108), 640–646.

Xie, W.-Q. & Jager, K. (1989) Cyanobacterial RNA Polymerase Genes *rpoCl* and *rpoC2* Correspond to *rpoC* of *Escherichia coli*. *J. BACTERIOL.* 1717.

Yang, W., Lee, J.Y. & Nowotny, M. (2006) Making and Breaking Nucleic Acids: Two-Mg2+-Ion Catalysis and Substrate Specificity. *Molecular Cell*. 22 (1), 5–13.

Yee, D., Armstrong, V.W. & Eckstein, F. (1979) Mechanistic studies on deoxyribonucleic acid dependent ribonucleic acid polymerase from *Escherichia coli* using phosphorothioate analog. Initiation and pyrophosphate exchange reactions. *Biochemistry*. 18 (19), 4116–4120.

Yuzenkova, Y., Bochkareva, A., Tadigotla, V.R., Roghanian, M., Zorov, S., Severinov, K. & Zenkin, N. (2010) Stepwise mechanism for transcription fidelity. *BMC Biology*. 854.

Yuzenkova, Y., Gamba, P., Herber, M., Attaiech, L., Shafeeq, S., Kuipers, O.P., Klumpp, S., Zenkin, N. & Veening, J.-W. (2014) Control of transcription elongation by GreA determines rate of gene expression in *Streptococcus pneumoniae*. *Nucleic Acids Research*. 42 (17), 10987–10999.

Yuzenkova, Y., Roghanian, M., Bochkareva, A. & Zenkin, N. (2013) Tagetitoxin inhibits transcription by stabilizing pre-translocated state of the elongation complex. *Nucleic Acids Research*. 41 (20), 9257–9265.

Yuzenkova, Y. & Zenkin, N. (2010) Central role of the RNA polymerase trigger loop in intrinsic RNA hydrolysis. *Proceedings of the National Academy of Sciences*. 107 (24), 10878–10883.

Zakharova, N., Bass, I., Arsenieva, E., Nikiforov, V. & Severinov, K. (1998) Mutations in and Monoclonal Antibody Binding to Evolutionary Hypervariable Region of *Escherichia coli* RNA Polymerase β′ Subunit Inhibit Transcript Cleavage and Transcript Elongation. *Journal of Biological Chemistry*. 273 (38), 24912–24920.

Zaychikov, E., Martin, E., Denissova, L., Kozlov, M., Markovtsov, V., Kashlev, M., Heumann, H., Nikiforov, V., Goldfarb, A. & Mustaev, A. (1996) Mapping of Catalytic Residues in the RNA Polymerase Active Center. *Science*. 273 (5271), 107–109.

Zenkin, N., Yuzenkova, Y. & Severinov, K. (2006) Transcript-assisted transcriptional proofreading. *Science (New York, N.Y.)*. 313 (5786), 518–520.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K. & Darst, S.A. (1999) Crystal Structure of *Thermus aquaticus* Core RNA Polymerase at 3.3 Å Resolution. *Cell*. 98 (6), 811–824.

Zhang, J. & Landick, R. (2016) A two-way street: regulatory interplay between RNA polymerase and nascent RNA structure. *Trends in biochemical sciences*. 41 (4), 293–310.

Zhang, J., Palangat, M. & Landick, R. (2010a) Role of the RNA polymerase trigger loop in catalysis and pausing. *Nature Structural & Molecular Biology*. 17 (1), 99–104.

Zuo, Y. & Steitz, T.A. (2016) A structure-based kinetic model of transcription. *Transcription*. 8 (1), 1–8.