

# Novel Types of RNA Polymerases

---

David Andrew Forrest

**A dissertation submitted in partial fulfilment of the requirements for the degree of  
Doctor of Philosophy**

Centre for Bacterial Cell Biology

Institute for Cell and Molecular Biosciences

Medical School

Newcastle University

September 2014

## Abstract

Transcription, the first step of gene expression, is accomplished in all domain of life by the multisubunit RNA polymerase (RNAP). Accordingly, the RNAP is an ancient enzyme that is highly conserved in all cellular organisms. However, in-depth bioinformatics has led to the identification of proteins distantly related to the multisubunit RNAP, such as YonO and ORF6 RNAP. Whilst being putative single subunit RNAPs, YonO and ORF6 RNAP are completely unrelated to T7 RNAP. YonO, encoded by the *Bacillus subtilis* SP $\beta$  prophage, has incredibly little similarity to RNAP. Conversely, ORF6 RNAP belonging to the *Kluyveromyces lactis* killer system contains half of the conserved RNAP active centre. YonO and ORF6 RNAP are potentially novel RNAPs and their characterisation will give new insights into mechanisms of transcription as well as the biology of the organisms which they belong to.

Despite lacking most of the conserved RNAP active centre, we have shown YonO is a very efficient DNA dependant RNAP. Striking, unlike all of its multisubunit relatives, YonO is able to initiate transcription on double stranded DNA without accessory factors (such as  $\sigma$  factors in bacteria). Furthermore, our work suggests YonO is expressed during induction of SP $\beta$  and functions to transcribe the SP $\beta$  late genes. This potentially makes YonO the first bacteriophage single subunit RNAP that resembles the multisubunit RNAP. On the other hand, ORF6 RNAP was very poor at polymerisation, with or without its putative accessory subunit. This suggests additional, currently unknown proteins are required for RNAP activity and potentially a new molecular mechanism of RNA polymerisation.

YonO homologues exist in a wide range of bacteria including firmicutes and cyanobacteria. YonO represents a new class of RNAP and our discoveries potentially open up a new area of research as well as being potentially useful for biotechnology and synthetic biology. In contrast, transcription by ORF6 RNAP is more complex than previously thought, with alternative lines of investigation required to identify additional factors that contribute to ORF6 RNAP activity.

## Acknowledgements

Firstly I would like to thank Professor Nikolay Zenkin for giving me this opportunity, for his advice, his patience and for his great scientific knowledge. I'd also like to thank the members of the Zenkin group (past and present), in particular Yulia Yuzenkova, Mohammad Roghanian, Aleksandra Bochkareva and Daniel Castro-Roa for all of their help. I'd also like to acknowledge all the members of The Centre for Bacterial Cell Biology who have shared their materials and methods with me.

Outside of science, I would like to thank the Lewis household for being so welcoming and for opening up their home every week. I'd also like to thank the Cooling family for their support over the years and for making me feel like one of their own. Finally, I'd like to thank my parents, Sue and Stephen for their love, support and encouragement.

## Contents

Abstract.....	i
Acknowledgements.....	ii
List of Figures .....	1
List of Tables.....	3
List of Abbreviations .....	3
Chapter 1: Transcription and the Multisubunit RNA Polymerase .....	6
1.1: Structure of RNAP .....	6
1.1.1: Bacterial RNAP .....	8
1.1.2: Nucleic acids within RNAP .....	10
1.1.3: Organisation of the Active centre .....	11
1.1.4: Mechanism of Catalysis.....	13
1.1.5: Double $\psi$ $\beta$ Barrel Domains .....	15
1.2: Single subunit RNA polymerases .....	17
1.3: The Stages of Transcription.....	17
1.4: Initiation of Transcription .....	17
1.5: Transcription Elongation .....	21
1.5.1: Nucleotide addition cycle .....	21
1.5.2: Pausing and Backtracking.....	22
1.5.3: Hydrolysis and proof reading .....	24
1.6: Transcription Termination .....	25
1.7: Evolution of multisubunit RNA Polymerases .....	27
Chapter 2: YonO: the hypothetical RNA polymerase of the <i>Bacillus subtilis</i> prophage SP $\beta$ .....	29
2.1. Resident prophages of <i>Bacillus subtilis</i> .....	29
2.2: SP $\beta$ prophage of <i>Bacillus subtilis</i> .....	31
2.3. YonO, a putative RNA polymerase of SP $\beta$ .....	34

Chapter 3: ORF6 RNAP of the <i>Kluyveromyces lactis</i> Killer System .....	41
3.1. Killer system of <i>Kluyveromyces lactis</i> .....	41
3.2. Linear Plasmids of the <i>K. lactis</i> Killer System.....	41
3.2.1: The pGKL1 Killer plasmid.....	42
3.2.2: The pGKL2 Killer plasmid.....	44
3.3: ORF6 RNAP .....	47
Chapter 4: Materials and Methods .....	51
4.1. General Methods .....	51
4.1.1. PCR .....	51
4.1.2. Growth media and Strains .....	51
4.1.2. Molecular Cloning .....	56
4.1.3. Buffers .....	57
4.2. Protein Methods .....	61
4.2.1. Purification of YonO .....	61
4.2.2. Purification of ORF6 RNAP by Ni <sup>2+</sup> NTA chromatography.....	62
4.2.3. Co-expression and purification of ORF6 RNAP and ORF7p in <i>S. cerevisiae</i> ..	63
4.2.4. Purification of ORF7p .....	64
4.2.5. Co-expression and purification of ORF6 RNAP and ORF7p in <i>E. coli</i> .....	64
4.2.6. SDS PAGE.....	65
4.2.7. Silver stain of SDS PAGE gel .....	65
4.2.8. Western Blot .....	66
4.2.7. Affinity purification of $\alpha$ -YonO Antibody from Serum.....	66
4.2.9. Electron Microscopy .....	67
4.3. <i>Bacillus subtilis</i> Methods.....	68
4.3.1. Transformation of <i>B. subtilis</i> .....	68
4.3.2. Isolation of Genomic DNA from <i>B. subtilis</i> .....	68
4.3.3. Construction of a <i>B. subtilis</i> $\Delta$ yonO strain.....	68

4.3.4. Induction of SP $\beta$ prophage from <i>Bacillus subtilis</i> .....	69
4.3.5. <i>B. subtilis</i> lysates for western blot .....	69
4.3.6. Isolation of total RNA from <i>B. subtilis</i> for RNAseq .....	69
4.3.7. Isolation of SP $\beta$ particles .....	70
4.4. X-ray Crystallography Methods.....	70
4.4.1 Initial Screening for Crystallisation Conditions .....	70
4.4.2. Determination of Crystal Quality .....	71
4.4.3. Optimisation of Crystallisation Conditions .....	71
4.4.4. Gel filtration calibration .....	71
4.4.5. Post crystallisation Optimisation of Crystal Diffraction .....	72
4.5. <i>In vitro</i> Transcription Methods .....	73
4.5.1. Assembled Elongation Complexes .....	73
4.5.2. 5' <sup>32</sup> P labelling of oligonucleotides.....	75
4.5.3. Transcription Initiation.....	75
4.5.4. Primer Extension .....	76
4.5.6. Potassium Permanganate DNA Footprinting.....	77
4.5.7. DNase I Footprinting .....	77
4.5.8. Sanger Sequencing of DNA .....	77
Chapter 5: Preliminary Characterisation of YonO, a Putative RNA polymerase of SP $\beta$ prophage .....	79
5.1. Introduction .....	79
5.2. Aims.....	79
5.3. Expression and purification of YonO.....	79
5.4. <i>In vitro</i> characterisation of YonO .....	82
5.4.1. Artificially Assembled Elongation Complexes .....	82
5.4.2. RNA Polymerase activity of YonO elongation complexes.....	83
5.4.3. YonO is a strict DNA Dependant RNA Polymerase .....	86

5.4.4. Determination of Optimum Hybrid Length.....	89
5.4.5. Rate of Transcription.....	91
5.4.6. Pyrophosphorolysis activity of YonO .....	94
5.4.7. Hydrolysis by YonO .....	95
5.4.8. Misincorporation by YonO .....	96
5.5. Discussion.....	99
Chapter 6: Investigation of Biological Function of YonO .....	104
6.1. Introduction .....	104
6.2. Aim .....	106
6.3. Expression of YonO during SP $\beta$ induction.....	106
6.4. Requirement of YonO for SP $\beta$ development .....	108
6.5. Identification of Genes Transcribed by YonO .....	109
6.6. Investigation of Transcription Initiation by YonO .....	115
6.6.1. Transcription initiation at the yonI predicted start site.....	115
6.6.2. Primer extension to confirm transcription initiation at the predicted yonI start site .....	118
6.6.3. KMnO <sub>4</sub> footprinting to observe DNA melting at the predicted yonI transcription start site.....	121
6.6.4. DNase I footprint to observe YonO closed promoter complex formation .	122
6.7. Discussion.....	124
Chapter 7: Attempts to Determine the Structure of YonO.....	129
7.1. Introduction .....	129
7.2. Aims.....	129
7.3. Screening for Crystallisation Conditions .....	130
7.4. Optimisation of YonO Crystallisation .....	131
7.4.1. Optimisation of YonO screening condition .....	131
7.4.2. Discovery and crystallisation of a possible YonO tetramer .....	133

7.5. Optimisation of YonO Crystal Diffraction.....	136
7.5.1. Microseeding to Produce Alternative Crystal Morphologies .....	136
7.5.2. Additive Screening .....	138
7.5.3. Removal of N-terminal His tag.....	139
7.5.4. Use of Artificial Transcription Scaffolds to Stabilise YonO RNAP .....	141
7.5.5. Lysine methylation .....	146
7.5.6. Twinned Data Set and Molecular Replacement .....	146
7.6. Dehydration of YonO crystals to Reproduce High Resolution Data Set .....	150
7.6.1. Dehydration by HC1 Dehumidifier .....	151
7.6.2. Dehydration through vapour diffusion and <i>In situ</i> data collection .....	154
7.7. Crystallisation of the YonO homologue VM RNAP.....	155
7.8. SAXS.....	156
7.9. Discussion.....	157
Chapter 8: ORF6 RNA Polymerase of the <i>Kluyveromyces lactis</i> Killer System.....	161
8.1: Introduction .....	161
8.2. Aim .....	162
8.3. Previous Attempts to Express ORF6 RNAP .....	163
8.4: Expression of Codon Optimised ORF6 in <i>Saccharomyces cerevisiae</i> .....	164
8.5. <i>In vitro</i> characterisation of ORF6 activity.....	166
8.5.1. Initial determination of ORF6 RNAP activity.....	166
8.5.2. Alternative nucleic acid scaffolds for in vitro transcription.....	167
8.5.3. Immobilisation of ORF6 RNAP Elongation complexes .....	170
8.5.4. ORF7p and the Trigger Loop .....	173
8.6. Co-expression of ORF6 RNAP and ORF7p in <i>S. cerevisiae</i> .....	175
8.7. Co-expression of ORF6 RNAP and ORF7p in <i>E. coli</i> .....	177
8.8. Discussion.....	179
Chapter 9: Concluding Discussion .....	184

References.....187

## List of Figures

Figure 1-1: Conservation of the multisubunit RNAP

Figure 1-2: Structure and layout of bacterial RNA polymerase

Figure 1-3: The active centre of RNAP

Figure 1-4: Mechanism of RNAP catalysis

Figure 1-5: The two double  $\Psi$   $\beta$  barrel (DPBB) folds of RNAP

Figure 1-6: Bacterial transcription initiation facilitated by  $\sigma^{70}$

Figure 1-7: The Nucleotide addition cycle

Figure 2-1: Life cycle of bacteriophages

Figure 2-2: The Distant relationship between YonO, rdRNAP and the multisubunit RNAP

Figure 2-3: Proposed Origins of YonO and ORF6 RNAP

Figure 3-1: Organisation of pGKL1 and pGKL2 plasmids of the *K. lactis* killer system

Figure 3-2: The conservation of multisubunit RNAP functional motifs in ORF6 RNAP

Figure 4-1: Calibration of Superdex 200 gel filtration column

Figure 5.1: Purification of YonO

Figure 5-2: Artificially Assembled Elongation Complexes

Figure 5-3: Determination of YonO RNAP activity

Figure 5-4: Investigating the RNA dependant RNAP and DNA polymerase activity of YonO

Figure 5-5: Influence of hybrid length on YonO elongation complexes

Figure 5-6: YonO is a fast and processive RNAP

Figure 5-7: Pyrophosphorolysis by YonO elongation complexes

Figure 5-8: Determination of YonO hydrolysis activity

Figure 5-9: Misincorporation by YonO elongation complexes

Figure 6-1: Expression of YonO during SP $\beta$  induction

Figure 6-2: YonO is essential for SP $\beta$  production *in vivo*

Figure 6-3: Fold changes of expression across the *B. subtilis* genome in response to Mitomycin C

Figure 6-4: Transcription profile of *yonO* and adjacent regions of SP $\beta$

Figure 6-5: Alignment of RNAseq reads to SP $\beta$  Cluster II DNA

Figure 6-6: Transcription initiation by YonO on upshift 1673 DNA

Figure 6-7: Orientation of transcription on the upshift 1673 template DNA

Figure 6-8: Primer extension to determine transcription start sites on 1673 template

Figure 6-9: DNA footprinting of YonO on upshift 1673 DNA

Figure 7-1: Optimisation of YonO crystallisation.

Figure 7-2: Purification and crystallisation of YonO tetramer

Figure 7-3: Microseeding of crystallisation screens

Figure 7-4: Additive screening for improved crystal diffraction

Figure 7-4: Additive screening for improved crystal diffraction

Figure 7-5: Crystallisation of tag free YonO

Figure 7-6: Open and closed conformation of QDE-1

Figure 7-7: Determination of optimum nucleic acid scaffold for YonO elongation complex crystallisation

Figure 7-8: High resolution diffraction from a twinned YonO crystal.

Figure 7-9: Effect of dehydration on crystal packing

Figure 7-10: SAXS analysis of YonO

Figure 8-1: The RNA polymerisation activity of ORF6 RNAP purified from *E. coli*

Figure 8-2: Ni<sup>2+</sup> - NTA Purification of ORF6 RNAP expressed from ORF6<sup>C+</sup> in *S. cerevisiae*

Figure 8-3: Activity of ORF6 RNAP purified from *S. cerevisiae*

Figure 8-4: Alternative scaffolds for the assembly of ORF6 RNAP elongation complexes

Figure 8-5: Stability of ORF6 RNAP elongation complexes

Figure 8-6: ORF7p does not increase the catalytic activity of ORF6 RNAP

Figure 8-7: FLAG – tag affinity purification of ORF6 RNAP expressed by *S. cerevisiae*

Figure 8-8: Co-purification of ORF6 RNAP and ORF7p expressed in *E. coli*.

### List of Tables

Table 4-1: Primers used in this study.

Table 4-2: Strains constructed and used in this study

Table 4-3: Composition of growth media

Table 4-4: Plasmids used and constructed through this work

Table 4-5: Composition of buffers

Table 4-6: Composition of Commercial buffers

Table 7-1: Data statistics for YonO data set 1 and 2

### List of Abbreviations

ATP – adenosine triphosphate

Bp – basepair

CNBr – cyanogen bromide

CTP – cytidine triphosphate

ddNTP – dideoxynucleotide

DLS – diamond light source

DNAP – DNA polymerase

dNMP – deoxynucleoside monophosphate

dNTP – deoxyribonucleoside triphosphate

DPBB – double psi beta barrell

DTT – Dithiothreitol

EC – elongation complex

EDTA – ethylenediaminetetraacetic acid

ETF – early transcription factor

FPLC – fast protein liquid chromatography

GFP – green fluorescent protein

GTP – guanoside triphosphate

ICE*bs1* – integrative conjugative element

IPTG - isopropyl  $\beta$ -D-1-thiogalactopyranoside

Kb – kilobase

KDa – kilodalton

LB – luria broth

mEC – mismatch elongation complex

MMB – modified m broth

NAC – nucleotide addition cycle

NMP – nucleoside monophosphate

NPH I – nucleoside triphosphate phosphohydrolase I

nt – nucleotide

NTP – nucleoside triphosphate

OD<sub>600</sub> – optical density at a wavelength of 600 nm

ORF – open reading frame

PAGE – polyacrylamide gel electrophoresis

PCR – polymerase chain reaction

PMF – peptide mass fingerprinting

PMSF - phenylmethylsulfonyl fluoride

PNK – polynucleotide kinase

PPi – pyrophosphate

rdRNAP – RNA dependent RNA polymerase

RNAP – multisubunit RNA polymerase

Rpm – revolutions per minute

SC - URA– Synthetic Complete - uracil

SDS PAGE – sodium dodecyl sulphate polyacrylamide gel electrophoresis

SKIN – *sigk* intervening element

SPBRE – SPβ repeated element

TIR – terminal inverted repeat

TP – terminal protein

UCS – upstream conserved sequence

UTP – uracil triphosphate

WT - wildtype

YPD – yeast peptone dextrose

## Chapter 1: Transcription and the Multisubunit RNA Polymerase

The first step of gene expression is the converting of the genetic information encoded in a DNA template into RNA. This template dependant synthesis of RNA is referred to as transcription. In all cellular organisms, the multisubunit DNA dependant RNA polymerase (RNAP) accomplishes transcription. RNAP synthesises RNA according to the sequence of template DNA using ribonucleoside triphosphates (NTPs) as substrates. The structure of RNAP and its mechanism of catalysis are highly conserved across all three domains of life. Such conservation implies an ancient, singular origin for this crucial enzyme.

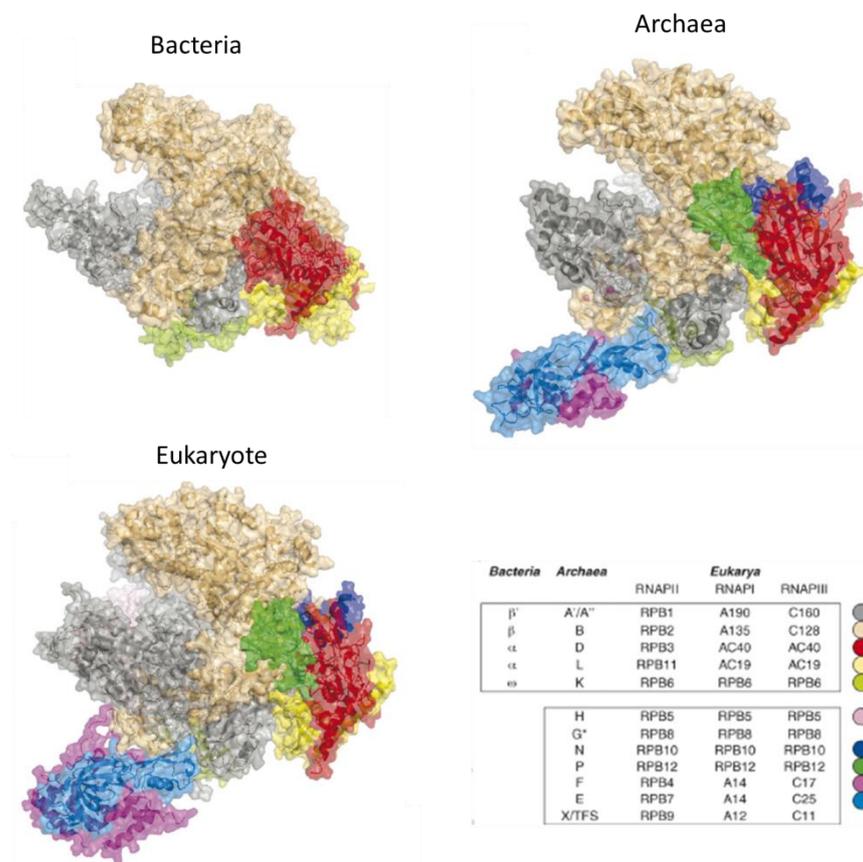
With the advancement of genome sequencing and bioinformatics, hypothetical proteins with distant and unusual homology to RNAP have been identified. Interestingly, these proteins appear to be related to multisubunit RNAPs, despite being single polypeptide chains. No homology with known single subunit RNAPs has been detected. These hypothetical proteins may be representatives of a novel class of RNAP with previously undescribed structures and mechanisms of catalysis.

This introduction aims to briefly describe the structure of multisubunit RNAP, the stages of transcription and the mechanism of catalysis. Subsequently, YonO and ORF6 RNAP, two examples of single proteins distantly related to RNAP, will be introduced and discussed.

### **1.1: Structure of RNAP**

Bacteria utilise a single RNAP comprised of 4 subunits:  $\beta'$ ,  $\beta$ ,  $\alpha_2$  and  $\omega$  (Figure 1-2a). This is in contrast to eukaryotes in which three RNAPs (RNAP I, RNAP II and RNAP III) accomplish transcription. Eukaryotic RNAPs are comprised of an increased number of subunits. RNAP I, II and III are formed by 14, 12 and 17 subunits, respectively. Over the past decade, it has become apparent that plants utilise two additional RNAP: RNAP VI and V (Reviewed in Haag and Pikaard (2011)). Both of these RNAP, whilst residing in the nucleus, do not function in canonical gene expression. Rather, they function in DNA methylation and gene silencing (Haag et al., 2012). Mass spectroscopy showed that whilst they are distinct enzymes, RNAP VI and V are clearly evolved from RNAP II (Ream et al., 2009). Like bacteria, archaea utilise a single RNAP. However this RNAP is

comprised of 12 subunits, resembling RNAP II. Despite such differences, a high degree of conservation exists in RNAP across all domains of life. The bacterial RNAP represents a minimal set of subunits that is conserved in eukaryotic and archaeal RNAPs (Figure 1-1). Eukaryotic and archaeal homologues for each of the five bacterial subunits have been identified. In addition to structural conservation, all RNAPs synthesise RNA by the same mechanism of catalysis. Therefore, the bacterial RNAP can be exploited as a model for the investigation into the key principles and processes of transcription. Unless stated, the remainder of this introduction will describe bacterial RNAP.

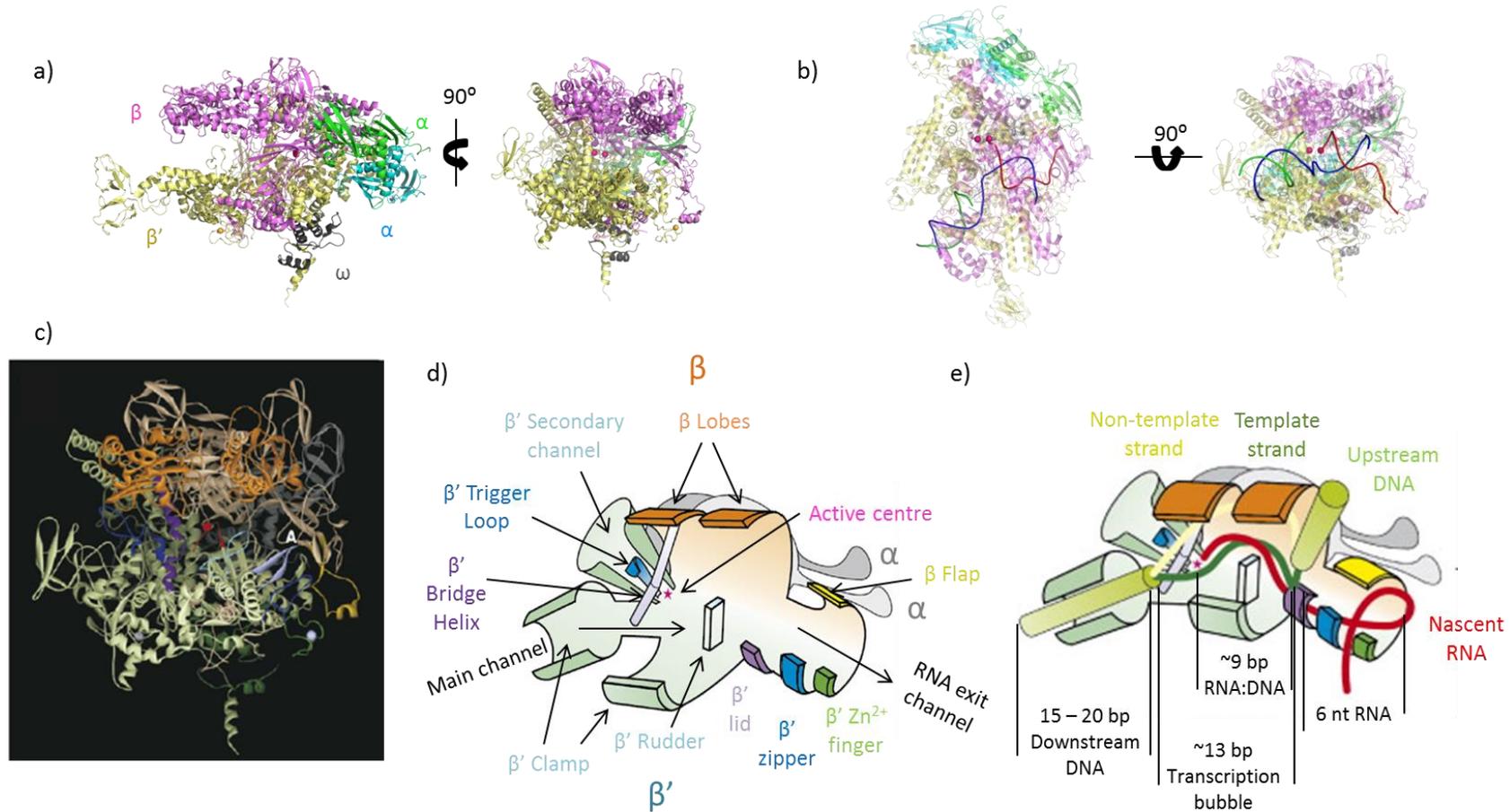


**Figure 1-1: Conservation of the multisubunit RNAP. Taken from Werner (2008).** The crystal structures of bacterial, archaeal and eukaryotic RNAP (RNAP II, *S. cerevisiae*) are shown to demonstrate conservation of the structure and core subunits. The bacterial RNAP represents a minimal set of subunits which is conserved in all domains of life, despite increased subunits in the archaeal and eukaryotic RNAPs. The legend shows the colouring scheme of the subunits.

### 1.1.1: Bacterial RNAP

Bacterial RNAP has a molecular weight of approximately 400 KDa (Zhang et al., 1999). The two largest subunits,  $\beta'$  (170 KDa) and  $\beta$  (124 KDa) form the bulk of the core enzyme.  $\beta'$  and  $\beta$  subunits form a crab claw structure, with each subunit representing a pincer of the claw. Between the two subunits is the main channel in which downstream duplex DNA enters the enzyme (Figure 1-2b). On the back wall of the main channel, towards the centre of the enzyme resides the active centre which is formed by amino acid residues of both  $\beta'$  and  $\beta$  subunits. The active centre is where catalysis occurs and is discussed in section 1.1.4.

The two  $\alpha$  subunits (35 KDa) interact by their N-terminal domains to form a homodimer ( $\alpha_2$ ) that is found towards the rear of the enzyme, behind the  $\beta'$  and  $\beta$  crab claw. The function of the  $\alpha_2$  dimer is to act as a platform for the assembly of  $\beta'$  and  $\beta$  (Zhang et al., 1999). The  $\omega$  subunit also functions in the assembly of RNAP by bringing together the C-terminal tail and N-terminal domain of  $\beta'$ . This has been suggested to increase the stability of the  $\beta'$  subunit and encourage its assembly into RNAP (Minakhin et al., 2001).



**Figure 1-2: Structure and layout of bacterial RNA polymerase. a) Subunit composition of bacterial RNAP.** *Thermus thermophilus* core RNAP structure (PDB accession number 2PPB. Modelled with pymol). The catalytic  $Mg^{2+}$  ions are coloured magenta. Structural  $Zn^{2+}$  ions are coloured yellow. Individual subunits are labelled and coloured. Image on the left is the profile view of RNAP and is rotated by  $90^\circ$  to give the front view. **b) Entry of DNA into RNAP.** Colouring and structure is as before. RNA, Template and non-template DNA strands are coloured red, blue and green, respectively. RNAP is transparent to emphasise the path of DNA. Image on the left is a top down view to reveal  $90^\circ$  kink in template DNA. **c), d) and e) are**

modified from Burukhov and Nudler, 2008. **c) Structure of *Thermus aquaticus* core RNAP, taken from Zhang et al., 1999.** Front view of core RNAP. Coloured based on functional motifs described in the text and shown in Figure 1d. **d) Schematic representation of the *Thermus aquaticus* core RNAP structure.** Schematic based on the structure presented in Figure 1c. The channels and functional motifs contacting nucleic acid chains, as discussed in the text, are indicated. The coloured motifs match those of Figure 1c. **e) The organisation of nucleic acid chains within RNAP.** Using the same schematic as Figure 1-2d, the pathway of DNA and RNA through the main channel and RNA exit channel is shown, along with approximations of nucleic acid lengths (bp – basepairs, nt – nucleotides).

### 1.1.2: Nucleic acids within RNAP

Double stranded DNA enters RNAP through the main channel, which accommodates 12 base pairs (bp) of the downstream DNA (Figure 1-2b). As the downstream duplex DNA enters RNAP, it is unwound into the template and non-template strands immediately before the active centre (Gnatt et al., 2001, Vassylyev et al., 2007a). The template DNA strand passes through the active centre, where it directs the synthesis of complementary RNA (Figure 1-2b) (Gnatt et al., 2001). The two DNA strands reform a duplex 11 nucleotides upstream from the active centre. The region of melted DNA is referred to as the transcription bubble. A transcription bubble of 13 nucleotides is maintained throughout transcription (Figure 1e). During transcription, RNAP can accommodate a 8-9 bp RNA:DNA hybrid (Bochkareva et al., 2012, Sidorenkov et al., 1998, Vassylyev et al., 2007a, Nudler et al., 1997). Upon reaching a length greater than 9 nucleotides, the 5' nascent RNA is threaded through the RNA exit channel (Figure 1.2e).

In addition to the main channel and the RNA exit channel, a third channel exists in RNAP. The secondary channel is a 12 Å wide channel which runs from the surface of the enzyme to the active centre (Gnatt et al., 2001). The opening of the funnel in the active centre is referred to as the pore. It is thought that NTPs diffuse through the secondary channel to reach the active centre where cognate NTPs are selected for and incorporated into the 3' end of elongating RNA (Batada et al., 2004, Westover et al., 2004). However, contradictory data exists which suggests substrate NTPs arrive into the active centre through the main channel (Gong et al., 2005).

Many contacts are made between RNAP and the nucleic acids running through it (Figure 1-2d). Referring back to the crab claw analogy, the majority of the  $\beta'$  'pincer' forms the clamp domain (Cramer et al., 2001). The clamp is a large mobile domain that swings on a hinge (called the switch region) during transcription initiation to narrow the width of the main channel by 8 Å, thereby increasing RNAPs hold on downstream DNA (Chakraborty et al., 2012). The clamp's grip on the downstream DNA accounts for the high stability and processivity of RNAP during elongation. Within the clamp, smaller functional motifs are present. The rudder increases the stability of the elongating RNAP whilst the lid and zipper are loops which function in maintenance of hybrid length and RNA displacement from the template DNA (Gnatt et al., 2001, Kuznedelov et al., 2002, Vassylyev et al., 2007a, Zenkin et al., 2006a). Furthermore, the zipper may also interact with upstream DNA (Yuzenkova et al., 2011) .

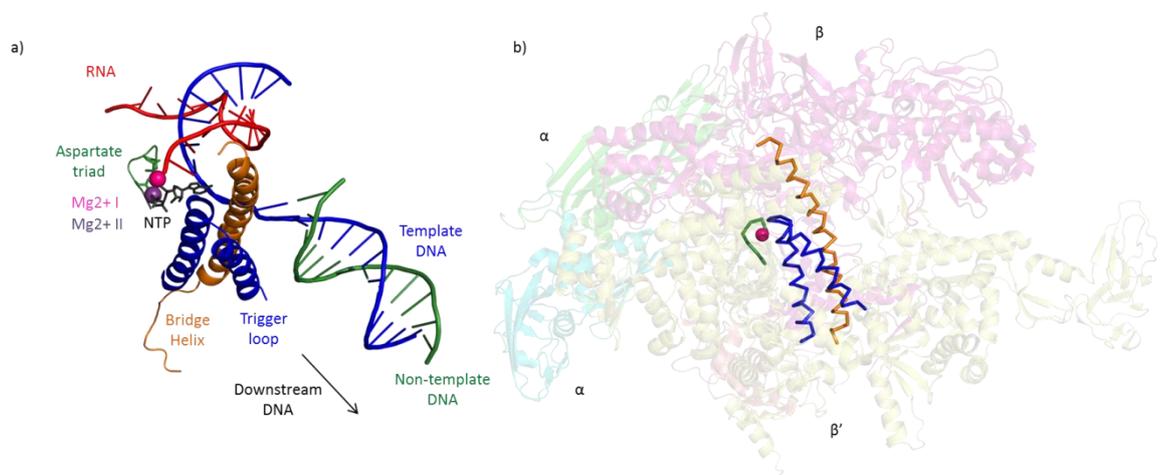
In addition to  $\beta'$ , the  $\beta$  subunit also makes contacts with nucleic acid chains. Approximately half of the RNA exit channel wall is formed by the  $\beta$  flap domain, with the  $\beta'$  zipper, lid and  $Zn^{2+}$  finger also contributing to the formation of the RNA exit channel. The  $\beta$  lobe domains are responsible for contacts with the non-template strand DNA (Korzheva et al., 2000). Typically, upstream non-template DNA is not included in nucleic acid scaffolds used for the crystallisation of RNAP. However, even when present, it is not visible in electron density suggesting a flexible, mobile conformation (Andrecka et al., 2009, Kettenberger et al., 2004). Cross linking and single molecule studies revealed that the non-template DNA passes through the polymerase and makes contact with  $\beta$  lobes 1 and 2 on the roof of the main channel (Korzheva et al., 2000, Andrecka et al., 2009).

### *1.1.3: Organisation of the Active centre*

At the heart of the RNAP active centre are two  $Mg^{2+}$  ions that are essential for catalysis (Sosunov et al., 2003). The first  $Mg^{2+}$  ion, Mg I, is permanently bound in the active centre. (Sosunov et al., 2003, Sosunov et al., 2005, Zaychikov et al., 1996). The aspartate triad motif, NADFDGD, is responsible for the strong chelation of Mg I. The aspartate triad motif is absolutely conserved in all RNAP and substitution of the aspartate residues results in abolishment of catalytic activity (Sosunov et al., 2005, Zaychikov et al., 1996). The second  $Mg^{2+}$  ion, Mg II is brought to the active centre

bound to each incoming substrate during elongation and is thus weakly bound in the active centre (Sosunov et al., 2003).

Two mobile domains contribute to the formation of the active centre (Figure 1-3). Spanning across the active centre, from the  $\beta'$  subunit to  $\beta$ , is the  $\beta'$  bridge helix (Zhang et al., 1999). In close proximity to the bridge helix is the  $\beta'$  trigger loop. Both the trigger loop and the bridge helix play key roles during catalysis by RNAP (Bar-Nahum et al., 2005, Wang et al., 2006, Yuzenkova et al., 2010, Zhang et al., 2010).



**Figure 1-3: The active centre of RNAP.** Made using the crystal structure of *Thermus thermophilus* RNAP elongation complex (PDB accession code: 2O5J. From Vassylyev et al., 2007). **a) Organisation of catalytic motifs of the active centre.** The prominent domains that function during catalysis are shown. The mobile domains, trigger loop (blue) and bridge helix (orange) are formed by residues of the  $\beta'$  domain. The trigger loop exists in two states, unfolded and folded. Here, the trigger loop is captured in a folded state as it interacts with the phosphate moieties of the incoming NTP (grey).  $Mg^{2+}$  II (purple) is chelated by the NTP whereas the permanent  $Mg^{2+}$  I is bound by the aspartate triad (green). The NTP is base paired with the template DNA (blue) and positioned next to the 3' NTP of the growing RNA chain (red). **b) Localisation of the active centre within RNAP.** The location of the bridge helix, trigger loop and aspartate triad in the RNAP structure is shown. The bridge helix spans across the active centre from the  $\beta'$  subunit (transparent yellow) to  $\beta$  (transparent pink). The aspartate triad is present at the centre of the enzyme, in close proximity to the catalytic residues of the trigger loop.

#### 1.1.4: Mechanism of Catalysis

The primary function of RNAP is to synthesise RNA in a template dependant manner. It does so through catalysing phosphodiester bond formation between the 3' OH of the growing RNA and the  $\alpha$  phosphate of the substrate NTP. A pyrophosphate molecule is released as a byproduct of phosphodiester bond formation. RNAP accomplishes this reaction using  $Mg^{2+}$  ions in a general two ion mechanism utilised by all nucleic acid polymerases (Sosunov et al., 2003, Steitz, 1998).

In the general two ion mechanism, Mg I and II orient the NTP substrate to the optimal position to trigger a nucleophilic attack on the  $\alpha$  phosphate of the substrate NTP by the 3' OH of the RNA. Specifically, Mg I activates the 3' OH group of RNA whilst coordinating the substrate NTP  $\alpha$  phosphate. Mg II orients all three NTP substrate phosphates into the optimal spatial position for the nucleophilic attack. The catalytic mechanism and amino acid residues involved are shown in Figure 1-4a. Together, Mg I and II stabilise the pentacovalent transition state, thus contributing to the catalysis of phosphodiester bond formation (Sosunov et al., 2003, Steitz, 1998, Steitz and Steitz, 1993).

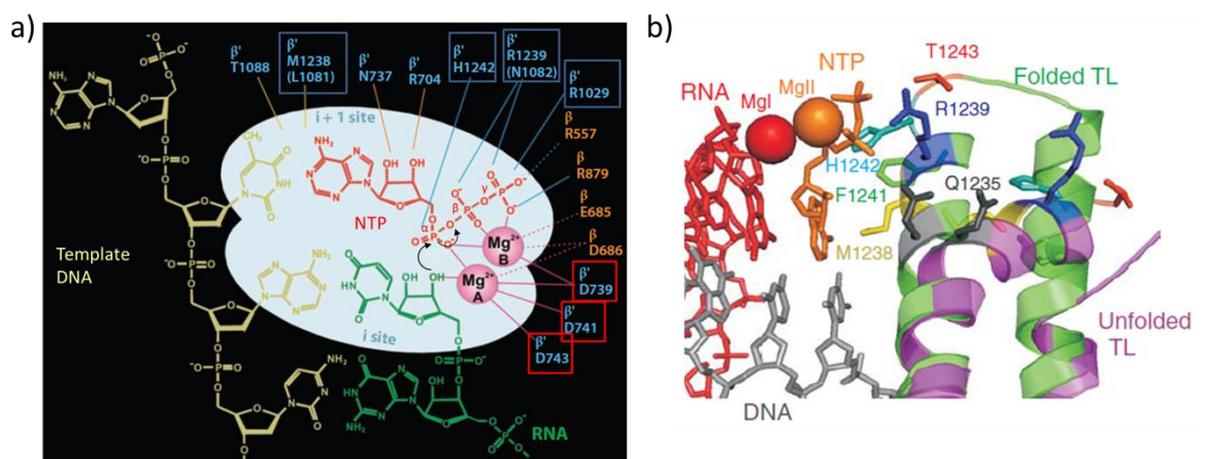
The trigger loop plays a crucial role in catalysis. Its deletion or impediment drastically diminishes any catalytic activity of RNAP (Temiakov et al., 2005, Yuzenkova et al., 2010) Structural studies show the trigger loop existing in an open and closed conformation (Figure 1-4b) (Vassilyev et al., 2007b, Wang et al., 2006). During catalysis, the trigger loop adopts a folded conformation upon binding of the correct NTP substrate. This is stabilised by interactions between the trigger loop M1238 residue and the base of the incoming NTP (Yuzenkova et al., 2010). Due to incorrect hydrogen bonding with the  $i+1$  template base, non-cognate substrates are not correctly positioned within the active centre. It has been proposed this prevents full folding of the trigger loop due to a steric clash between the M1238 of the trigger loop and the base of the non-cognate substrate. Since a closed off active centre is required for efficient catalysis, the trigger loop contributes to the fidelity of transcription via an induced fit mechanism (Yuzenkova and Zenkin, 2010).

Folding of the trigger loop positions its catalytic residues into the active centre, bringing them in close proximity with the Mg I, Mg II and the substrate NTP. The H1242 and R1239 residues (*T. aquaticus* numbering) are central to the trigger loops

participation in catalysis. H1242 and R1239 contribute to catalysis by coordinating the triphosphate moieties of the substrate NTP to a position optimal for transition state stabilisation (Yuzenkova et al., 2010).

In addition to phosphodiester bond formation, RNAP is able to catalyse pyrophosphorolysis and phosphodiester bond hydrolysis through a general metal ion mechanism. Pyrophosphorolysis is the direct reverse reaction of phosphodiester bond formation and results in the removal of the 3' NMP of the RNA chain. In this reaction, Mg II is bound by the pyrophosphate, which carries out the nucleophilic attack on the phosphodiester bond between the final and penultimate NMP of the RNA chain. The role of Mg II is reversed for pyrophosphorolysis and hydrolysis, as this ion rather than Mg I activates the attacking nucleophile (Sosunov et al., 2003).

As with phosphodiester bond synthesis, the trigger loop also plays a crucial role in phosphodiester bond hydrolysis. During RNAP hydrolysis, the 3' NMP of the RNA chain contributes to catalysing the hydrolysis of the penultimate phosphodiester bond by reorienting in the active centre to chelate Mg II (discussed further in section 1.4.3) (Zenkin et al., 2006b). The trigger loop functions in hydrolysis by positioning the 3' NMP into the active centre. As with polymerisation, H1242 of the trigger loop is central to hydrolysis, with the residue being responsible for the repositioning of the 3' NMP (Yuzenkova and Zenkin, 2010).



**Figure 1-4: Mechanism of RNAP catalysis. a) Catalysis of phosphodiester bond formation.** Adapted from Nudler, 2009. The active centre is shown in schematic representation. *T. thermophilus* residue numbering is used, with β' subunit residues in blue and β in orange. The aspartate triad residues (highlighted by a red box) permanently bind Mg<sup>2+</sup>I (shown as Mg<sup>2+</sup> A in the Figure) in the active centre. Mg<sup>2+</sup> II

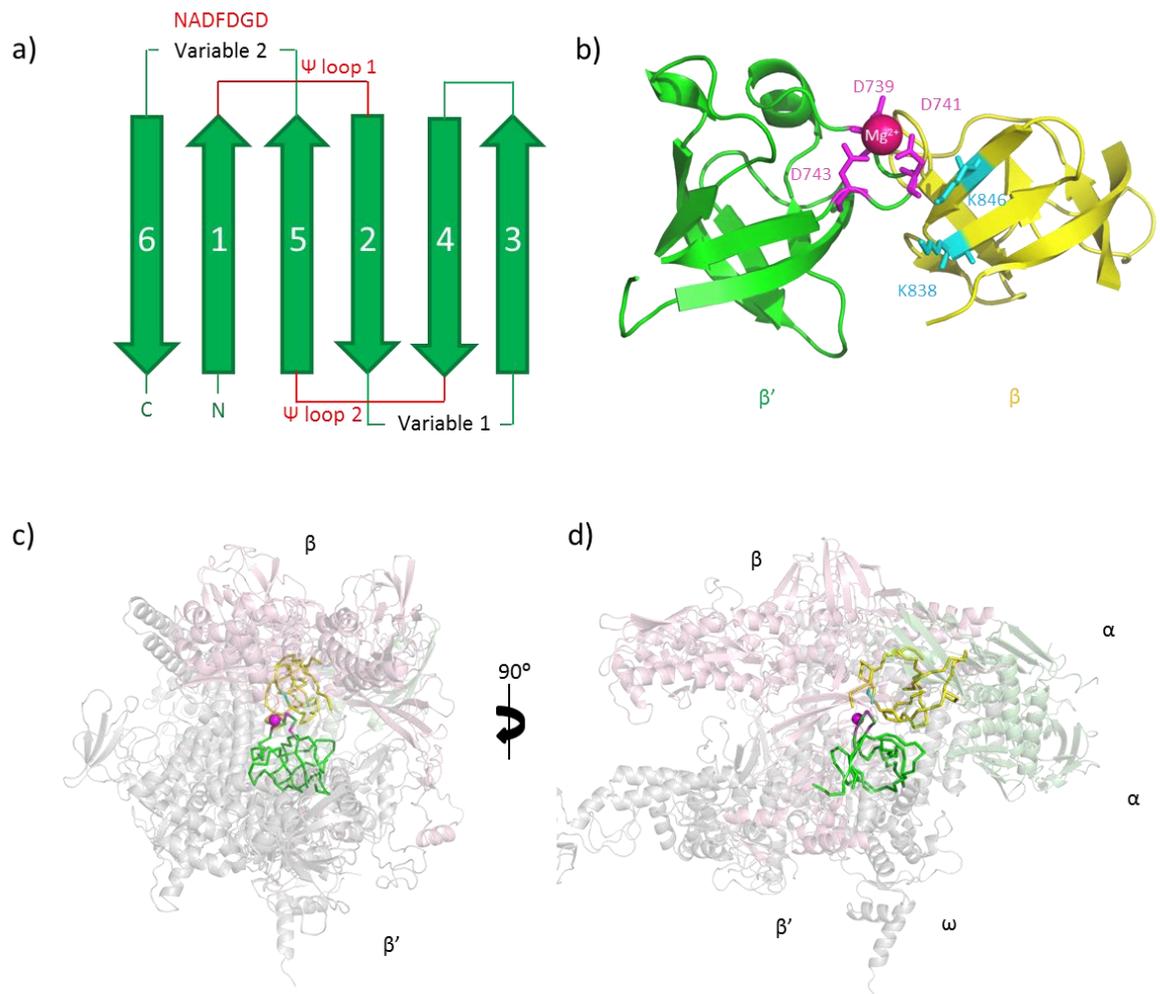
(Shown as  $Mg^{2+}$  B) is delivered to the active centre by substrate NTP (red).  $Mg^{2+}$  ions and the trigger loop (residues in blue boxes) coordinate the phosphates of the NTP.  $Mg^{2+}$  activation of 3' OH group on the RNA coupled to optimal orientation of the NTP catalyses a nucleophilic attack (curved arrows), resulting in the synthesis of a phosphodiester bond and release of PPI. **b) Folding of the trigger loop to contact NTP substrate** Taken from Yuzenkova et al., 2010. Shown is the active centre of *Thermus thermophilus* elongation complex (PDB 2PPB and 2O5J). When the correct NTP is present, the trigger loop undergoes conformation change, from an unfolded (purple) to folded (green) state. In doing so, residues such as R1239 and H1242 contact the phosphates of the NTP (orange) to contribute to the substrate orientation.

#### 1.1.5: Double $\psi$ $\beta$ Barrel Domains

Analysis of the structure of the  $\beta'$  subunit revealed the aspartate triad is present on a double  $\Psi$   $\beta$  barrel fold (DPBB). The fold is characterised by the topology of the  $\beta$  sheets, which involves 6  $\beta$  sheets and resembles two inverted  $\Psi$  symbols. A schematic of this fold is shown in Figure 1-5a (Castillo et al., 1999). Each DPBB has two additional variable regions situated between strands 2 and 3 and strands 5 and 6. Two psi-loops exist in the fold, between  $\beta$  strands 1 and 2 and strand 5 and 4. In the  $\beta'$  subunit DPBB, the catalytic aspartate triad is present on the second variable region which protrudes into the active centre.

A second DPBB motif was identified in the  $\beta$  subunit of RNAP. There is no sequence conservation between the  $\beta'$  and  $\beta$  DPBB folds, with the  $\beta$  DPBB only being identified through visual inspection of structure (Iyer et al., 2003). The  $\beta'$  and  $\beta$  DPBB folds are directly adjacent to each other, with a cleft running between the two folds. The aspartate triad is projected into this cleft (Figure 1-5b). Just as  $\beta'$  DPBB provides functional residues, so does the  $\beta$  DPBB. Two lysine residues protrude into the cleft and make contacts with the incoming NTP substrate (Iyer et al., 2003). The aspartate triad and the  $Mg^{2+}$  ion it chelates are regarded as the active centre, thus the two DPBB folds contribute to the formation of RNAP active centre.

Along with the aspartate triad, two DPBB folds are conserved across all multisubunit RNAPs. This conservation highlights the substantial role the two folds play in the structural formation of the active centre. The presence of an aspartate triad and two DPBB folds can serve as a hallmark of RNA polymerase activity, aiding in the identification of distantly related RNAPs.



**Figure 1-5: The two double  $\Psi$   $\beta$  barrel (DPBB) folds of RNAP. a) Topology of a DPBB fold.** Modified from Castillo et al., 1999. The topology of a standard DPBB fold is shown. The amino acid chain and  $\beta$  sheets are green, whilst the  $\Psi$  loops are labelled and coloured red. The  $\beta$  sheets are numbered according to the order in the amino acid chain. The  $Mg^{2+}$  chelating aspartate triad is shown in red above the variable loop 2, where it occurs in the  $\beta'$  DPBB. **b) Structure of the  $\beta'$  and  $\beta$  DPBB folds.**  $\beta'$  DPBB (green) and  $\beta$  DPBB (yellow) are shown, with the essential aspartate triad residues and conserved lysines in pink and blue, respectively. In the structure, the triad is chelating  $Mg^{2+}$ , represented by a pink sphere. The active centre is present in the cleft running between the two DPBB folds. The Figure was made using the crystal structure of *Thermus thermophilus* RNAP elongation complex (PDB accession code: 2PPB. From Vassylyev et al., 2007). **c) and d)  $\beta'$  and  $\beta$  DPBB folds in the context of the RNAP.** DPBB folds are in ribbon representation and coloured as before. The DPBB are present at the back wall of the main channel formed by  $\beta'$  and  $\beta$  subunits. In **d)** the placement of the aspartate triad, shown chelating  $Mg^{2+}$ , into the active centre by the protusion of  $\beta'$  variable loop 2 can be seen.

## **1.2: Single subunit RNA polymerases**

Despite multisubunit RNAP accomplishing canonical transcription in all cellular organisms, single subunit DNA dependant RNA polymerases have been identified. The most studied example is T7 RNAP of the T7 coliphage. T7 RNAP specifically transcribes T7 genes during the late phase of *E. coli* infection. Comparisons of T7 RNAP with multisubunit RNAPs have revealed the two classes of enzymes are extremely unlike, with no sequence and structural similarities. The structure and amino acid alignments of T7 RNAP and other related single subunit RNAPs show they resemble Pol I class DNA polymerases and HIV reverse transcriptase (Sousa et al., 1993). Thus single subunit RNAPs are members of a discrete clade of related nucleic acid polymerases distinct to the multisubunit RNAP (Delarue et al., 1990, Sousa et al., 1993, Salgado et al., 2006).

The structure of T7 RNAP and related protein resembles a right hand, a striking contrast to the crab claw like structure of the multisubunit RNAP (Sousa et al., 1993). The comparison to a right hand extends to the nomenclature of the T7 RNAP domains, which include the finger, palm and thumb domains. Like multisubunit RNAPs, right handed RNAPs perform catalysis using a general two ion mechanism (Steitz, 1998, Steitz and Steitz, 1993). However, the mechanics and structural domains participating in catalysis are very different to multisubunit RNAPs. No aspartate triad is present in T7 RNAP. Instead, two aspartates (D537 and D812), 250 amino acid sequences apart on the polypeptide chain, are responsible for chelating  $Mg^{2+}$  ions. The presence of the aspartate triad motif is therefore solely characteristic of multisubunit RNAP and further highlights the differences between the two classes of RNAP.

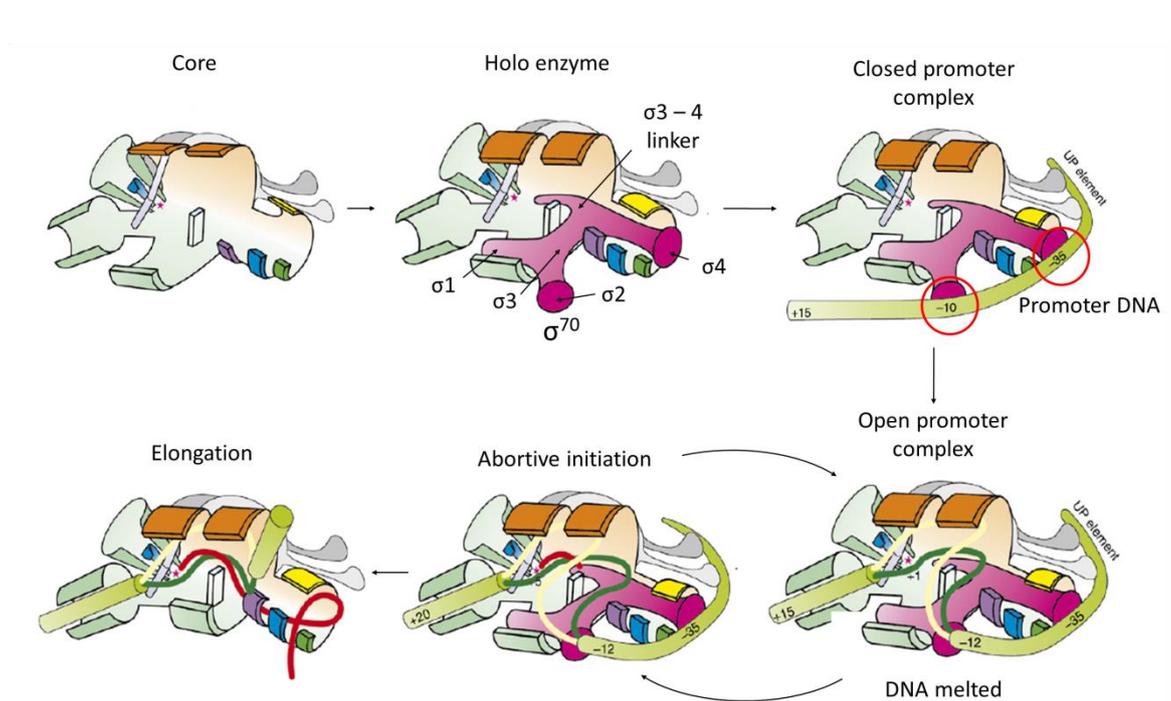
## **1.3: The Stages of Transcription**

Each round of transcription can be divided into three discrete phases: initiation, elongation and termination. Below, each phase will briefly be discussed.

### **1.4: Initiation of Transcription**

For transcription to begin, RNAP is required to find and bind specific promoter double stranded DNA sequences. Alone, RNAP cannot accomplish this task. All RNAPs require additional proteins, transcription factors, for recruitment to specific promoters and melting of the duplex DNA.

In bacteria, a  $\sigma$  factor binds the core RNAP to form the holoenzyme. The  $\sigma$  factors function as transcription factors by recruiting RNAP to promoters through recognising specific DNA sequences. Once RNAP is bound to double stranded DNA,  $\sigma$  factors are then responsible for the melting of the DNA duplex. Multiple  $\sigma$  factors exist in most bacterial species, with each  $\sigma$  factor recruiting RNAP to a different class of genes. In *E. coli*,  $\sigma^{70}$  is required to initiate transcription of housekeeping genes. Homologues of  $\sigma^{70}$  carry out the same function across other species of bacteria. In the *Bacillus* and *Thermus* genera, the housekeeping  $\sigma$  factor is  $\sigma^A$ . The details of  $\sigma^{70}$  aided transcription initiation are described below and shown in figure 1-6.



**Figure 1-6: Bacterial transcription initiation facilitated by  $\sigma^{70}$ .** Adapted from Nudler and Borukhov, 2008. The domains of RNAP are coloured as in figure 1-2, whilst  $\sigma^{70}$  is shown in pink. At the beginning of the transcription cycle, core RNAP binds to  $\sigma^{70}$  to become the holoenzyme. As shown, the domains of  $\sigma^{70}$  thread through the active centre of RNAP, with  $\sigma 2$  and  $\sigma 4$  domains being on the surface of the holoenzyme. This allows the  $\sigma 2$  and  $\sigma 4$  domains to contact the -35 and -10 elements (encircled in red), respectively. The holoenzyme bound to the double stranded DNA is referred to as the closed promoter complex. However, upon  $\sigma 2$  facilitating the melting of duplex DNA downstream of the -10 element, the complex becomes known as an open promoter complex. At this stage, RNAP undergoes abortive initiation. This is a cyclic process where short transcripts are synthesised and released, until eventually a nascent transcript enters the RNA exit channel. As  $\sigma^{70}$  occupies the RNA exit channel, this extension of RNA results in the eventual release of  $\sigma^{70}$  which allows RNAP to pass through promoter clearance and enter productive elongation.

$\sigma^{70}$  is composed of 4 conserved domains:  $\sigma 1$ ,  $\sigma 2$ ,  $\sigma 3$  and  $\sigma 4$ , with each domain being connected by flexible linkers (Campbell et al., 2002). Upon binding to RNAP,  $\sigma$  domains 2 and 4 are solvent exposed, allowing them to recognise and bind their cognate promoter DNA sequences (Murakami et al., 2002b).  $\sigma^{70}$  promoters are centred around two DNA sequences, the -35 element and -10 element. The -35 element is 35 bp upstream from the transcription start site (+1) and has a consensus sequence of TTGACA that interacts with the  $\sigma 4$  domain. The -10 element is 10 bp upstream of the start site and has a consensus sequence of TATAAT that is bound by the  $\sigma 2$  domain (Murakami et al., 2002a). The holoenzyme bound to double stranded DNA is referred to as the closed complex. Subsequent to closed complex formation, the  $\sigma 2$  domain triggers the melting of the DNA duplex 11 basepairs (-11 site) upstream from the transcription start site (+1 site) (Murakami et al., 2002a, Juang and Helmann, 1994, Panaghie et al., 2000). This DNA melting extends downstream and results in the placement of the +1 transcription start site on the template strand into the active centre, ready for RNAP to begin transcription. A second outcome of DNA melting is the formation of the transcription bubble which stretches from the -11 position to +2. Holoenzyme complexed with melted DNA is referred to as the open complex due to the open conformation of the DNA strands. Most open complexes are very stable and are capable of withstanding high ionic strength.

RNAP is not able to enter the processive elongation phase immediately after open complex formation. With  $\sigma^{70}$  still bound, RNAP undergoes a cycle of synthesising and releasing small RNA chains 2 - 8 nt in length, a process referred to as abortive initiation. In addition to binding and melting promoter DNA,  $\sigma^{70}$  functions during abortive initiation.  $\sigma 3$  and  $\sigma 4$  domains are linked by the flexible  $\sigma 3.2$  domain which features a loop that protrudes towards the active centre (Murakami et al., 2002b, Vassylyev et al., 2002). Correlating with its insertion into the active centre,  $\sigma 3.2$  has been shown to be required for binding the first nucleotides and the formation of the phosphodiester bond between them (Kulbachinskiy and Mustaev, 2006, Pupov et al., 2014).

Domain  $\sigma 3.2$  also facilitates promoter clearance, the event in which RNAP breaks contact with the promoter element to enter processive elongation. After numerous cycles of abortive initiation, RNAP eventually synthesises RNA which extends into the

RNA exit channel. The  $\sigma^{3.2}$  domain is partially blocking the RNA exit channel and so competes with the growing RNA chain (Murakami et al., 2002b). It is thought that upon clashing with the RNA chain, the removal of  $\sigma^{3.2}$  from the RNA exit channel triggers a series of steps which ultimately results in the breaking of contacts between  $\sigma^{70}$  and the -35 element, allowing RNAP to transcribe away from the promoter (promoter clearance) (Kulbachinskiy and Mustaev, 2006, Pupov et al., 2014). During promoter clearance,  $\sigma^{70}$  is displaced from RNAP. However, experimental evidence has revealed the expulsion of  $\sigma^{70}$  is not obligatory for RNAP to convert into a processive elongation complex (Bar-Nahum and Nudler, 2001). Rather, the conserved domains of RNAP, including  $\beta'$  clamp, contribute to processivity by tightening interactions on the nucleic acid upon transitioning into the elongation phase (Vassilyev et al., 2007a).

Eukaryotic transcription initiation is a complex processes, involving a multitude of transcription factors (Reviewed in Vannini and Cramer, 2012). Initiation of RNAP II transcription requires the assembly of the pre-initiation complex (PIC). Briefly, the PIC is a complex formed by numerous transcription factors: TBP, TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH. The PIC functions in recruitment of RNAP to promoter DNA, DNA unwinding, open complex stabilisation, transcription start site selection, and stimulation of early polymerisation. RNAP I and III require fewer external proteins for initiation; the additional core subunits of RNAP I and RNAP III have homology to the transcription factors utilised by RNAP II. Archaeal transcription is reminiscent of eukaryotic RNAP II transcription. However, the PIC is reduced to three transcription factors: TBP, TFB (TFIIB homologue) and TFE (TFIIE homologue) (Werner and Weinzierl, 2005). TBP recognises the TATA box promoter element whilst RNAP is recruited to TBP and promoter DNA by TFB. TFE stimulates the formation of the open promoter complex. TFB resembles  $\sigma^{70}$  in that it has a B finger domain which protrudes into the active centre of the RNAP and stimulates the initial catalytic activity of RNAP (Werner and Weinzierl, 2005).

Broadly speaking, the principles of transcription initiation are shared across all three domains of life despite increased complexity in eukaryotic organisms. All multisubunit RNAPs are dependent on transcription factors for recognition and binding of promoter DNA, followed by melting of the DNA to form the open complex. This is demonstrated by the resemblance shared between the  $\sigma^{3.2}$  of bacterial  $\sigma^{70}$  and the B finger of TFIIB

and TFB in eukaryotes and archaea, respectively. All three proteins feature a domain which protrudes towards the active centre and stimulates the initial catalytic activity of RNAP. Furthermore, all three domains are displaced from the RNA exit channel by the growing RNA chain which leads to processive transcription (Kulbachinskiy and Mustaev, 2006, Sainsbury et al., 2013, Werner and Weinzierl, 2005).

The disparity between RNAP and single subunit RNAPs continues to transcription initiation. T7 RNAP is able to initiate transcription independent of additional factors. The 300 amino acid N Terminal domain (promoter binding domain) and specificity loop of T7 RNAP are able to recognise the 17 base pair promoter sequence. Transcription initiation by T7 RNAP is accompanied by gross conformational rearrangements (Durniak et al., 2008, Yin and Steitz, 2002). During progression to elongation, the N terminal domain contacting the promoter DNA undergoes a large structural change to break DNA contacts, allowing promoter clearance. The conformational change results in the opening of the enzyme to accommodate an 8bp RNA:DNA hybrid (Yin and Steitz, 2002). Prior to this structural rearrangement, T7 RNAP can only accommodate a hybrid of 3 bp. This reflects the scale of the change in conformation. An additional outcome of the rearrangement is the formation of the RNA exit channel. Together, the increased hybrid length and RNA exit channel allow for an increase in the processivity of the enzyme.

## **1.5: Transcription Elongation**

Unlike transcription initiation, the elongation phase of transcription is conserved across all multisubunit RNAPs. Below, bacterial elongation will be discussed.

### *1.5.1: Nucleotide addition cycle*

During the elongation phase of transcription, RNAP catalyses the rapid and processive incorporation of substrate NMPs into the growing 3' end of RNA. The process of NMP incorporation is repeated for each position along the template DNA. This process is referred to as the nucleotide addition cycle (NAC). The steps of the NAC are shown in Figure 1-7.

The RNAP active centre has two binding sites, *i* and *i*+1 (insertion site), which the 3' end of RNA and the incoming substrate NTP can occupy. At the beginning of the NAC, elongation complexes are in a post-translocated state. In this state the 3' end of RNA

occupies the  $i$  site and  $i+1$  is vacant. The incoming NTP arrives to the active centre and is positioned into the  $i+1$  site where it forms hydrogen bonds with the complementary template base, ready for catalysis. Crystal structures, supported by biochemistry, have revealed additional nucleotide binding sites within the active centre, notably the pre-insertion site (Kettenberger et al., 2004, Vassylyev et al., 2007b). The pre-insertion site is thought to serve as an initial binding site for substrate NTPs. If the correct NTP is bound, the NTP repositions into the  $i+1$  site. This is coupled to the folding of the trigger loop which closes off the active centre to bring about catalysis.

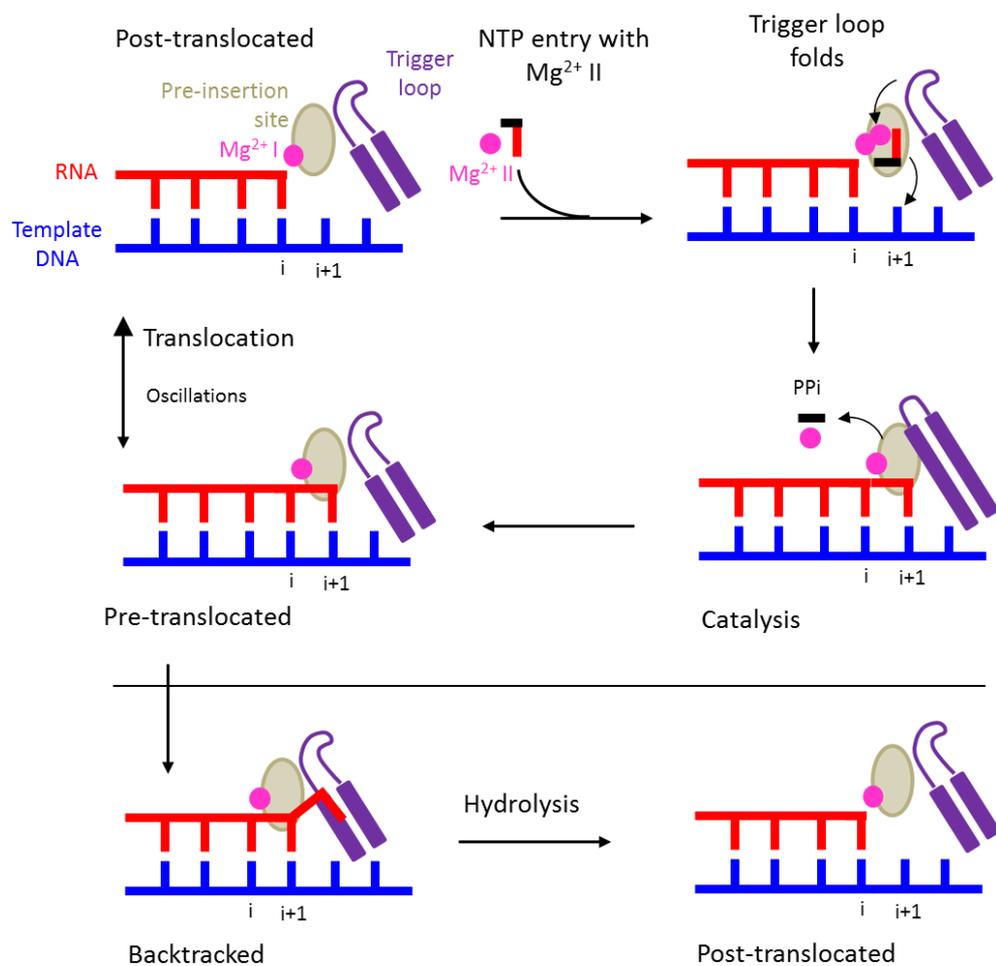
The result of catalysis is the incorporation of substrate NMP present in the  $i+1$  site into the 3' end of the transcript. Therefore, after catalysis the new 3' end of RNA occupies the  $i+1$  site, leaving the complexes in a pre-translocated state. Translocation is the lateral movement of the RNAP by one position on the template DNA strand. To carry on elongation, RNAP translocates to position the new 3' RNA end into the  $i$  site thereby freeing the  $i+1$  site for the next NTP and returning the complexes to a post-translocated state. At any position on the DNA template, RNAP is able to oscillate from a pre-translocated to post-translocated state due to thermal motion. In the presence of NTPs, forward translocation and stabilisation of the post-translocated state is thought to occur by a Brownian ratchet mechanism as the binding of the next NTP in the now empty  $i+1$  site favours the forward position of RNAP (Bar-Nahum et al., 2005).

### *1.5.2: Pausing and Backtracking*

Despite primarily moving in a forward direction, occasionally RNAP can pause on the DNA. The pausing of RNAP on the DNA is considered to control the overall speed of elongation and so pausing features in various regulatory mechanisms including transcription termination and attenuation (reviewed in Landick, 2006). The pause itself is induced by the DNA sequence, which can be recognised by RNAP (Bochkareva et al., 2012, Kireeva and Kashlev, 2009). One outcome of transcribing a pause sequence is reverse movement of the RNAP along the RNA:DNA hybrid. This reverse movement is backtracking (bottom of Figure 1-7). For backtracking to occur, the 3' end of RNA disengages from the active centre and is fed into the secondary channel as RNAP slides backwards along the hybrid. In addition to pause sequences, the incorporation of the wrong nucleotide (misincorporation) can trigger backtracking. As the 3' end of RNA is

no longer within the active centre, backtracking events need to be resolved before the resumption of transcription.

Given the importance of transcription, auxiliary factors exist to modulate the rate of elongation. Two examples of elongation factors are NusA and NusG, which control elongation speeds by promoting and suppressing RNAP pauses, respectively (Artsimovitch and Landick, 2000, Herbert et al., 2010). Backtracking events are resolved through the removal of the 3' nascent RNA. RNAP has an intrinsic RNA hydrolysis activity which can accomplish the removal of nascent RNA (Orlova et al., 1995). However, *in vivo*, Gre proteins in bacteria and TFIIS for RNAP II exist to cleave the RNA and rescue backtracked RNAP (Borukhov and Nudler, 2008, Cipres-Palacin and Kane, 1994).



**Figure 1-7: The Nucleotide addition cycle.** At the beginning of the cycle, the RNAP is in a post-translocated state. The 3' RNA (red) occupies the *i* site. The incoming substrate NTP, chelating  $Mg^{2+}$  II, binds to the preinsertion site. The trigger loop (purple) folds, positioning the NTP in the *i+1* site. The folded trigger loop and  $Mg^{2+}$  ions catalyze the

formation of a phosphodiester bond. The reaction results in the release of a pyrophosphate (PPi). After catalysis the RNAP is in a pre-translocated state. RNAP translocates to vacate the *i*+1 site, ready for another round of nucleotide addition. RNAP can oscillate between the pre and post-translocated states, as shown by the double headed arrow. RNAP is able to backtrack on the RNA:DNA hybrid. Backtracked RNAP cannot continue RNA synthesis. Through its own hydrolytic activity, or a cleavage factor, the nascent 3' RNA is removed to return the RNAP to a post-translocated state.

### 1.5.3: Hydrolysis and proof reading

Misincorporation, incorporation of the wrong nucleotide, leads to the production of inaccurate transcripts that can have a potentially drastic effect on gene expression. Therefore maintaining high fidelity is an important aspect of transcription. *In vitro* studies have helped to delineate the determinants of transcription fidelity. The prevention of misincorporation occurs in a multistep process during the binding of the substrate (Yuzenkova et al., 2010). The major selection step occurs upon the binding of the substrate and the folding of the trigger loop. Key residues of the trigger loop, including the catalytic R1239 and H1242 residues, are able to discriminate against incorrect NTPs such as non-complementary or 2' deoxyNTPs (dNTPs) that have wrongly been positioned in the *i*+1 site. In the instance of an incorrect NTP in *i*+1, steric clashes with the folding trigger loop cause catalysis to occur very slowly which favours the expulsion of the NTP from the active centre (Yuzenkova et al., 2010).

In addition to substrate discrimination, RNAP contributes to fidelity by removing misincorporated NTPs through catalysing hydrolysis. In the event of misincorporation, RNAP backtracks by one position, placing the penultimate NMP of the RNA into the *i*+1 site. Upon backtracking, the 3' misincorporated NMP is flexible as it is not base paired with the template DNA. The E site is an additional nucleotide binding site in the active centre in which the incoming NTP substrate is thought to bind first. The flexibility of the 3' misincorporated NMP allows it to bind in the E site and provide chemical groups which function in the catalysis of hydrolysis, including the stabilisation of Mg II (Zenkin et al., 2006b). The trigger loop, in particular the H1242 residue is thought to facilitate the positioning of the 3' NMP into the E site (Yuzenkova and Zenkin, 2010). The result of hydrolysis is the release of a dinucleotide and a return to the post-translocated state. This process is referred to as transcript assisted hydrolysis because the misincorporated NMP aids in its own removal by stimulating back tracking and participating in the hydrolysis reaction (Zenkin et al., 2006b).

Although RNAP possesses an intrinsic hydrolysis activity, conserved external transcription factors exist to catalyse the hydrolysis of RNA in the event of misincorporation or transcription arrest *in vivo* (Orlova et al., 1995). The Gre factors in bacteria (GreA/GreB in *E. coli*) and their eukaryotic counterparts (TFIIS for RNAP II, A12.2 and C11 core subunits for RNAP I and III) provide catalytic residues for hydrolysis by protruding into the active centre through the secondary channel (Kettenberger et al., 2003, Roghanian et al., 2011, Laptenko et al., 2003, Sosunova et al., 2003). *In vitro* experiments have revealed that Gre factors replace the trigger loop when supplying catalytic residues to chelate Mg II and position the attacking water molecule (Roghanian et al., 2011).

### **1.6: Transcription Termination**

Two independent mechanisms exist in bacteria for the termination of transcription: intrinsic termination and rho-dependant termination. In intrinsic termination, RNAP transcribes a G-C rich inverted repeat followed by a stretch of thymine residues. The A:U hybrid is weak and induces a pause in transcription, allowing the transcribed G-C rich inverted repeat to form an RNA hairpin (Gusarov and Nudler, 1999). The hairpin extends into the main channel, reducing the length of the RNA:DNA hybrid, clashing with the trigger loop and opening the  $\beta'$  DNA clamp domain (Epshtein et al., 2007). As the  $\beta'$  clamp is responsible for RNAPs stable grasp on DNA, its opening causes a release of the transcript and the removal of RNAP from the DNA.

The second mechanism of transcription termination requires Rho. Rho is a homohexamer, with each subunit having the ability to perform ATP hydrolysis. During transcription, nascent RNA is fed through the RNA exit channel and out of RNAP. Rho binds to the nascent RNA, in particular to cytidine rich sequences and hydrolyses ATP to thread the nascent RNA through its ring-like structure (Peters et al., 2009). Threading of RNA allows Rho to translocate up the RNA in a 5' to 3' direction towards the transcribing RNAP. Rho then interacts with RNAP to alter its conformation, trapping the elongation complex in an inactive state. Release of the RNA and DNA then ensues, thus terminating transcription (Epshtein et al., 2010).

Termination of transcription varies for each of the eukaryotic RNAPs, with two mechanisms operating to terminate RNAP II (reviewed in Kuehner et al., 2011). Protein coding mRNA transcribed by RNAP II undergoes cleavage and polyadenylation at the 3' end during the final stages of transcription. The termination of transcription is coupled to these 3' end processing events. Briefly, the proteins required for cleavage and polyadenylation are recruited to the C-terminal domain (CTD) of RNAP II. After the transcription of the polyadenylation signal, the mRNA is cleaved and polyadenylated at newly exposed 3' end. The other outcome of cleavage is exposure of the 5' end of the nascent RNA chain still bound by RNAP II. Rat1, a 5'-3' exonuclease bound to the CTD proceeds to degrade the nascent RNA which results in the removal of RNAP II from the DNA through a collision event. The second mechanism of RNAP II termination applies to the transcription of non-coding genes. This mechanism occurs through the recruitment of the Sen1 helicase which unwinds the RNA:DNA hybrid in the active centre, leading the release of RNA and removal of RNAP II from DNA.

RNAP I and RNAP III termination are simpler processes. *In vitro*, a stretch of thymine bases and the Reb1 roadblock protein bound to its corresponding DNA element downstream is sufficient to terminate RNAP I transcription (Lang et al., 1994). However, results of *in vivo* work has implicated pre-rRNA processing factors in RNAP I termination. Evidence suggests as for RNAP II, RNAP I transcripts are cleaved with downstream RNA remaining bound to RNAP I. This RNA is then degraded by Rat1 exonuclease which upon contacting the RNAP I, triggers the removal of RNAP I from DNA (reviewed in Nemeth et al., 2013)) RNAP III termination resembles the intrinsic termination of bacterial RNAP in that both mechanisms utilise RNA secondary structure and a downstream poly U tract (Zenkin, 2014a). Upon transcribing the poly U tract, RNAP III backtracks and becomes catalytically inactive (Nielsen and Zenkin, 2013). RNAP III transcribes structured RNA including tRNA. In a backtracked state, the nascent RNA folds into its secondary structure which brings about the removal of the RNAP III from the template DNA.

It remains unclear how archaeal transcription is terminated. The limited experimental evidence suggests that the mechanism is similar to that of bacterial intrinsic termination (Santangelo and Reeve, 2006, Zenkin, 2014b). Whilst a bacterial termination signal featuring an RNA hairpin and a downstream thymine rich DNA

sequence was recognised by archaeal RNAP, the RNA hairpin was shown not to be essential for termination. Interestingly, archaeal transcription termination is influenced by the sequence of upstream DNA but the cause of this remains to be investigated.

### **1.7: Evolution of multisubunit RNA Polymerases**

The  $\beta'$  and  $\beta$  DPBB folds are conserved in all multisubunit RNAPs, leading to the suggestion that the two DPBB form an ancient core from which the multisubunit RNAP evolved (Iyer et al., 2003). The two DPBB folds are only structurally homologous and do not share any similarities in amino acid sequence. Despite this, it has been proposed that the  $\beta'$  and  $\beta$  subunits originate from a single protein containing the DPBB fold, which duplicated in the last universal common ancestor (LUCA). The individual DPBB folds then underwent divergent evolution to loose sequence similarity and gain differing motifs, with  $\beta'$  DPBB gaining the asp triad. Finally, the two DPBB folds split to form two separate subunits. (Iyer et al., 2003, Iyer and Aravind, 2012).

Single subunit RNA dependant RNAPs (rdRNAPs) exist in a wide variety of eukaryotic organisms. Rather than bearing homology to other single subunit RNAPs such as T7 RNAP, bioinformatics revealed a distant relationship to the multisubunit RNAP (Iyer et al., 2003). All eukaryotic rdRNAPs contain an aspartate triad and the presence of two DPBB domains has been confirmed by the structure of one such rdRNAP, QDE-1 (Salgado et al., 2006, Iyer et al., 2003). Therefore RNAPs exist which contain the  $\beta'$  and  $\beta$  DPBB folds on a single polypeptide. It is thought that these enzymes diverged prior to the splitting of the two DPBB into individual subunits (Iyer and Aravind, 2012, Salgado et al., 2006, Iyer et al., 2003).

The presence of two DPBB folds and an aspartate triad has been used to identify a number of hypothetical, non-canonical RNAPs (Aravind et al., 2012). Patterned based PSI BLAST searches identified hypothetical bacteriophage encoded proteins with distant homology to eukaryotic rdRNAPs and multisubunit RNAPs (Iyer et al., 2003). One example is the YonO protein encoded by the *Bacillus subtilis* prophage SP $\beta$  (discussed in the next section). Recently, the identification of two DPBB folds has further suggested that the ORF6 protein of the *Kluyveromyces lactis* killer system is an RNA polymerase (Ruprich-Robert and Thuriaux, 2010). Previously, this protein was identified as a putative RNAP through amino acid sequence similarities (Wilson and

Meacock, 1988). Furthermore, when projected onto the crystal structure of RNAP II, the amino acid homology mapped onto the two DPBB folds suggesting ORF6 RNAP contains these characteristic RNAP folds (Ruprich-Robert and Thuriaux, 2010).

In the next chapters, YonO and ORF6 RNAPs will be discussed in more detail, including their discovery, what is currently known about these proteins and descriptions of the transcription systems they are thought to operate within.

## Chapter 2: YonO: the hypothetical RNA polymerase of the *Bacillus subtilis* prophage SP $\beta$

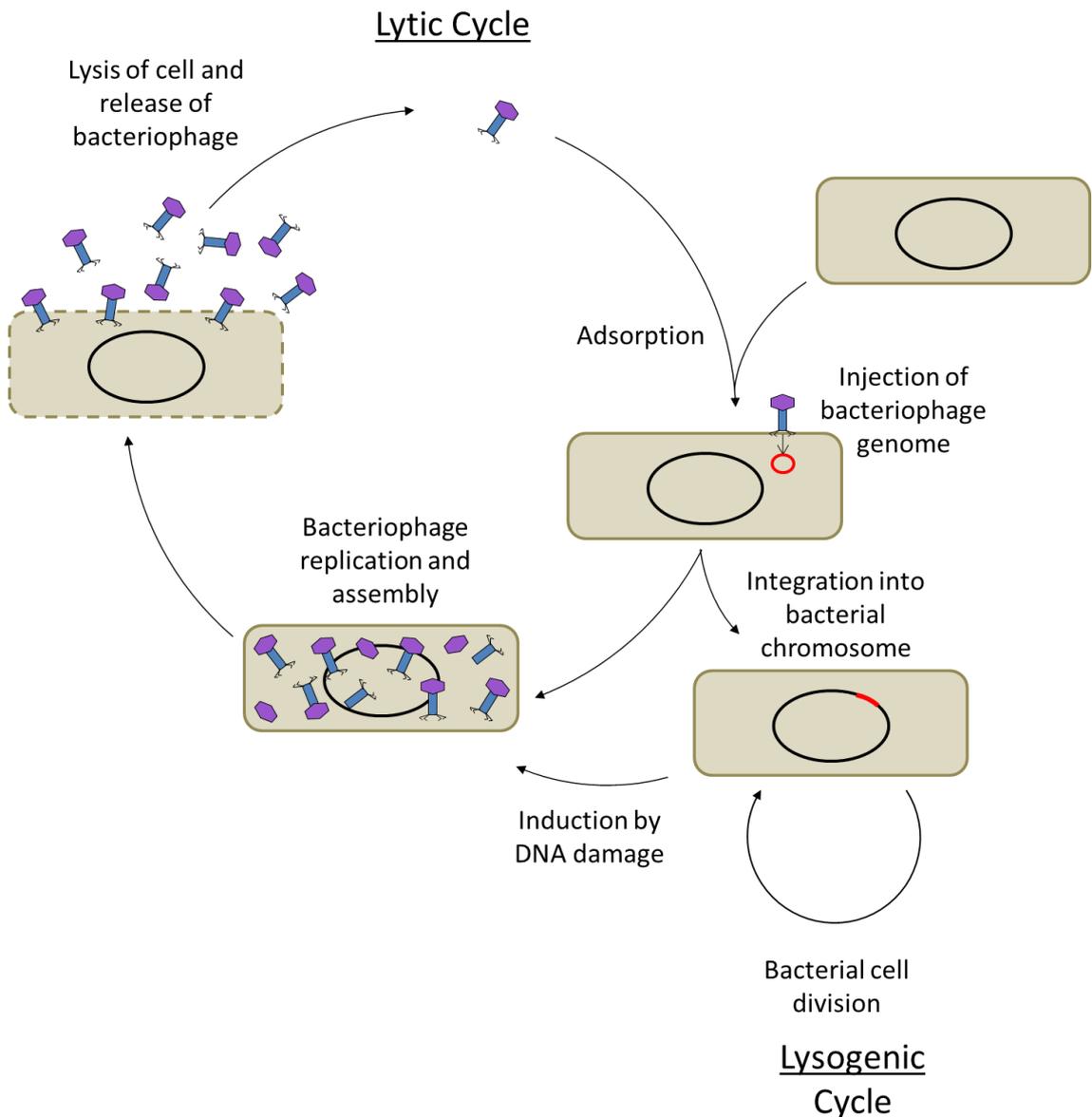
### 2.1. Resident prophages of *Bacillus subtilis*

*Bacillus subtilis* is a soil dwelling, gram positive rod bacteria. Many aspects of its physiology have made *B. subtilis* a widely used model organism for the study of gram positive bacteria. The *B. subtilis* 168 strain is widely regarded as a wildtype strain. The chromosome of this strain and its many derivatives harbours two prophages and 7 prophage-like elements (Westers et al., 2003, Kunst et al., 1997). Prophages are genomic integrations of temperate bacteriophage DNA. Prophage DNA remains in the host chromosome until induction stimulates the replication and release of the bacteriophage particle. This life cycle of temperate bacteriophages is referred to as the lysogenic cycle. In addition to the lysogenic life cycle, bacteriophages can also replicate by a lytic life cycle. In contrast to the temperate bacteriophages, lytic bacteriophages begin replication immediately after the injection of their DNA into the host bacteria. This makes the lytic life cycle the quickest mode of replication, with the T7 bacteriophage producing approximately 180 phage particles in less than 20 minutes under optimal conditions (Nguyen and Kang, 2014). The two life cycles are outlined in Figure 2-1.

The two prophages residing in the *B. subtilis* chromosome are PBSX and SP $\beta$  (Seaman et al., 1964, Warner et al., 1977). Spontaneous induction of the prophages can occur at very low frequencies (1 in  $10^6$  for SP $\beta$ ), resulting in cell lysis and the release of phage particles (Warner et al., 1977). Induction of PBSX and SP $\beta$  is controlled by the RecA dependant SOS response triggered by DNA damage (Goranov et al., 2006). Experimentally, the DNA crosslinker Mitomycin C is used to stimulate the induction of the prophages (Warner et al., 1977, Seaman et al., 1964).

PBSX is a defective bacteriophage as upon its induction, *B. subtilis* DNA rather than the PBSX genome is packaged into the phage particles. Furthermore, PBSX particles do not inject their DNA into susceptible *B. subtilis* cells. Instead, PBSX particles cause the rapid lysis of susceptible bacteria upon adsorption (Seaman et al., 1964). In contrast, SP $\beta$  forms phage particles capable of lysogenising sensitive strains (Warner et al., 1977).

Therefore SP $\beta$  is a functional temperate bacteriophage which stably resides in the chromosome of *B. subtilis* 168 strain.



**Figure 2-1: Lifecycle of bacteriophages.** Bacteriophages can be classified according to the life cycle they undertake during infection of a host bacterium. Lytic bacteriophages such as T7 bacteriophage replicate by the lytic life cycle. In this life cycle, the bacteriophage directly progresses to replication and assembly of progeny phage particles subsequent to injecting the host with its DNA. This is in contrast to temperate bacteriophages which replicate by a lysogenic life cycle. In this cycle, the bacteriophage DNA integrates into the host chromosome to become a prophage. The prophage stably resides in the chromosome and is therefore replicated during bacterial cell division. In the event of unfavourable conditions, such as DNA damage, the prophage is induced and progresses to replication, particle assembly and eventual host cell lysis.

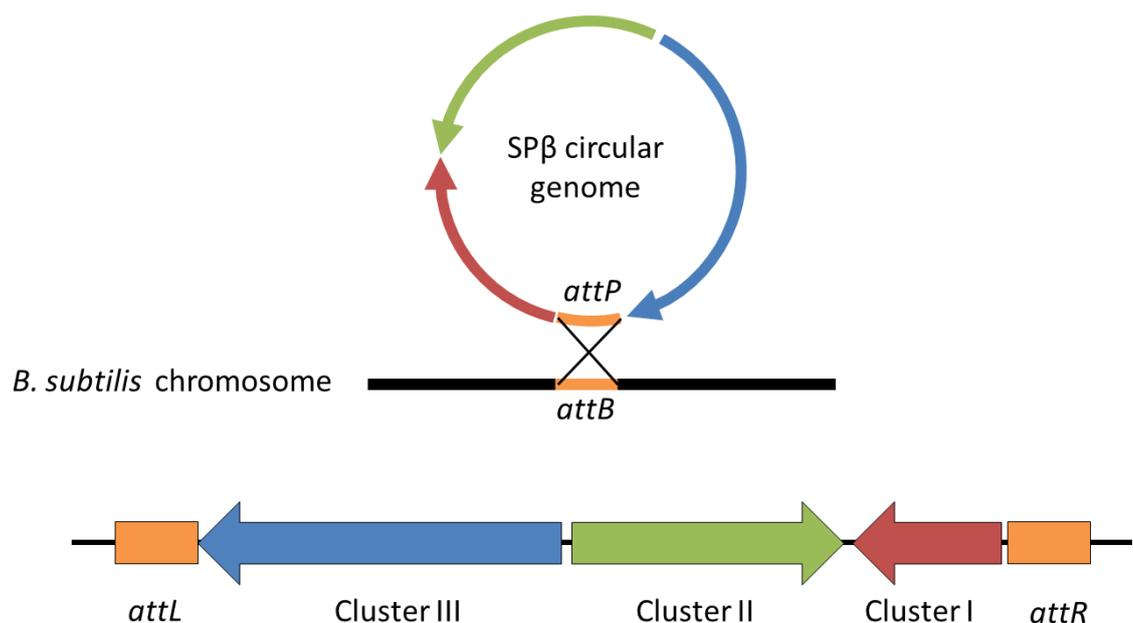
The 7 prophage-like elements in the genome include the *ICEbs1* (integrative conjugative element) and the SKIN (*SigK* intervening) element. *ICEbs1* is a 20 kb mobile element encoding its own excisionase and integrase, allowing for its horizontal transfer to recipient cells through conjugation (Auchtung et al., 2005, Lee et al., 2007). The SKIN element is a 48 kb prophage-like element, with 3 of its operons resembling the early, middle and late operons of PBSX (Krogh et al., 1996). The remaining genes of the SKIN element are predicted to function in arsenate metabolism (Takemaru et al., 1995, Krogh et al., 1996). An unusual property of the SKIN element is that its insertion into the genome disrupts the *sigK* gene coding for the sigma factor  $\sigma^K$ . Upon sporulation, the SKIN element is excised from the mother cell chromosome by the action of a site-specific recombinase, which itself is encoded by the *spoIVCA* gene on the SKIN element (Stragier et al., 1989, Sato et al., 1990, Kunkel et al., 1990). The excision results in the reconstitution of the *sigK* gene and expression of  $\sigma^K$ , which directs transcription in the sporulating mother cell (Kunkel et al., 1990, Sato et al., 1990). Despite having homology to PBSX, expression of SKIN element genes is not induced by the triggering of the SOS response by Mitomycin C (Krogh et al., 1996). This is in contrast to *ICEbs1*, whose expression is triggered by the SOS response or high cell densities (Auchtung et al., 2005).

## **2.2: SP $\beta$ prophage of *Bacillus subtilis***

SP $\beta$  has a very large 134 Kb genome that integrates into the host genome via recombination during lysogeny. Recombination occurs between the SP $\beta$  attachment site (*attP*) and the *B. subtilis* attachment site (*attB*) located towards the replication terminus of the *Bacillus* chromosome (Figure 2-2) (Zahler et al., 1977, Lazarevic et al., 1999). Generally, SP $\beta$  remains stably integrated into the genome until induction of the phage. However, spontaneous excision of the phage does occur, leading to the isolation of cured *B. subtilis* strains. One such strain, Su<sup>3+</sup> proved invaluable for the initial characterisation of SP $\beta$ . Su<sup>3+</sup> was susceptible to lysogenisation by SP $\beta$ , providing investigators a means to investigate SP $\beta$  induction and lysogenisation (Warner et al., 1977).

The whole sequence of the prophage DNA uncovered the organisation of the genome. 187 ORFs have been predicted. The majority of the proteins (~75%) encoded by the 187 ORFs on SP $\beta$  have unknown functions (Lazarevic et al., 1999). Putative functions

have been assigned to encoded proteins based on their amino acid sequence similarities. The ORFs have been organised into three clusters based on their orientation on the genome, as shown in Figure 2-2. Cluster I and III are directed towards the origin of replication. Cluster II is transcribed in the direction away from the origin. Cluster III contains many short ORFs predicted to function in DNA replication and metabolism (Lazarevic et al., 1999). Therefore, it has been posited that Cluster III contains the operons transcribed early during the induction of SP $\beta$ . Cluster II contains ORFs encoding for proteins with homology to the lytic enzymes of PBSX (Lazarevic et al., 1999). Additionally, homologues of PBSX tail fibre proteins are encoded in Cluster II. Tail fibre proteins are structural proteins of the phage particle. Tail fibre proteins are expressed at the end of phage particle maturation along with lytic enzymes which degrade the cell wall of the host bacterium. Therefore Cluster II contains the late operons.



**Figure 2-2: Genome organisation of SP $\beta$ .** **Top: Integration of SP $\beta$  genome into the *B. subtilis* chromosome.** Upon infection of a *B. subtilis* cell, the circular SP $\beta$  genome integrates into the host chromosome by homologous recombination between the *attP* and *attB* attachment sites on the SP $\beta$  and *B. subtilis* genomic DNA, respectively. Clusters I, II and III are shown in red, green and blue. In the circularised form, Cluster II is orientated in the opposite direction of Clusters I and III. **Bottom: Genome organisation of SP $\beta$  prophage.** Homologous recombination and subsequent integration of SP $\beta$  duplicates the attachment site, giving *attL* and *attR* either side of SP $\beta$ . Upon integration, Cluster II is the only cluster to be coded on the plus DNA strand.

Unlike Cluster II and III, Cluster I and the genes it contains are not categorised based on when they are expressed during SP $\beta$  induction (i.e. early or late). A number of the proteins encoded by SP $\beta$  DNA are expressed during the prophage state. Examples of such proteins are found encoded on Cluster I. *B. subtilis* cells harbouring SP $\beta$  produce the bacteriocin sublancin 168 (Paik et al., 1998). Sublancin 168 has bactericidal activity on a broad range of gram positive organisms including *Staphylococcus aureus*. Sublancin 168 is encoded by the *sunA* gene in Cluster I (Lazarevic et al., 1999, Paik et al., 1998). Immediately downstream lies an operon coding for the specific sublancin 168 exporter along with the enzymes required for sublancin synthesis. The ability to kill other competing bacteria is an advantage conferred to *B. subtilis* cells by SP $\beta$ , explaining the continued maintenance of the large prophage in the *B. subtilis* genome. Beneficial proteins are not exclusively encoded by ORFs on cluster I. The SspC protein encoded by the *sspc* gene of Cluster III has been incorporated into the complex process of sporulation. SspC binds to DNA and offers the spore protection against UV radiation (Tovar-Rojo and Setlow, 1991).

The detailed mechanism of SP $\beta$  induction by DNA damage has not been elucidated. Through transcriptome analysis, it has been shown that induction of SP $\beta$  and PBSX is dependent on RecA and LexA (also referred to as DinR in *B. subtilis*) (Goranov et al., 2006). RecA mediates the SOS response in bacteria. Upon DNA damage, RecA stimulates the autocleavage of DinR. DinR is a transcriptional repressor of the SOS regulon and binds the SOS box element to prevent transcription. The nucleotide sequence of SP $\beta$  has revealed the locations of four SOS box sequences in the SP $\beta$  genome (Lazarevic et al., 1999). The SOS response and cleavage of DinR may result in the derepression of numerous SP $\beta$  operons and/or ORFs. An SP $\beta$  encoded protein synthesised as a consequence of the SOS response may inhibit an SP $\beta$  repressor protein, allowing the full induction of the prophage (Lazarevic et al., 1999). Hypothetical SP $\beta$  repressor binding sites have been identified. SPBRE (SP $\beta$  repeated element) are 30 bp sequences found in intragenic spaces upstream from ORFs. 8 SPBRE are present in Cluster III. One is present upstream of the first ORF of Cluster II. Predicted  $\sigma^A$  promoters are located within or directly adjacent to SPBRE sequences. Any repressor of SP $\beta$  induction may bind to the SPBRE to occlude the  $\sigma^A$  promoter. A repressor of SP $\beta$  lytic growth has been identified, the *c* gene. Mutations in the *c* gene

prevent SP $\beta$  from lysogenising susceptible cells, with the mutant SP $\beta$  being restricted to lytic growth on susceptible strains (Warner et al., 1977). A temperature sensitive mutation in the c gene, defined as c2, allows induction of SP $\beta$  when the lysogenic cells are shifted to a non-permissive temperature (Rosenthal et al., 1979). How the c gene protein regulates lytic growth is unknown. It also not known whether there is a connection between the c gene protein and the predicted SPBRE elements.

### **2.3. YonO, a putative RNA polymerase of SP $\beta$**

Single subunit RNA dependant RNAPs (rdRNAPs) function in post transcriptional gene silencing across a wide variety of eukaryotic organisms. In an effort to investigate the origins of the rdRNAP, pattern based protein data base searches (PSI – BLAST) were conducted to identify potential distant homologues (Iyer et al., 2003). In PSI-BLAST, the patterns of conservation in the amino acid alignments produced by the first PSI-BLAST query are used for the second database search (iteration). This process can be repeated multiple times, with each iteration producing amino acid alignments of increasingly distantly related proteins. Specifically, an 800 residue core common to rdRNAPs was used as the PSI-BLAST query sequence. The second iteration revealed distant homology between the rdRNAP core and the protein coded by the *YonO* ORF of SP $\beta$  (Iyer et al., 2003). Furthermore, a reciprocal PSI-BLAST using YonO amino acid sequence as the query resulted in the discovery of YonO homologues. Primarily, YonO homologues reside in firmicute bacteria including *Bacillus* and *Clostridium* species. However, distant homologues have been also been identified in cyanobacteria (Ruprich-Robert and Thuriaux, 2010, Iyer and Aravind, 2012).

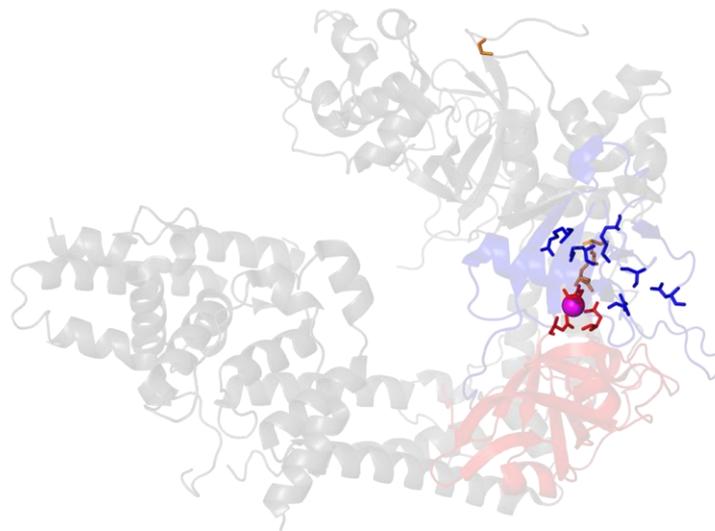
The homology between rdRNAP and YonO had a significant e value in the PSI BLAST search. However, the two proteins are only distantly related, with amino acid sequence similarities limited to 9 motifs and 12 invariant amino acids (Iyer et al., 2003). Secondary structure prediction showed conservation of 20  $\beta$  sheets and 16  $\alpha$  helices conserved across YonO and rdRNAPs. 3 of the 12 invariant amino acids are those of the aspartate triad motif, which is regarded as the RNAP active centre (Ruprich-Robert and Thuriaux, 2010). In addition to the aspartate triad, 7 of the remaining amino acids conserved in YonO also map to the active centre of rdRNAPs. This was shown by identifying the conserved residues in the structure of QDE-1, a well-studied example of a eukaryotic rdRNAP from *Neurospora crassa* (Figure 2-3a). The

conservation of the aspartate triad and other residues of the rdRNAP active centre strongly suggests YonO is an RNAP functioning in an undetermined stage in the SP $\beta$  life cycle.

The PSI-BLAST search also uncovered a very distant homology between YonO and rdRNAP to the multisubunit RNAP (Iyer et al., 2003). Amino acid sequence similarity between YonO and multisubunit RNAP is limited to the aspartate triad of  $\beta'$  and a conserved lysine and aspartate residue of  $\beta$ . The aspartate triad motif shared between rdRNAPs and YonO has less conserved amino acids than the motif shared between multisubunit RNAP, with the conserved motif being DbDGD (b represents a bulky amino acid). In multisubunit RNAPs the conserved motif is extended to NADFDGD. Figure 2-3b shows a comparison of the aspartate triad in YonO, QDE-1 and  $\beta'$ . In multisubunit RNAPs, the aspartate triad is present on a loop between two  $\beta$  strands of a DPBB fold in the  $\beta'$  subunit. The calculated secondary structure of YonO and rdRNAPs predicted the aspartate triad to be buried in a region enriched with  $\beta$  sheets, suggesting the presence of a DPBB fold. Since this initial prediction, the structure of QDE-1 was solved. This confirmed that the aspartate triad does indeed exist on a DPBB fold that corresponds to that of the  $\beta'$  subunit.

The structure of QDE-1 also revealed the presence of the second DPBB fold corresponding to the  $\beta$  subunit DPBB fold. Two closely spaced lysines present in the  $\beta$  DPBB fold are extremely conserved in multisubunit RNAPs. However, amino acid sequence comparisons show that only the second lysine is present in YonO and eukaryotic rdRNAPs. The first lysine has been substituted for an extremely conserved arginine residue (Iyer et al., 2003, Ruprich-Robert and Thuriaux, 2010). The remaining YonO residues that map to the QDE-1 active centre are present on the  $\beta$  DPBB, giving further evidence that both DPBB folds are present in YonO.

a)



b)

Predicted secondary structure	$\beta$	$\beta$	$\beta$	$\beta$	$\beta$	$\beta$
QDE-1	948 ESFTLLSDCD----VLVARSP	AHFPSD-IQRVRAVFKPELHS	-LKDVIIIFSTKG---DVPLAKKL--SGGDYDGL-MAVVCWD			
YonO	473 LSGSDVHCSLYD--EGYIDIL	1 SPHLFR-EHGVWNKKNSEYE	1 WFITPGVYTSI---HDPISKLL--Q-FDNDGL-KALIISD			
$\beta'$ <i>E. coli</i>	350 SGRSVITVGPYLR--LHQGLP	44 EVIREHVPVLLNRAPTLHRLGI	QAFEPVLIIEGKA---IQHLPLVCAAYNAFDGL-QMAVHVP			
$\beta'$ <i>T. thermophilus</i>	626 SGRSVIVVGPQLK--LHQGLP	47 EVIHGKVVLLNRAPTLHRLGI	QAFQPVIVVEGQS---IQHLPLVCEAFNAFDGL-QMAVHVP			

**Figure 2-3: The Distant relationship between YonO, rdRNAP and the multisubunit RNAP. a) Amino acid shared between YonO and rdRNAP mapped on the structure of QDE-1.** The invariant amino acids shared amongst YonO and rdRNAPs were identified on the structure of the eukaryotic rdRNAP, QDE-1. Of the 12 residues, 10 were present on the two highly conserved DPBB folds. The DPBB folds corresponding to the  $\beta'$  and  $\beta$  DPBB are coloured red and blue, respectively. The residues conserved in YonO are coloured according to the DPBB which they map to. Those not present on DPBB folds are coloured orange. **b) Conservation of the aspartate triad in YonO, rdRNAP and multisubunit RNAPs.** Adapted from Iyer et al., 2003. Amino acid sequences corresponding to the  $\beta'$  DPBB fold in rdRNAPs, YonO and RNAP were aligned. QDE-1 from *N. crassa* was used to represent rdRNAP whilst  $\beta'$  subunits of *E. coli* and *T. thermophilus* were used for to represent the multisubunit RNAP. The absolutely conserved aspartate triad is shown in red. Secondary structure predictions show that the aspartate triad of YonO and rdRNAP is embedded in a region enriched in  $\beta$  strands. These predictions are shown above the corresponding amino acid sequences.

YonO has not been investigated beyond bioinformatics and so the enzymatic activity of this protein remains to be seen. Assuming it is an RNAP, it is not known whether YonO would function as an RNA dependant or DNA dependant RNAP. Eukaryotic rdRNAPs function in post transcription gene silencing. In addition to an RNAP, these complex systems require a cohort of additional proteins to bring about gene silencing (Iyer et

al., 2003). To date no homologues of these proteins have been identified in bacteriophages, suggesting that YonO does not function as an rdRNAP in a novel bacterial gene silencing system. SP $\beta$  is a DNA based bacteriophage, therefore it is likely that, despite resembling RNA dependant RNAPs, YonO utilises DNA as a template if it is indeed an RNAP. QDE-1 rdRNAP has been observed carrying out transcription utilising both RNA and DNA as a template, implying that the homology shared between YonO and rdRNAP does not preclude YonO from functioning as a DNA dependent RNAP (Aalto et al., 2010, Lee et al., 2009).

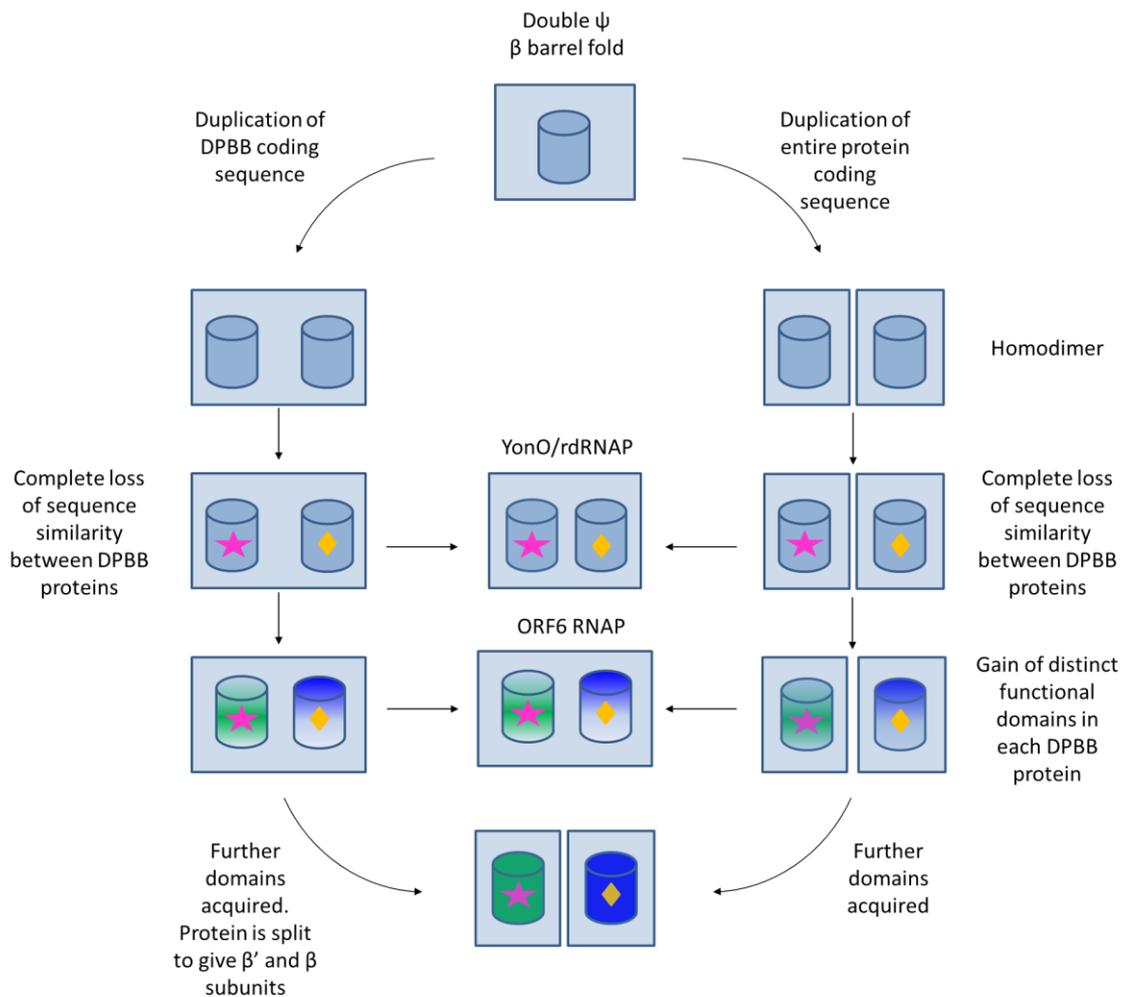
RNA based bacteriophages are known to utilise rdRNAPs for gene expression and replication. All single subunit bacteriophage RNAPs, including rdRNAPs, resemble the 'right handed' T7 RNAP and so do not feature an aspartate triad embedded on a DPBB fold (reviewed in (Ferrer-Orta et al., 2006). Therefore, YonO is potentially the first example of a single subunit bacteriophage RNAP equipped with the aspartate triad and resembling the multisubunit RNAP.

The presence of the aspartate triad suggests that YonO carries out nucleotide polymerisation using the same catalytic mechanism used by all RNAPs; chelating Mg<sup>2+</sup> ions and catalysing nucleophilic attack between the 3' hydroxyl of RNA and the  $\alpha$  phosphate of the substrate NTP. Besides this, nothing is known regarding YonO catalysis. According to amino acid sequence similarities, no other conserved multisubunit RNAP motifs are present in YonO, including the trigger loop. This is striking, as the trigger loop is critical for catalysis in multisubunit RNAPs. It is possible that the structure of functional motifs is conserved rather than the amino acid sequence.

The distant relationship between YonO to other RNAPs raises interesting questions regarding its evolution and occurrence in SP $\beta$ . YonO does not have increased similarity to an rdRNAP of a particular eukaryotic organism (Iyer et al., 2003). This undermines the hypothesis that the presence of a eukaryotic related RNAP in a bacteriophage is due to recent horizontal transfer. Rather, YonO and rdRNAPs may represent an ancient intermediate in the evolution of the multisubunit RNAP. Figure 2-2 shows the proposed origin of YonO in relation to the evolution of the multisubunit RNAP. It is possible the ancestor of YonO and rdRNAP may have diverged from the multisubunit

RNAP ancestor after the duplication of the DPBB folds but before the individual folds gained differing motifs and split (Iyer and Aravind, 2012, Iyer et al., 2003, Salgado et al., 2006). An implication of this hypothesis is that the ancestral YonO/rdRNAP in LUCA was lost in Archaea but kept in eukaryotes and certain bacteria. An alternative hypothesis is that the YonO/rdRNAP ancestor was completely lost by LUCA but was preserved in ancient bacteriophages and mobile elements (Iyer et al., 2003). Assuming the YonO/rdRNAP ancestor functioned in RNA metabolism in the RNA world, it would have become redundant upon the emergence of the DNA world allowing it to be lost over time. However, the YonO/rdRNAP could have been reintroduced at an early point in the eukaryotic lineage by a horizontal gene transfer event (Iyer et al., 2003).

In summary, it appears YonO is a putative RNAP of the *B. subtilis* bacteriophage SP $\beta$  which is only distantly related to rdRNAP and multisubunit RNAP. The identification of YonO led to the uncovering of eukaryotic rdRNAP-like proteins previously unseen in bacteriophages. Investigation into YonO and the transcription system it operates within could reveal an exciting, previously undescribed RNAP utilising a novel mechanism of catalysis. The identification of an uncharacterised single subunit RNAP is an exciting prospect. The discovery and subsequent characterisation of T7 RNAP opened up a whole new field within transcription, a field that has been intensely investigated over the past four decades (Chamberlin et al., 1970). Furthermore, exploitation of T7 RNAP in gene expression has completely revolutionised molecular biology by introducing a tractable transcription system for the expression of proteins.



**Figure 2-3: Proposed Origins of YonO and ORF6 RNAP.** Based on Salgado et al., 2006. Two possible explanations for the origin of the multisubunit RNAP, YonO and ORF6 RNAP are shown. The proposals are based on the conservation of two characteristic DPBB folds in multisubunit RNAP, rdRNAP, YonO and ORF6 RNAP (Iyer and Aravind, 2012, Iyer et al., 2003, Salgado et al., 2006, Ruprich-Robert and Thuriaux, 2010). **Left)** The original DPBB fold existed in a pre-LUCA organism, possibly in an RNA based world. The coding sequence for the DPBB was duplicated, giving rise to a single protein with two DPBB folds. This protein may have been a cofactor to a ribozyme RNAP. The two DPBBs began to diversify (represented by changes in colour), gaining catalytic motifs such as the aspartate triad (pink star) and conserved lysine residues (orange diamond) over extended periods of time. In an RNA world, such a protein may have become the primary rdRNAP enzyme. At this point, it is theorised YonO and rdRNAP may have diverged. If rdRNAP and YonO were expunged from the LUCA, they may have been preserved in ancient bacteriophages before reintroduction into the eukaryotic lineage. Alternatively they may have remained in LUCA and were subsequently lost in bacteria and archaea. The two DPBB continued to diversify and the RNAP they constitute became DNA dependent as the DNA world emerged. DNA dependent, single polypeptide RNAPs with two DPBB folds such as ORF6 RNAP may be remnants of this ancestral RNAP. Eventually, prior to the LUCA split, the DPBB folds split to produce two subunits which are the precursors of  $\beta'$  and  $\beta$ . Diversification removed all sequence similarity between  $\beta'$  and  $\beta$  DPBB folds. This RNAP would be the ancestor of all

multisubunit RNAPs. **Right)** An alternative origin of RNAPs also features a duplication of the DPBB fold. However, unlike the previous origin, the entire coding sequence of the DPBB fold protein may have been duplicated rather than only the DPBB coding sequence. This would result in a homodimeric complex rather than a single protein with two DPBB folds. Over time, the amino acid sequences of the two DPBB may have diversified, explaining the absence of sequence similarities between  $\beta'$  and  $\beta$  DPBB folds. Prior to further diversification, including the accretion of the functional domains identified in the multisubunit RNAP, YonO and rdRNAPs could have arisen due to the fusion of the two DPBB coding sequences. Between the divergence of YonO and the evolution of the LUCA multisubunit RNAP, a second fusion event may have occurred. This could explain the presence of single proteins, such as ORF6 RNAP which have increased sequence similarity to both RNAP subunits.

## Chapter 3: ORF6 RNAP of the *Kluyveromyces lactis* Killer System

### 3.1. Killer system of *Kluyveromyces lactis*

Killer yeast species have been identified in many yeast genera. Killer yeast secrete toxins which either kill or halt the growth of yeast competing in the same environment. These toxins are synthesised by the killer systems (reviewed in Magliani et al., 1997). In most cases the killer system also grants immunity to the killer yeast. Generally, the killer systems are defined as the proteins of the system, including the toxins, and the genetic determinants encoding them. The genetic determinants of killer systems vary across yeast species. In most cases the genetic material of the killer system is extra-chromosomal and has a cytoplasmic localisation. All three identified killer systems of the well-studied *Saccharomyces cerevisiae* are encoded by three different double stranded RNA of cytoplasmically inherited satellite viruses (Magliani et al., 1997). Another yeast with a killer system encoded by cytoplasmic nucleic acid is *Kluyveromyces lactis*. However, this system is encoded by two cytoplasmic DNA plasmids rather than a satellite virus.

*K. lactis* is a yeast species closely related to *Saccharomyces cerevisiae* and is characterised by its unusual ability to use lactose as its sole carbon source, hence its isolation from dairy products and nick name 'diary yeast' (reviewed in Schaffrath and Meacock (1996)). The *K. lactis* killer system is encoded by two linear double stranded DNA plasmids, pGKL1 and pGKL2 (discussed in section 2.1.1) (Gunge et al., 1981). As with other killer systems, pGKL1 and pGKL2 are located in the cytoplasm and encode the proteins of the toxin. The toxin, zymocin, is comprised of three subunits ( $\alpha$ ,  $\beta$  and  $\gamma$ ), with the  $\alpha$  and  $\beta$  subunits functioning to transport the effector subunit,  $\gamma$ , across the cell wall of susceptible yeast species (Stark and Boyd, 1986, Butler et al., 1991). The  $\gamma$  subunit of zymocin cleaves the anticodon of a subset of tRNA molecules, causing an arrest in cell growth (Auchtung et al., 2005)

### 3.2. Linear Plasmids of the *K. lactis* Killer System

pGKL1 and pGKL2 are 8.9 Kb and 13.4 kb long plasmids that encode the proteins of the *K. lactis* killer system (Gunge et al., 1981, Stark et al., 1984). The plasmids were first identified and isolated during a search for novel yeast plasmids (Gunge et al., 1981). In addition to being associated to the killer phenotype, the first investigation into pGKL1

and pGKL2 revealed that the plasmids have a linear structure. A subsequent study, aimed at increasing the efficiency of plasmid purification, revealed that pGKL1 and pGKL2 reside in the cytoplasmic fraction (Stam et al., 1986). As the DNA replication and transcription machinery resides in the nucleus, the cytoplasmic localisation of the plasmids implies they are self-maintained (this is discussed below).

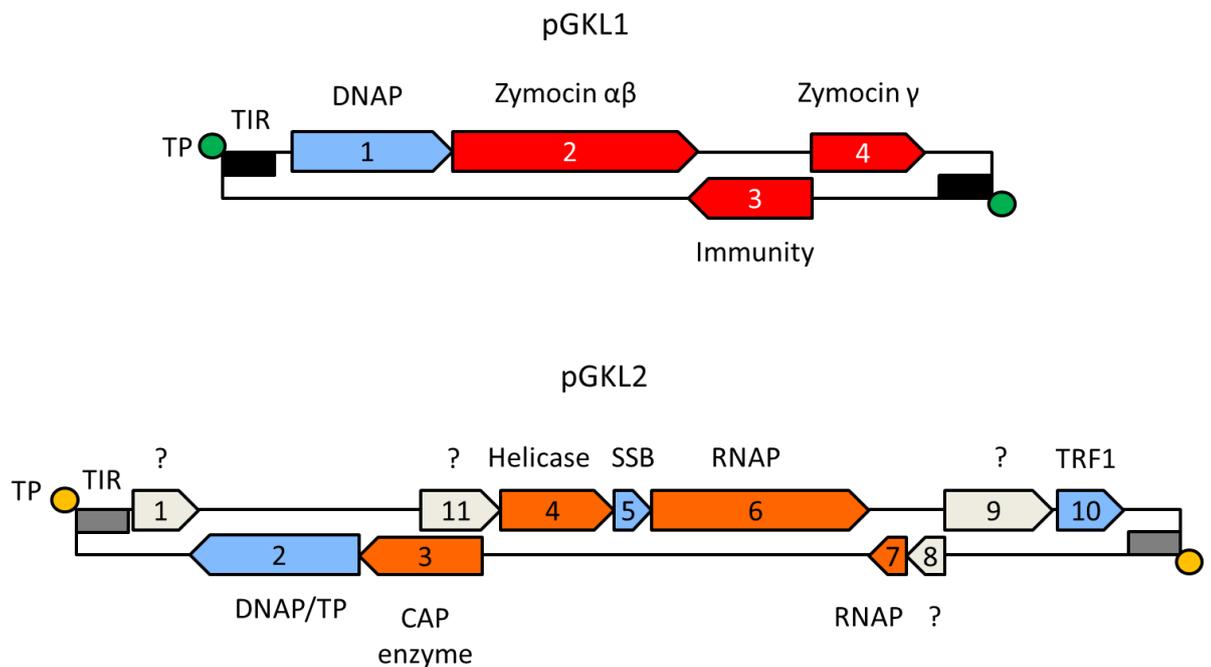
Further analysis of the plasmids by DNA sequencing revealed additional features. Both plasmids are extremely AT rich, having an AT content of ~74% (Hishinuma et al., 1984, Tommasino et al., 1988). Terminal inverted repeats (TIR) were identified at the 5' ends of both plasmids (Hishinuma et al., 1984). However, the TIRs differ in sequence and in length on each plasmid. The TIRs of pGKL1 and pGKL2 are 220 bp and 180 bp in length, respectively. The presence of TIRs strongly suggests a protein primed mode of DNA replication similar to that of the adenovirus and  $\phi$ 29 *Bacillus* phage (Mcneel and Tamanoi, 1991). Indeed, two terminal proteins (TP) covalently bound to the 5' termini of the plasmids have been observed. As with the TIRs, the two bound proteins were different for each plasmid, with 28 KDa and 36 KDa proteins binding the TIR of pGKL1 and pGKL2, respectively (Stam et al., 1986).

Sequencing allowed the identification of the open reading frames which encode for the proteins of the killer system (Hishinuma et al., 1984, Stark et al., 1984, Tommasino et al., 1988). Arrangement of the ORFs on the plasmid is very compact, with ORFs present on both strands of the DNA (Gunge et al., 1981). It was calculated that 95-97 % of the DNA sequence codes for protein (Tommasino et al., 1988). Investigating the proteins encoded by the ORFs on pGKL1 and pGKL2 revealed a clear organisation of the two plasmids based on the function of their encoded proteins. The ORFs encoding the zymocin subunit localised to pGKL1. Through amino acid sequence alignments, the hypothetical proteins of pGKL2 were given putative roles in plasmid replication and gene expression (see below) (Stark et al., 1990, Tommasino et al., 1988, Wilson and Meacock, 1988)

### *3.2.1: The pGKL1 Killer plasmid*

Initially, the gene encoding zymocin was linked to pGKL1 through genetic mutations which led to a loss of the killer phenotype (Wesolowski et al., 1982). Sequencing allowed the four ORFs present on pGKL1 to be identified (Figure 3-1) (Stark et al., 1984). The amino acid sequences of proteins encoded by ORF2 and ORF4 bore

similarities to the N-terminal secretion signal, leading to the conclusion that the two ORFs encode the toxin subunits. However, further studies indicated that the toxin consists of 3 subunits:  $\alpha$ ,  $\beta$  and  $\gamma$ . Amino acid sequencing of the individual toxin subunits revealed that ORF2 actually codes for two subunits,  $\alpha$  and  $\beta$ . ORF4 encodes for the active toxin subunit,  $\gamma$  (Stark and Boyd, 1986). Genetic studies revealed that ORF3 is required for self-immunity to the  $\gamma$  subunit of zymocin (Tokunaga et al., 1987). However, its mode of action has not yet been determined. The amino acid sequence of ORF1 gives further support to the hypothesis that the killer plasmid replicate via a protein primed replication. ORF1p (ORF1p refers to the protein encoded by ORF1, and unless stated otherwise, this nomenclature will be used to refer to proteins encoded by pGKL1 and pGKL2) has been assigned as a pGKL1 specific DNA polymerase (DNAP) based on its homology to the DNAPs operating in the protein-primed DNA replication of  $\phi$ 29 and the adenovirus (Jung et al., 1987, Fukuhara, 1987).



**Figure 3-1: Organisation of pGKL1 and pGKL2 plasmids of the *K. lactis* killer system.** Adapted from Schaffrath and Breunig, 2000. The organisation of the ORFs on pGKL1 and pGKL2 are shown. Hypothetical or confirmed functions are displayed adjacent to each ORF. The ORFs are coloured according to their predicted function. Red ORFs – toxin and immunity function. Blue – DNA replication. Orange – transcription. Grey – unknown. Terminal inverted repeats (TIR) at the 5' terminus of each plasmid are represented as black (pGKL1) and grey (pGKL2) boxes. Terminal proteins (TP) bound to the TIR are shown as green and yellow circles.

### 3.2.2: The pGKL2 Killer plasmid

pGKL2 has been shown to be essential for the maintenance of pGKL1, as pGKL1 cannot be isolated from strains lacking pGKL2 (Tommasino et al., 1988). Its localisation in the cytoplasm, away from the nuclear replication and transcription machinery, suggested that pGKL2 was responsible for encoding proteins which perform replication and gene expression. Sequencing of pGKL2 confirmed this hypothesis, as in addition to allowing the identification of 10 ORFs, it showed that the majority of these ORFs encode proteins homologous to replication and transcription machinery.

Initially 10 ORFs were identified on pGKL2. However, an updated search identified an additional ORF, raising the total number of ORFs present to 11 (Tommasino et al., 1988, Larsen and Meinhardt, 2000). The organisation of the pGKL2 ORFs is shown in Figure 3-1. ORFs 2 – 10 are essential for pGKL2 maintenance. ORF1 is dispensable and it has not yet been determined whether or not ORF11 is required (Larsen and Meinhardt, 2000, Schaffrath et al., 1992). 7 ORFs have been assigned putative functions based on amino acid sequence similarities and limited experimental evidence. The proteins can be categorised based on the two distinct processes in which they may function. ORF2p, ORF5p and ORF10p most likely function in DNA replication whilst ORF3p, ORF4p, ORF6p, ORF7p may carry out transcription (Romanos and Boyd, 1988, Tommasino et al., 1988, Wilson and Meacock, 1988, Mcneel and Tamanoi, 1991, Schaffrath and Meacock, 1995, Takeda et al., 1996, Larsen et al., 1998, Schaffrath and Meacock, 2001, Schaffrath et al., 1997). ORF1p, ORF8p, ORF9p and ORF11p do not have sequence similarity to any known proteins.

The amino acid sequences of proteins coded by pGKL2 suggests that the killer system utilises plasmid specific protein primed DNA replication, as ORF2p of pGKL2 was shown to be homologous to the DNAP encoded by ORF1 on pGKL1 (Tommasino et al., 1988). Additional observations support the hypothesis of plasmid specific replication machinery. The 5' TIR present on pGKL1 and pGKL2 bear no sequence similarity (Hishinuma et al., 1984). Furthermore, the terminal proteins covalently bound to the 5' TIR appear to be different. Whilst the TP bound to pGKL2 TIR has been identified as the protein product of pGKL2 ORF2, the ORF coding for pGKL1 TP has not been identified (Takeda et al., 1996). ORF2 of pGKL2 is unusual in that it codes for the TP and a DNAP. TP are utilised to prime DNA replication in many viruses, including the  $\phi$ 29 phage of *B.*

*subtilis*. The coding sequence of the  $\phi$ 29 phage TP is also present in the coding sequence of the  $\phi$ 29 specific DNAP.

If plasmid specific replication systems do exist, then ORF5p and ORF10p encoded by pGKL2 are common to both systems. As with the other replication proteins of the killer system, ORF5p is homologous to a protein of the  $\phi$ 29 bacteriophage, the p5 single strand DNA binding (SSB) protein. During  $\phi$ 29 DNA replication, p5 coats the displaced single stranded DNA. *In vitro* experiments show that ORF5p functions similarly to p5 (Schaffrath and Meacock, 1995, Schaffrath and Meacock, 2001). ORF10 encodes Terminal recognition factor 1 (TRF1). TRF1 specifically binds to the TIR DNA of both killer plasmids; however a binding consensus sequence was not identified (Mcneel and Tamanoi, 1991). TRF1 is possibly functionally similar to the P6 protein of  $\phi$ 29, which also binds to TIRs. However, P6 is considerably different to TRF1 in that it recognises and binds a specific DNA sequence. Both P6 and TRF1 are thought to be involved in the initiation of TP primed DNA replication.

In summary, it is clear that pGKL1 and pGKL2 replicate independent of the nuclear replication machinery. The replication system operating on the plasmids closely resembles that of viruses, in particular the  $\phi$ 29 bacteriophage.

The presence of a cytoplasmic transcription system operating on pGKL1 and pGKL2 was first hypothesised when nuclear transcription machinery failed to express the killer toxin (Romanos and Boyd, 1988, Wilson and Meacock, 1988). The whole native pGKL1 sequence was cloned into a circular plasmid which localises to the *K. lactis* nucleus. The constructed plasmid was then transformed into a *K. lactis* strain defective in toxin synthesis and assayed for killer activity. No killer activity was observed. Investigating the lack of toxin expression revealed truncated transcripts coding for  $\alpha$  and  $\beta$  subunits of zymocin. The most fitting explanation for these observations is an independent, cytoplasmic transcription system for gene expression of pGKL1 and pGKL2 ORFs (Romanos and Boyd, 1988, Tanguy-Rougeau et al., 1990, Wilson and Meacock, 1988).

The proteins accomplishing the transcription of pGKL1 and pGKL2 ORFs were identified through amino acid sequence homology. ORF3p, ORF4p, ORF6p and ORF7p encoded by their corresponding ORF on pGKL2 share similarity with proteins of previously

studied transcription systems (Wilson and Meacock, 1988, Schaffrath et al., 1997, Larsen et al., 1998). ORF3p is a 67 KDa protein which contains conserved residues from the three catalytic domains of the Vaccinia virus mRNA capping enzyme (Larsen et al., 1998). In eukaryotes, 5' 7-methylguanosine mRNA caps are necessary to stimulate translation and prevent degradation. Vaccinia virus resides in the cytoplasm throughout its life cycle and therefore utilises its own transcription system, including an mRNA capping enzyme (reviewed in Moss, 1990 ). *In vitro* assays demonstrated ORF3p has 5' RNA triphosphatase and guanylyltransferase activity (Tiggemann et al., 2001). The third enzymatic activity required for mRNA capping is methyltransferase, it was not determined if ORF3p has this activity due to the lack of a suitable assay. However, amino acid homology suggests ORF3p is able to perform this activity.

The similarities between the Vaccinia virus and the pGKL2 transcription system are not limited to ORF3p. ORF4p shares amino acid sequence similarity to two individual proteins of the Vaccinia virus, the D6R subunit of the ETF (Early Transcription Factor) and the NPH I (nucleoside triphosphate phosphohydrolase I) (Stark et al., 1990, Wilson and Meacock, 1988). Both proteins function in transcription and have DNA dependant ATPase activity (Broyles, 2003). The ETF recruits the Vaccinia encoded RNAP to DNA. In contrast to this, NPH I is a helicase involved in transcription termination. ORF4p has not been investigated experimentally and so it is unknown if it functions during initiation or termination of transcription.

Central to any transcription system is RNAP, which catalyses the synthesis of RNA. ORF6 of pGKL2 encodes a large 115 KDa protein with significant homology to the conserved multisubunit RNAP (Wilson and Meacock, 1988). Furthermore, the small 17 KDa protein product of ORF7 has a small region of weak homology to the  $\beta'$  subunit (Schaffrath et al., 1997). Therefore the current consensus is that ORF6p, potentially complexed with ORF7p, is the RNAP of the cytoplasmic transcription system of the *K. lactis* killer system. The homology between ORF6 RNAP and the multisubunit RNAP is unusual and raises many questions about how the enzyme functions. ORF6 RNAP has amino acid sequence similarity to both catalytic subunits ( $\beta'$  and  $\beta$  in bacteria, RPB1 and RPB2 in RNAP II). However, not all of the conserved domains which operate in RNAP during transcription appear to be present in ORF6 RNAP. This, in addition to the absence of recognisable nuclear promoters, further implies that the killer system

utilises a novel, previously uncharacterised transcription system featuring an unusual RNAP.

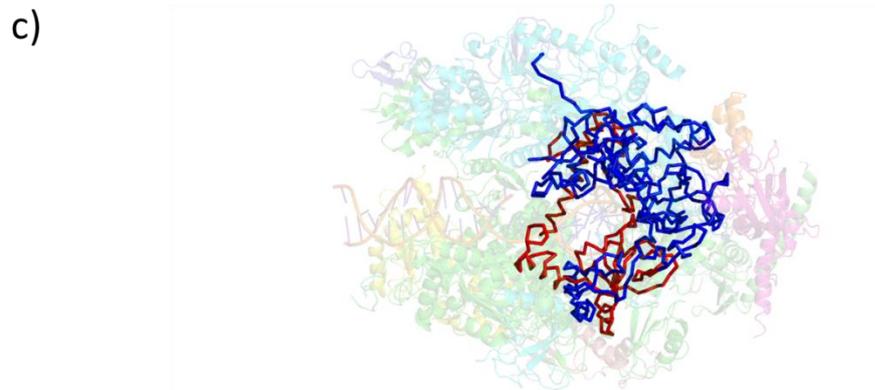
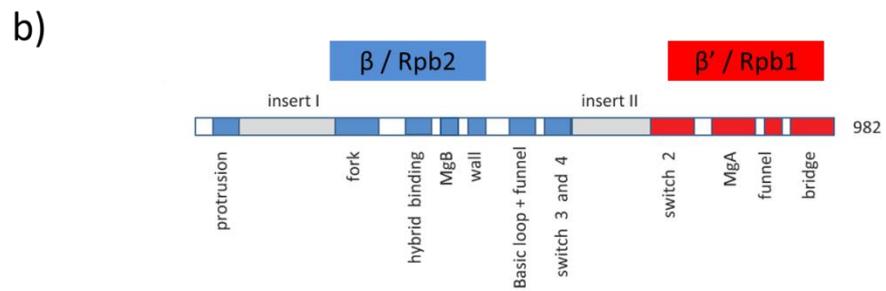
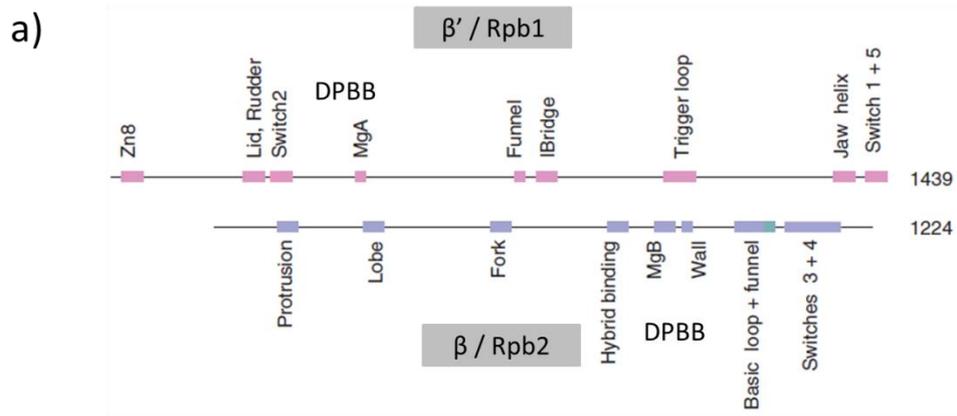
### **3.3: ORF6 RNAP**

ORF6 of pGKL2 encodes a 982 amino acid, 115 KDa protein. Upon searching the protein database with the amino acid sequence of ORF6p (referred to as ORF6 RNAP), Wilson and Meacock uncovered homology to the multisubunit RNAP (Wilson and Meacock, 1988). Unusually, ORF6 RNAP, a single polypeptide chain, has sequence similarity to both catalytic subunits of RNAP. At the time, this was considered unique as no other known single subunit RNAP bore resemblance to both  $\beta'$  and  $\beta$  (Stark et al., 1990). Other examples of such RNAPs are now known (Ruprich-Robert and Thuriaux, 2010). Upon first inspection, ORF6 RNAP had three small regions of homology, one region (denoted region A) homologous to  $\beta$  and two regions (denoted region II and III) homologous to  $\beta'$ . Region III contains the essential aspartate triad motif (Wilson and Meacock, 1988). In 2010 a greater level of conservation between ORF6 RNAP and the  $\beta'$  and  $\beta$  subunits was observed (Ruprich-Robert and Thuriaux, 2010). The homology is now known to extend to 250 amino acids that span the length of ORF6 RNAP. In the 2010 study, the authors defined 21 functional motifs present in the two large catalytic subunits ( $\beta'$  and  $\beta$  in bacteria) of all multisubunit RNAPs. These domains include the previously discussed switch region, trigger loop, bridge helix and aspartate triad (these domains are shown schematically in Figure 3-2). The 250 amino acids shared between ORF6 RNAP and the multisubunit RNAP corresponds to 11 of the 25 motifs (Figure 3-2) (Ruprich-Robert and Thuriaux, 2010). Approximately half of the motifs involved in transcription in all cellular organisms are absent in ORF6 RNAP. Incredibly, the trigger loop is absent from ORF6 RNAP. The trigger loop plays a pivotal role in all reactions catalysed by RNAP (Wang et al., 2006, Yuzenkova et al., 2010, Yuzenkova and Zenkin, 2010, Temiakov et al., 2005). Its absence, along with the majority of the  $\beta'$  motifs, raises interesting questions regarding the mechanism by which ORF6 RNAP carries out the nucleotide addition cycle and catalysis. It also raises the exciting possibility that ORF6 RNAP is a novel RNAP with a minimal architecture consisting of only the domains absolutely essential for transcription. The minimal nature of ORF6 RNAP, including its reduced size compared to the multisubunit RNAP, is illustrated in Figure 3-2. Alternatively, the functional motifs may not be conserved in the amino acid sequence

but may perhaps be conserved structurally. Two amino acid inserts are present in the ORF6 RNAP polypeptide sequence (Ruprich-Robert and Thuriaux, 2010). The residues in these inserts may fold to structurally resemble absent conserved motifs.

Modelling the homology shared between ORF6 RNAP and multisubunit RNAP onto the structure of RNAP II revealed two DPBB folds, a conserved feature of all multisubunit RNAPs (Ruprich-Robert and Thuriaux, 2010). In addition to further confirming that ORF6 encodes an RNAP, the presence of two DPBB fold led to speculation regarding the origin of ORF6 RNAP. ORF6 RNAP may be a representative of an intermediate ancestral RNAP, an RNAP in which the two DPBB folds had partially differentiated to gain different functional motifs, but not split into the individual proteins (Iyer and Aravind, 2012). Splitting of the DPBB folds into two proteins, further divergence and gain of motifs could account for the increase in functional motifs observed in the multisubunit RNAP. An alternative hypothesis is that ORF6 RNAP is a result of the two large RNAP subunits fusing together, with subsequent divergence and loss of the unessential functional motifs (Ruprich-Robert and Thuriaux, 2010).

ORF6 RNAP remains totally uncharacterised beyond the analysis of its amino acid sequence. However, the promoter elements driving transcription on pGKL1 and pGKL2 ORFs have been identified. Once it became clear that the leader sequences of pGKL1 transcripts were very short (~2 - 27 nucleotides) compared to nuclear mRNA (50 nucleotides on average), the DNA sequences upstream of each ORF were analysed (Stark et al., 1984). A 13 bp conserved sequence (ACT(A/T)AATATATGA), 35 bp upstream from the translation start codon of ORF1, ORF2 and ORF3 was identified. This DNA element is known as the upstream conserved sequence (UCS). An 11 bp UCS with reduced conservation is present 49 bp upstream of ORF4 (Romanos and Boyd, 1988).



**Figure 3-2: The conservation of multisubunit RNAP functional motifs in ORF6 RNAP.** Adapted from Ruprich-Robert and Thuriax, 2010. **a) Conserved motifs of the multisubunit RNAP.** The 21 conserved motifs of the two large catalytic subunits ( $\beta'$  and  $\beta$  in bacteria, Rpb1 and Rpb2 in eukaryotic RNAP II) are displayed along the polypeptide of their subunit. The amino acid length of each subunit is shown on the left. The approximate locations of the two DPBB folds are shown. **b) RNAP motif conservation in ORF6 RNAP.** The multisubunit RNAP motifs present in ORF6 RNAP are displayed along the ORF6 polypeptide chain. ORF6 RNAP has homology to both  $\beta'$  (Rpb1) and  $\beta$  (Rpb2). The motifs homologous to  $\beta'$  and  $\beta$  are coloured red and blue, respectively. **c) Structural representation of RNAP motifs conserved in ORF6 RNAP.** The motifs shown in b) were identified in the structure of the RNAP II elongation complex to demonstrate the minimal nature of ORF6 RNAP compared to the multisubunit RNAP. The motifs conserved in ORF6 RNAP are shown in ribbon representation and coloured as for b).

UCSs are also present on pGKL2 but they display marked differences to their pGKL1 counterparts (Tommasino et al., 1988). pGKL2 UCSs are shorter (~7 bp) and display greater variation with a consensus of TA(A/-)TNTGA. On pGKL2, transcription starts 10 – 17 bp downstream of the UCSs. However, the distance between the UCS and the translation start codon of the corresponding ORF can vary widely from 27 to 110 bp. Therefore, in contrast to pGKL1, pGKL2 ORFs have long 5' leader sequences. An UCS precedes every ORF on pGKL1 and pGKL2, suggesting that each ORF is transcribed independently and not as part of an operon. Transcripts for each ORF have been detected, confirming this (Romanos and Boyd, 1988, Stark et al., 1984, Sor and Fukuhara, 1985). Positioning of the glucose dehydrogenase reporter gene or kanamycin resistance cassette downstream of a UCS on the pGKL1 plasmid resulted in detectable activity and resistance, respectively (Tanguy-Rougeau et al., 1990, Schickel et al., 1996). This provided direct evidence that UCSs are genuine promoter elements directing expression in the *K. lactis* killer system.

In summary, ORF6 RNAP is regarded as the enzyme responsible for cytoplasmic gene expression in the *K. lactis* killer system. Approximately half of the conserved motifs in RNAP are not present in ORF6 RNAP according to amino acid sequence. Interestingly, one such motif is the trigger loop. Investigation into ORF6 RNAP may reveal an RNAP with a minimal architecture which updates our understanding of the basic requirements for catalysis. Alternatively, novel functional motifs may be uncovered, leading to the elucidation of undescribed mechanisms of catalysis. Furthermore ORF6 RNAP utilises UCSs, novel promoter elements to initiate transcription. The investigation of ORF6 RNAP and UCSs may lead to the discovery of a novel transcription system which can be exploited for the expression of AT rich ORFs. Aside from studies which have shown ORF6 RNAP to be essential for plasmid maintenance, no experimental investigations into ORF6 RNAP have yet been published (Schaffrath et al., 1995b). Therefore a primary aim of this thesis is to begin the characterisation of the ORF6 RNAP transcription system using an *in vitro* approach.

## Chapter 4: Materials and Methods

### 4.1. General Methods

#### 4.1.1. PCR

Each 50  $\mu$ l PCR reaction contained the following: 0.5  $\mu$ l DNA template (typically 100 ng  $\mu$ l<sup>-1</sup>), 1 unit of Phusion DNA polymerase (New England Biolabs), 0.4 mM of each primer, 200  $\mu$ M dNTPS mixed in High Fidelity Buffer (New England Biolabs). Sequences of all primers can be found in Table 4-1.

Each PCR was carried out as follows:

Temperature (°C)	Time (s)	Cycles
98	30	
98	9	30
variable	15	
72	30 s per kb	
72	300	
10	pause	

#### 4.1.2. Growth media and Strains

All bacterial and yeast strains used throughout this work are listed in Table 4-2. Unless stated otherwise, all bacterial strains were grown in Luria – Bertani (LB) medium supplemented with the required antibiotic which is stated in the relevant materials and methods. Yeast strains were grown in Synthetic Complete – uracil (SC-URA) medium supplemented with a specific carbon source detailed in the corresponding method. The composition of growth media can be seen in Table 4-3.

**Table 4-1: Primers used in this study.**

Oligonucleotide	Sequence (5' - 3')	Annealing temperature °C	Restriction Enzyme (If applicable)
yonO pET28 L	CATATGGCTAGCTTGAAAGGAAAAAAGACGG	58	NheI
yonO pET28 R	GTAGTTCTCGAGCTAACGGTGGTTAATTC		XhoI
<i>yonO</i> SDM L	GCTGCTGCAGTTTGACAATAACGGGAATAAGGCCTTAATTATTC	68	
<i>yonO</i> SDM R	GAAATAATTAAGGCCTTATCCCGTTATTGTCAAAGTGCAGCAGC		
pBEST:UP L	GAAATATGAAGCTTCTAGACTTGGTCTGAAAC	60	HindIII
pBEST:UP R	CATGTGTGGTCGACCCCTTATTACTTTTATGC		XhoI
pBEST:DOWN L	GAATATGGCGCCGCGTGCCTTGAGCCTTACTC	65	NotI
pBEST:DOWN R	CACACATGGGATCCCAAGCACAAAAGGCCTTC		BamHI
pBESTdx L	GATGGATCCACTATCATTTACTTTTCC	55	
pBESTdx R	CTCGTAAACACATTCCGGCTG		

pBEST seq L	GAGACTGCTATTATTGAAGGAACTG		
pBEST seq R	CTCCCTTATGACGATTATCTCC		
pBEST internal seq L	GATGTGTGTCATGTCAACAGAGG		
pBEST internal seq R	CCTCTGTTGACATGACACACATC		
1673 L	GCGACTCCAATCTAAATAAAAATTAAG	52	
1673 R	GTTTTGAACCCTTTAAAATCACTGG		
1673 L + 50	GTTCAGGTATCTTTTTGCTAGTTCTTC	52	
1673 R + 50	CTTCTTCATCCTCAATCGTTTTTTTG	52	
VM RNAP L	GCATATGGCTAGCTTGAAAGGAAAAAAGACGG	60	NheI
VM RNAP R	CATATATGCTCGAGCTAAATGCTAACAACCTTCTTTC		XhoI
ORF6 <sup>C+</sup> pYES L	GTACATGATCAGAGGATCCATGATCGTCTTGGACATCC	59	BamHI
ORF6 <sup>C+</sup> pYES2 R	CTATACGTATGCACTCGAGTCATAAGTACATTTTCGTTGC		XhoI
ORF7 pET21 L	GTACGTCATATGAATGAAAATATTATTC	53	NheI

ORF7 pET21 R	GTAATGAACTCGAGTAAAAAGAATATTTTCTCG		XhoI
ORF6 <sup>C+</sup> pESC L	GTCGTATGCAACTAGTAACACAATGGATTACAAGGATGACGACGATAAGATCATGATCGTCTTGGA CATCCC	68	SpeI
ORF6 <sup>C+</sup> pESC R	GGATGTCGAGGAGCTCTCATAAGTACATTTTCGTTGCAG		XhoI
ORF7 pESC L	GGATGTCGAGGGATCCAACACAATGAATGAAAATATTATTTTC	60	BamHI
ORF7 pESC R	TGCATACGACGTCGACTAAAAAGAATATTTTCTCG		Sall
ORF6 <sup>C+</sup> pACYC L	GAAATATTGGATCCGATGATCGTCTTGGACATC	58	BamHI
ORF6 <sup>C+</sup> pACYC R	CATATATGAAGCTTTCATAAGTACATTTTCGTTG		HindIII
ORF7 pACYC L	GAGCTATGCATATGAATGAAAATATTATTTTC	56	NdeI
ORF7 pACYC R	CATATATGCTCGAGTTATAAAAAAGAATATTTTC		XhoI

**Table 4-2: Strains constructed and used in this study**

Species	Strain	Genotype	Reference
<i>B. subtilis</i>	168 (WT)	<i>trpC2</i>	Kunst et al., 1997
	$\Delta$ <i>yonO</i>	<i>trpC2</i> $\Delta$ <i>yonO</i> :: neo <sup>R</sup>	This work
	CU1065	<i>trpC2</i> SP $\beta$ <sup>S</sup>	Zahler et al., 1977
	$\Delta$ SP $\beta$	<i>trpC2</i> $\Delta$ SP $\beta$	Dorenbos et al., 2002
<i>B. vallismortis</i>	DV1-F-3	Wildtype isolate	Roberts et al., 1996
<i>E. coli</i>	T7 Express	<i>fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--Tet<sup>S</sup>)2 [dcm] R(zgb-210::Tn10--Tet<sup>S</sup>) endA1 <math>\Delta</math>(mcrC-mrr)114::IS10</i>	New England Biolabs
	DH5 $\alpha$	<i>fhuA2 <math>\Delta</math>(argF-lacZ)U169 phoA glnV44 <math>\Phi</math>80 <math>\Delta</math>(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>	New England Biolabs
	BL21-CodonPlus (DE3)-RIPL	F <sup>-</sup> <i>ompT hsdS(r<sub>s</sub><sup>-</sup> m<sub>s</sub><sup>-</sup>) dcm<sup>+</sup> Tet<sup>r</sup> gal <math>\lambda</math> (DE3) endA Hte (argU proL Cam<sup>r</sup>) (argU ileY leuW Strep/Spec<sup>r</sup>)</i>	Agilent
<i>K. lactis</i>	AWJ137	MAT $\alpha$ <i>leu2 trp1</i> [pGKL1 <sup>+</sup> pGKL2 <sup>+</sup> ]	Kämper et al., 1991
<i>S. cerevisiae</i>	INVSc1	<i>his3<math>\Delta</math>1/his3<math>\Delta</math>1 leu2/leu2 trp1-289/ trp1-289 ura3-52/ura3-52;</i>	Invitrogen
	YPH449	MAT $\alpha$ <i>ura3-52 lys2-801<sup>amber</sup> ade2-101<sup>ochre</sup> trp1-<math>\Delta</math>63 his3-<math>\Delta</math>200 leu2-<math>\Delta</math>1</i>	Agilent

**Table 4-3: Composition of growth media**

Media	Concentration / Amount	Component
Luria - Bertani	1%	Tryptone
	0.50%	Yeast extract
	1%	NaCl
Spizizen Minimal Salts (Anagnostopoulos and Spizizen, 1961)	1.4%	K <sub>2</sub> HPO <sub>4</sub>
	0.6%	KH <sub>2</sub> PO <sub>4</sub>
	0.1%	Na <sub>3</sub> C <sub>6</sub> H <sub>5</sub> O <sub>7</sub>
MM (Minimal Media)	Spizizen minimal salts supplemented with:	
	1ug ml <sup>-1</sup>	FeNH <sub>4</sub> citrate
	0.50%	Glucose

	6mM	MgSO <sub>4</sub>
	0.02%	Casamino acids
MMB (Modified M Broth)	1%	Bacto-tryptone
	0.50%	Bacto-yeast extract
	1%	NaCl
	5mM	MgCl <sub>2</sub>
	0.1mM	MnCl <sub>2</sub>
Nutrient Agar	0.50%	Peptone
	0.30%	Yeast Extract
	2.00%	Agar
	0.50%	NaCl
SC- URA	1 x	yeast nitrogen base (formedium)
	1 x	SC-URA dropout supplement (Formedium)
	2%	Glucose
SC-URA agar	1 x	yeast nitrogen base (formedium)
	1 x	SC-URA dropout supplement (Formedium)
	2%	Glucose
	2%	Agar
SC-URA:gal/raf	1 x	Yeast nitrogen base (formedium)
	1 x	SC-URA dropout supplement (Formedium)
	2%	galactose
	1%	raffinose
Starvation Media	Spizizen minimal salts supplemented with:	
	0.50%	Glucose
	6mM	MgSO <sub>4</sub>

#### 4.1.2. Molecular Cloning

All insert DNA fragments for cloning were generated through PCR (see section 4.1.1). The Primers and restriction enzymes used can be seen in Table 4-1. Amplified insert DNA was purified using the Qiagen Gel Extraction kit, following the protocol provided. Plasmid and insert DNA were restricted using the appropriate restriction enzymes (see Table 4-1) for 1 hour at 37 °C before being purified as before. For ligations, 3 times molar excess of insert DNA to plasmid DNA was mixed and ligated overnight at 15 °C using T4 DNA ligase (New England Biolabs). The following day, 5 µl of ligation mix was transformed into 50 µl of DH5α competent *E. coli* cells (New England Biolabs). Cells were transformed according to the supplied protocol. Cells were plated on nutrient agar containing the selective antibiotic, either 100 µg ml<sup>-1</sup> ampicillin or 50 µg ml<sup>-1</sup> kanamycin.

To identify positive clones, colonies were picked and grown overnight in LB supplemented with the selective antibiotic. Plasmid DNA was isolated using the Qiagen Mini Prep Kit and restricted using the same restriction enzymes used for the cloning. The digest was resolved on a 1% agarose gel to determine which plasmids contained the insert. Positive plasmids were then sent for sequencing. All sequencing was carried out by GATC using their standard primers or the primers used for molecular cloning. Resultant sequences were visualised using the software Vector NTI (Invitrogen).

All plasmids used and constructed can be seen in Table 4.

**Table 4-4: Plasmids used and constructed throughout this work**

Plasmid	Source	Antibiotic Marker
pBEST501	Itaya et al., 1989	Kanamycin
pBEST:U:D	This work	Kanamycin
pET-21	Novagen	Ampicillin
pET-21: yonO <sup>STOP</sup>	This work	Ampicillin
pET-28	Novagen	Kanamycin
pET-28: yonO	This work	Kanamycin
pYES2/NT	Novagen	Ampicillin
pYES2/NT:ORF6 <sup>C+</sup>	This work	Ampicillin
pESC/URA	Agilent	Ampicillin
pESC-URA:O6 <sup>C+</sup> :O7	This work	Ampicillin
pACYC-Duet	Noavgen	Chloramphenicol
pACYC:ORF6 <sup>C+</sup> :ORF7	This work	Chloramphenicol

#### 4.1.3. Buffers

Composition of all buffers used can be seen Table 4-5. All chromatography buffers were filtered through 0.45 µm pore PVDF filters (Helena) prior to use. Where available, the compositions of commercial buffers are listed in Table 4-6.

**Table 4-5: Composition of buffers.** Buffers are listed in alphabetical order

Buffer	Concentration	Component
Acetate buffer	100mM	NaAcetate pH 4.8
	10mM	MgCl <sub>2</sub>
	500mM	NaCl
	0.10%	Triton X-100
Binding buffer	10mM	Tris - HCl pH 6.8
	100mM	NaAcetate pH 4.8
	10mM	MgCl <sub>2</sub>
	50mM	NaCl
Blocking buffer	200mM	Tris - HCL pH 8
	10mM	MgCl <sub>2</sub>
	500mM	NaCl
	0.10%	Triton X-100
Breaking buffer	50mM	Sodium phosphate buffer
	50%	Glycerol
	1mM	PMSF
<i>B. subtilis</i> lysis buffer	20mM	Tris-HCl pH 8
	2mM	EDTA
	1.20%	Triton X-100
Buffer I	10mM	Tris - HCl pH 7.2
	10mM	MgCl <sub>2</sub>
	1M	NaCl
	0.10%	Triton X-100
Buffer II	10mM	Tris - HCl pH 7.2
	10mM	MgCl <sub>2</sub>
	150mM	NaCl
	0.10%	Triton X-100
Coupling buffer	100 mM	NaHCO <sub>3</sub> pH 8.3
	10mM	MgCl <sub>2</sub>
	500mM	NaCl
	0.10%	Triton X-100
Developing solution	235mM	Sodium Carbonate
	0.02%	37% w/v Formaldehyde
Dilution buffer	10mM	Tris - HCl pH 7.4
	0.10%	Triton X-100
Elution buffer I	100 mM	Glycine pH 2
	0.10%	Triton X-100
Elution buffer II	2M	Tris-HCl pH 8
Fixation solution	40%	Ethanol
	10%	Acetic acid
Gel Filtration buffer	50mM	Tris - HCL pH 7.9
	500mM	NaCl

Grinding buffer	50mM	Tris-HCl pH 7.9
	200mM	NaCl
HEPES storage buffer	20 mM	HEPES KOH pH 7.9
	50%	Glycerol
	0.5mM	EDTA
	1mM	$\beta$ mercaptoethanol
Laemmli buffer	100 mM	Tris - HCl pH 6.8
	2%	SDS
	20%	Glycerol
	4%	$\beta$ -mercaptoethanol
ORF6 RNAP Heparin buffer	10 mM	Tris-HCl pH 7.9
	5%	Glycerol
	0.1 mM	EDTA
	0.1 mM	DTT
ORF6 RNAP Heparin elution buffer	10 mM	Tris-HCl pH 7.9
	5%	Glycerol
	0.1 mM	EDTA
	0.1 mM	DTT
	1M	NaCl
ORF6 RNAP Ni <sup>2+</sup> buffer	20 mM	Tris-HCl pH 7.9
	600 mM	NaCl
	5%	Glycerol
ORF6 RNAP Ni <sup>2+</sup> elution buffer	20 mM	Tris-HCl pH 7.9
	600 mM	NaCl
	5%	Glycerol
	200mM	Imidazole pH 8
ORF7p Storage buffer	20mM	Tris - HCL pH 7.9
	50%	Glycerol
	200mM	KCl
pESC Breaking buffer	50 mM	Tris-HCl pH 7.4
	150 mM	NaCl
	1mM	EDTA
	1%	Triton X-100
Sensitising solution	30%	Ethanol
	4.0%	5% w/v Sodium thiosulphate
	830mM	Sodium acetate
Silver solution	10%	2.5% w/v Silver nitrate
	0.04%	37% w/v Formaldehyde
Transcription Buffer	20mM	Tris-HCl pH 8.5
	40mM	KCl
	10mM	MgCl <sub>2</sub>
Tris Buffered Saline	50 mM	Tris-HCl pH 7
	150 mM	NaCl

Western blot buffer	1 X	Phosphate buffered saline (Oxoid)
	0.10%	Tween
Western blocking buffer	1 X	Phosphate buffered saline (Oxoid)
	0.10%	Tween
	5%	Milk powder
Western transfer buffer	1 X	Tris- glycine (Fisher Scientific)
	0.20%	SDS
	20%	Methanol
YonO Heparin buffer	10mM	Tris-HCl pH 7.8
YonO Heparin elution buffer	10mM	Tris-HCl pH 7.9
	1M	NaCl
YonO Ni <sup>2+</sup> buffer	20mM	Tris-HCl pH 7.9
	600mM	NaCl
YonO Ni <sup>2+</sup> elution buffer	20mM	Tris-HCl pH 7.9
	600mM	NaCl
	200mM	Imidazole pH 8
YonO Storage buffer	20mM	Tris-HCl pH 7.9
	200mM	KCl
	50%	Glycerol
	1mM	DTT
	0.1mM	EDTA

**Table 4-6: Composition of Commercial buffers.**

Commercial Kit	Buffer	Concentration	Component
Qiagen Miniprep Kit	P1	50 mM	Tris - HCl pH 8
		10 mM	EDTA
		100 $\mu\text{l ml}^{-1}$	RNase A
	P2	200 mM	NaOH
		1%	SDS
	N3	5 M	Guanidinium chloride
0.9 M		Potassium acetate pH 4.8	
Qiagen Gel Extraction Kit	QG	5.5 M	Guanidine thiocyanate
		20 mM	Tris - HCl pH 6.6
	PE	10 mM	Tris - HCl pH 7.5
		80%	ethanol

## 4.2. Protein Methods

### 4.2.1. Purification of YonO

DNA encoding for YonO was amplified from *B. subtilis* genomic DNA using primers yonO pET28 L and yonO pET28 R and cloned into pET-28a. An overnight culture of the T7 express *E. coli* strain (New England Biolabs) carrying pET-28:yonO was used to inoculate 12 L of LB media at a ratio of 1:1000 and grown at 37 °C until an OD<sub>600</sub> of 0.4 was reached. At this point the temperature was shifted to 18 °C and IPTG (Melford) was added to a final concentration of 1mM. Cultures were grown overnight before being harvested by centrifugation using a Beckman JLA 8.1000 at 6000 rpm for 10 minutes. Cell pellets were resuspended in 120 ml of grinding buffer and sonicated on ice by a Braun sonicator at 60% amplitude, 2 seconds on, 2 seconds off for 9 minutes in total. Lysates were clarified by centrifugation at 20 000 rpm using a JA25.50 rotor followed by filtration through a 0.45 µM PVDF filter (Merck-Milipore). Prior to FPLC, imidazole pH 8 was added to the lysate to a final concentration of 20 mM.

The lysate was applied to a 5 ml His-trap Ni<sup>2+</sup>-NTA column (GE Healthcare) equilibrated in grinding buffer at a flow rate of 5 ml min<sup>-1</sup> using the p900 pump on the AKTA Explorer FLPC (GE Healthcare). The column was washed with Ni<sup>2+</sup> elution buffer containing 25 mM imidazole to remove any unbound proteins. YonO was eluted by washing the column with Ni<sup>2+</sup> elution buffer containing 100 mM imidazole followed subsequently with 200 mM imidazole. SDS-PAGE was used to determine which elution fraction contained YonO.

The elution fraction containing YonO were pooled and applied to a 5 ml HiTrap Heparin HP column (GE Healthcare) equilibrated in 600 mM NaCl Heparin Elution Buffer. Contaminating proteins were eluted by washing the column with 700 mM NaCl Heparin Elution Buffer. This was followed by elution using a shallow gradient of 700 mM to 1M NaCl in Heparin Elution Buffer over 2 hours at a flow rate of 2 ml min<sup>-1</sup>. The elution fraction containing YonO (determined by SDS PAGE) was concentrated using a Centricon device with a 30 000 MW cut off (Merck-Millipore). The concentrated protein was clarified by centrifugation at 15 000 rpm at 4 °C and applied to a Superdex 200 gel filtration column (GE Healthcare) equilibrated in gel filtration buffer. Gel filtration buffer was run through the column at a rate of 1 ml min<sup>-1</sup>. The absorbance of the eluate at 280 nm revealed that YonO eluted after 70 ml of buffer had passed

through the column; this was confirmed by SDS PAGE. Elution fractions which contained YonO were pooled and concentrated as previously described.

Purified YonO intended for use in screening for crystallisation conditions was concentrated to 10 mg ml<sup>-1</sup> and applied to the condition screens. For biochemistry, purified YonO was dialysed into YonO storage buffer overnight and stored at -20 °C.

#### 4.2.2. Purification of ORF6 RNAP by Ni<sup>2+</sup> NTA chromatography

The DNA sequence of ORF6 was codon optimised for expression in *Saccharomyces cerevisiae*. The optimised sequence, denoted as ORF6<sup>C+</sup> was synthesised by Eurogentec and was provided cloned into pUC57. ORF6<sup>C+</sup> was amplified using primers ORF6<sup>C+</sup> pYES L and ORF6<sup>C+</sup> pYES R and cloned into pYES2/NT.

The construct pYES2/NT:ORF6<sup>C+</sup> was transformed into *S. cerevisiae* strain INVSc1 using the standard lithium acetate protocol (Gietz et al., 1992). Positive transformants were selected for by plating onto SC-URA agar medium. The transformant colonies were purified by restreaking individual colonies on SC-URA agar medium.

An INVSc1 transformant was grown overnight at 30 °C in SC-URA medium. The overnight culture was used to inoculate 1.5 litres of SC-URA which was grown until an OD<sub>600</sub> of 3 was reached. At this point, the culture was pelleted and washed in SC-URA:gal/raf medium before being used to inoculate 20 litres of SC-URA:gal/raf media to an OD<sub>600</sub> of 0.3. Cultures were grown at 30 °C overnight before being harvested by centrifugation (6000 rpm for 10 minutes in Beckman JLA8.1000 rotor) at stored at -80 °C until required.

The weight of the pelleted cells was measured and cells were resuspended in 10 times as much pYES2 breaking buffer. The resuspended cells were French pressed twice at 35 Kpsi. The lysate was clarified by centrifugation at 16 000 rpm for 20 minutes at 4 °C in a Beckman JA25.50 rotor. The lysate was made to 5mM imidazole and 600mM NaCl and filtered through a 8µM pore filter (Whatman) before being loaded onto a 10 ml Ni<sup>2+</sup> - NTA column (GE Healthcare) at a rate of 5 ml min<sup>-1</sup> using AKTA Explorer FLPC (GE Healthcare). Ni<sup>2+</sup> - NTA column was equilibrated in ORF6 RNAP Ni<sup>2+</sup> buffer. Once loaded, the column was washed with ORF6 RNAP Ni<sup>2+</sup> elution buffer containing increasing concentrations of imidazole (5 mM, 25 mM, 50 mM, 100 mM and 200 mM). The protein content of each elution fraction was visualised by SDS PAGE (sodium

dodecyl sulphate polyacrylamide gel electrophoresis). Peptide mass fingerprint (PMF) carried out by Pinnacle proteomics (Newcastle University) was used to confirm the presence of ORF6 RNAP in the 200 mM imidazole elution fraction. This fraction was concentrated using a Centricon device with a 10 000 MW cut off (Merck-Millipore) and made to 50% glycerol before storage at -20 °C.

#### 4.2.3. Co-expression and purification of ORF6 RNAP and ORF7p in *S. cerevisiae*

ORF6<sup>C+</sup> and ORF7 DNA fragments were amplified from pUC57 and isolated pGKL2 using primer pairs ORF6<sup>C+</sup>pESC L + ORF6<sup>C+</sup> pESC R and ORF7 pESC L + ORF7 pESC R respectively. pGKL2 DNA was obtained from *Kluyveromyces lactis* strain AWJ137 grown overnight in YPD media. Cells were harvested and resuspended in 250 µl of buffer P1 of the Qiagen Miniprep kit. 100 µl of acid washed glass beads and vortexed for 5 minutes. pGKL2 was then purified according to the Qiagen miniprep manufacturer's instructions. The fragments were sequentially cloned into pESC – URA. The resulting construct, pESC-URA:ORF6<sup>C+</sup>:ORF7 was transformed in YPH499 *S. cerevisiae* strain.

YPH499 carrying the pESC-URA:ORF6<sup>C+</sup>:ORF7 construct was grown and induced by SC-URA galactose as in previous section (4.2.1.1). Cells were harvested by centrifugation in JLA 8.1000 rotor (Beckman) at 6000 rpm for 10 minutes. Pellets were weighed and resuspended in 10 x volume of pESC breaking buffer and French pressed twice at 35 Kpsi. Lysates were clarified by centrifugation at 16 000 rpm in a JA25.50 rotor. The lysate was made to 600 mM NaCl and the pH was adjusted to pH 7 through gradual addition of 1% NaOH. Lysate was filtered through 0.2 µm filter (Helena Biosciences) before 1 ml of Anti – FLAG M2 (Sigma Aldrich) resin was added. The lysate and resin were incubated at 4 °C for 90 minutes with gentle shaking. Resin was collected on filter paper disc (Whatman) by filtration through a Buckner flask. Resin was washed in a total of 90 ml Tris Buffered Saline pH 7 (TBS) by resuspension followed by filtration. The resin was resuspended in 10 ml TBS and decanted into a gravity flow column (Bio Rad). Once the resin had settled, protein was eluted by 6 x 1 ml aliquots of glycine pH 2. Eluate was collected into 2 ml microcentrifuge tubes containing 25 µl Tris – HCl pH 8. The protein content of the eluate was visualised through SDS PAGE followed by silver staining.

#### 4.2.4. Purification of ORF7p

pGKL2 DNA was used as template DNA for the amplification of ORF7 with primers ORF7 pET21 L and ORF7 pET21 R. pET-21: ORF7 was transformed into the T7 express *E. coli* strain (New England Biolabs). All growth was done in the presence of 100  $\mu\text{g ml}^{-1}$  ampicillin. An overnight culture of T7 express *E. coli* carrying pET-21: ORF7 was used to inoculate 6 L of LB media at a ratio of 1:1000. Cultures were grown at 37 °C with shaking until an OD<sub>600</sub> of 0.4 was reached. 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG, Melford) was added to cultures which were then left to grow for a further 3 hours. Cells were harvested by centrifugation in a Beckman JLA 8.1000 rotor at 6000 rpm for 10 minutes. Isolation of ORF7 p inclusion bodies was based on the method presented by Borukhov and Goldfarb (1993). Cell pellets were resuspended in 40 ml of grinding buffer supplemented with 0.2 mg ml<sup>-1</sup> lysozyme (Fluka) and 0.2% sodium deoxycholate. Cells were incubated on ice for 20 minutes before disruption by a Braun sonicator for 3 minutes (2 seconds on, 2 seconds off) at 60 % amplitude. Inclusion bodies were pelleted by centrifugation in a Beckman JA25.50 rotor at 15 000 rpm for 20 minutes at 4 °C. Inclusion bodies were resuspended in 40 ml grinding buffer with 0.2 mg ml<sup>-1</sup> lysozyme and 0.2 % n-octyl  $\beta$ -D-glucopyranoside (Sigma Aldrich) and sonicated as before. After centrifugation, pelleted inclusion bodies were resuspended in 40 ml grinding buffer with 0.2 % n-octyl  $\beta$ -D-glucopyranoside and sonicated. Isolated inclusion bodies were dissolved in 30 ml 8 M urea followed by clarification by centrifugation for 20 minutes at 15 000 rpm. Dissolved ORF7p was subjected to syringe driven Ni<sup>2+</sup> - NTA chromatography under denaturing conditions (8M urea). ORF7p was loaded onto a 5 ml Ni<sup>2+</sup> - NTA column (GE Healthcare) pre-equilibrated with 8M urea. The bound proteins were eluted from the column by washing with increasing concentrations of imidazole (0 mM, 25 mM, 100 mM and 200 mM). SDS-PAGE analysis revealed that ORF7p eluted at 25 mM imidazole. Therefore, this elution fraction was diluted 16 x in 8M urea and dialysed overnight against ORF7p storage buffer.

#### 4.2.5. Co-expression and purification of ORF6 RNAP and ORF7p in *E. coli*

ORF6<sup>Ct</sup> and ORF7 DNA fragments were amplified from pUC57 and pGKL2 using primer pairs ORF6<sup>Ct</sup> pACYC L + ORF6<sup>Ct</sup> pACYC R and ORF7 pACYC L + ORF7 pACYC R, respectively. The ORF6<sup>Ct</sup> and ORF7 PCR fragments were sequentially cloned into pACYC duet vector. An overnight culture of T7 Express *E. coli* cells (New England Biolabs)

carrying the pACYC:ORF6<sup>C+</sup>:ORF7 construct was used to inoculate 2 litres of LB supplemented with 33 µg ml<sup>-1</sup> chloramphenicol. The cultures were grown with shaking at 37 °C until an OD<sub>600</sub> of 0.4 was reached, at which point the cultures were shifted to 18 °C and grown overnight.

Disruption of the cells and the following Ni<sup>2+</sup>-NTA chromatography were performed as described in section 4.2.1, with the following alterations. Harvested cells were resuspended in 40 ml of grinding buffer and the YonO Ni<sup>2+</sup> buffer was exchanged for ORF6 RNAP Ni<sup>2+</sup> buffer with a NaCl concentration of 300 mM. According to SDS PAGE, ORF6 RNAP was present in the 100 mM imidazole elution fraction. This fraction was diluted 1 in 4 with ORF6 RNAP heparin buffer and subjected to Heparin affinity chromatography. The diluted protein was loaded onto a 5 ml Heparin HiTrap column (GE Healthcare) pre-equilibrated with ORF6 RNAP heparin buffer containing 100 mM NaCl using the p900 pump on the AKTA Explorer FLPC system (GE Healthcare) at a flow rate of 2 ml min<sup>-1</sup>. Protein was eluted by washing the column with ORF6 RNAP heparin elution buffer containing increasing concentration of NaCl (100 mM, 300 mM, 600 mM and 1 M NaCl). The protein content of each elution fraction was determined by SDS PAGE. The elution fractions were made to 50% glycerol and stored at -20 °C.

#### 4.2.6. SDS PAGE

Proteins were suspended in Laemmli buffer, boiled for approximately 3 minutes and loaded onto a precast 4-12% SDS PAGE gel (Expedeon). SDS PAGE gels were ran in Rapid Run buffer (Expedeon) at 150 v for 45 minutes. Gels were stained overnight with the coomassie blue based Instant Blue stain (Expedeon).

#### 4.2.7. Silver stain of SDS PAGE gel

Proteins were resolved on SDS PAGE as above. Gels were incubated for 30 minutes in fixation solution. Fixation was followed by 30 minute incubation in sensitising solution. The gel was washed 3 times in distilled water for a total of 15 minutes. Water was decanted off and the gel was incubated for 20 minutes in silver solution. After a brief 2 minute wash with water, protein was visualised through incubation in the developing solution. Development was stopped after protein became visible by the removal of developer and washing in distilled water.

#### 4.2.8. Western Blot

Lysates were run on a 4-12% SDS-PAGE gel (Expedeon) as described above. The SDS-PAGE gel was blotted onto Amersham Hybond – P membrane (GE Healthcare) using wet transfer apparatus (Bio-Rad). The membrane was equilibrated in transfer buffer (1X Tris-Glycine, 0.2% SDS and 20% methanol) for 10 minutes prior to blotting. The membrane was washed with western blot buffer before being left shaking in western blocking buffer for 1 hour at room temperature or overnight at 4°C. The blocking buffer was replaced with 5 ml of fresh blocking buffer (2% milk powder) containing the primary antibody (1 in 1000 dilution for  $\alpha$ -YonO antibody) was left incubating at room temperature for 1 hour. The membrane was washed thoroughly with blot buffer for a total of 20 minutes before blocking buffer (2% milk powder) containing the secondary antibody (1 in 10000 dilution. Goat anti rabbit HRP conjugate, Sigma Aldrich) was left shaking with the membrane for 1 hour. The membrane was washed with blot buffer for 30 minutes. The ECL reaction (GE Healthcare) was carried out following manufactures' instructions and exposed using an Image Quant gel doc system (GE Healthcare).

#### 4.2.7. Affinity purification of $\alpha$ -YonO Antibody from Serum

YonO was expressed and purified as above. The N-terminal His tag was removed using thrombin cleavage capture kit (Novagen). Thrombin cleavage was performed following the provided manufacturer's instructions with 1 unit of biotin tagged thrombin per 1 mg of purified YonO (10 mg in total). Cleavage reactions were incubated at room temperature for 16 hours. Visible precipitation formed during incubation, this was removed by centrifugation. Biotinylated thrombin was removed through immobilisation on streptavidin agarose beads. To separate tag free and N-terminal His tag YonO, soluble YonO was loaded onto 1 ml Ni<sup>2+</sup>-NTA column. YonO present in the flow through and 0 mM imidazole wash was pooled, made to 50 % glycerol and sent to the Eurogentec to raise antibodies. Antibodies were raised by immunising 2 rabbits over 28 days. Serum from preimmune bleed, middle bleed (3 weeks after immunisation) and final bleed (4 weeks after immunisation) was supplied.

Antibodies were purified by affinity chromatography with cyanogen bromide (CNBr) activated sepharose coupled to YonO. YonO was purified as before except Tris-HCl pH

7.9 was replaced with HEPES-KOH pH 7.9 in all purification buffers. Following purification, thrombin cleavage capture kit was used to remove the N-terminal His tag.

Coupling of YonO to CNBr and subsequent purification was adapted from Banzhaf et al. (2012). 0.8g of CNBr activated sepharose (Sigma Aldrich) was poured into a gravity flow column (Bio-rad) and washed with 200 ml of 1 mM HCl for 15 minutes. 4 mg YonO was added to 10 ml coupling buffer. This was in turn incubated with the CNBr sepharose beads overnight with mixing. Resin was washed with 25 ml coupling buffer. The uncoupled active groups of the CNBr sepharose were blocked through overnight incubation with 10 ml blocking buffer. The following day, the resin was washed with 20 ml of acetate buffer and 20 ml of blocking buffer. This washing was repeated a total of three times. The column was then washed and stored in 5 ml binding buffer. When required for use, the column was initially washed with 1 column volume (10 ml) of elution buffer I to remove weakly bound YonO and prevent contamination of subsequently eluted antibodies. Immediately after, 30 ml of buffer I was flowed through the column. 10 ml of serum from the final bleed of rabbit 2 was added to 35 ml of diluent buffer before clarification by centrifugation at 4500 rpm for 10 minutes. Resin was decanted from the column into a 50 ml falcon tube and incubated with diluted serum for 20 hours with gentle mixing. Resin was transferred into the gravity flow column and the diluted serum was allowed to flow through. Column was washed with 20 ml of buffer I and buffer II. Antibodies were eluted by flowing 10 x 1 ml aliquots of elution buffer I through the column. Eluate was collected in 2 ml microcentrifuge tubes containing 200  $\mu$ l elution buffer II. 300  $\mu$ l 100% glycerol was added and the elution fractions were stored at -80  $^{\circ}$ C. SDS – PAGE was used to determine the antibody content of each elution fraction. Elution fractions 2 – 3 was pooled and aliquoted. All antibody aliquots were stored at -80  $^{\circ}$ C.

#### *4.2.9. Electron Microscopy*

Negative stain transmission electron microscopy was performed as previously described by Prof Robin Harries of the Electron microscopy research services at Newcastle University (Harris, 1997, Harris, 2008). 2% w/v uranyl acetate was used for negative staining and microscopy was performed on a Zeis EM900 electron microscope.

### 4.3. *Bacillus subtilis* Methods

#### 4.3.1. Transformation of *B. subtilis*

Transformation procedure was based on a two-step starvation procedure (Anagnostopoulos and Spizizen, 1961). 300 µl of overnight culture of strain 168 grown in MM media at 30°C was inoculated into 5 ml of MM media. The culture was grown at 37°C for 3 hours before an equal volume of starvation media was added. After a further 2 hours of growth, 200 – 300 ng of DNA was added to 600 µl of culture and grown for 1 hour before plating onto nutrient agar containing 2 µg ml<sup>-1</sup> kanamycin.

#### 4.3.2. Isolation of Genomic DNA from *B. subtilis*

An overnight culture of the relevant *B. subtilis* strain grown in LB at 30 °C was diluted 1 in 100 and grown at 37°C until an OD<sub>600</sub> of 0.5 was reached. 2ml of culture was spun down at 13 000 rpm and resuspended in 180 µl *B. subtilis* lysis buffer. 5 µl of lysozyme was added and the cells were incubated at 37°C for 30 minutes. 25 µl of proteinase K was added along with 200 µl of Buffer AL from the Qiagen DNeasy kit and incubated for 30 minutes at 56 °C with shaking. 200 µl of 100% ethanol was added and the lysate was applied to the Qiagen column. From this point the genomic DNA was isolated as per the instructions of the Qiagen DNeasy Blood and Tissue kit.

#### 4.3.3. Construction of a *B. subtilis* $\Delta$ yonO strain

Double cross over homologous recombination was utilised to generate a  $\Delta$ yonO strain. Using genomic DNA from strain 168 as template DNA, primer pairs pBEST:UP L + pBEST:UP R and pBEST:DOWN L + pBEST:DOWN R were used to amplify 2kb regions upstream and downstream of the *yonO* open reading frame, respectively. The PCR products were purified and cloned into the pBEST 501 plasmid. This generated the plasmid pBEST:U:D, which contained a neomycin resistance cassette flanked by DNA sequence homologous to the upstream and downstream region of the *yonO* open reading frame. The plasmid was transformed into strain 168. Transformants were plated onto nutrient agar supplemented with 2µg ml<sup>-1</sup> kanamycin (Melford). Individual transformant colonies were purified by restreaking onto 2µg ml<sup>-1</sup> kanamycin nutrient agar. PCR on genomic DNA using primers pBESTdx L and pBESTdx R was used to confirm the resistance cassette had integrated into the chromosome by double crossover homologous recombination. Additionally, sequencing using primer pairs pBEST seq L + pBEST seq R and pBEST internal seq L + pBEST internal seq R was used to

confirm the deletion of *yonO*. Genomic DNA from the  $\Delta$ *yonO* strain was isolated and used to back cross the *yonO* mutation into the parental strain twice.

#### 4.3.4. Induction of SP $\beta$ prophage from *Bacillus subtilis*

The following protocol is adapted from Molecular Biological Methods for *Bacillus* (Harwood and Cutting, 1991). Overnight cultures of the relevant *B. subtilis* strain were grown at 30°C in MMB medium. These cultures were used to inoculate fresh MMB medium to an OD<sub>600</sub> of 0.02. The cultures were grown at 37°C until they reached mid log phase (approximately 0.5 OD<sub>600</sub>). To induce SP $\beta$ , Mitomycin C (Melford) was added to a final concentration of 0.5  $\mu$ g ml<sup>-1</sup>. Lysis of induced cells occurred at approximately 2 hours post induction.

#### 4.3.5. *B. subtilis* lysates for western blot

*B. subtilis* cells were grown and induced by Mitomycin C. At the required time point, cells were harvested by centrifugation at 9000 rpm for 5 minutes. Pellets were frozen in liquid nitrogen and stored at -20 °C. Pellets were resuspended in grinding buffer. The volume of grinding buffer was one tenth of the volume of cells harvested, typically 5 ml. Cells were then sonicated using a Braun sonicator at 60% amplitude, with 2 seconds on, 2 seconds off, for a total of 4 minutes. The lysates were clarified by centrifugation for 5 minutes at 14 000 rpm. The protein concentration of the lysates was determined using Bradford reagent. The protein concentration was then used for equal loading of the samples on an SDS PAGE gel. The western blot and probing of the membrane were carried out as described in section 4.2.6.

#### 4.3.6. Isolation of total RNA from *B. subtilis* for RNAseq

Fresh MMB media was inoculated with cultures of *B. subtilis* grown in MMB media grown overnight at 30 °C. Cultures were grown at 37 °C until mid-log phase (0.5 OD<sub>600</sub>) was reached, at which point the cultures were split in half. To one half, SP $\beta$  was induced by the addition of Mitomycin C at final concentration of 0.5  $\mu$ g ml<sup>-1</sup>. The other half of the culture was not induced. 60 minutes post induction, cultures were harvested by centrifugation at 14 000 rpm in a Table top centrifuge and the pellets were frozen in liquid nitrogen. To extract total RNA from cell pellets, the Total RNA Plus kit (Norgen) was used as per the manufactures instructions.

To determine the quality of isolated total RNA, an RNA 6000 nano chip and Bioanalyser 2100 (Agilent) were used to the manufacture instructions. For RNAseq, total RNA was

sent to Primbio in the United States. Once there, rRNA was depleted prior to library construction. Sequencing was carried out using Ion torrent technology (Life technologies).

#### *4.3.7. Isolation of SP $\beta$ particles*

Method adapted from Molecular Biology Methods for *Bacillus* (Harwood and Cutting, 1991). To produce SP $\beta$  phage particles, *B. subtilis* strain 168 was grown to 0.5 OD<sub>600</sub> before Mitomycin C induction of SP $\beta$  (see previous section). After lysis ~120 minutes after induction, the cultures were centrifuged at 13000 rpm for 2 minutes. The supernatant, containing phage particles, was then filtered using a syringe driven 0.45  $\mu$ M PVDF filter (Merck-milipore).

The SP $\beta$  sensitive strain, A100A was grown until mid-log phase (0.5 OD<sub>600</sub>). 0.3ml of A100A culture was then incubated for 2 minutes at room temperature with 0.1ml of isolated SP $\beta$  particles. 3ml of MMB overlay agar (MMB media plus 0.5% agar) medium was added before plating out onto bottom agar (MMB plus 2% agar). Plates were incubated overnight at 37 °C before being inspected for plaque formation.

### **4.4. X-ray Crystallography Methods**

#### *4.4.1 Initial Screening for Crystallisation Conditions*

Screening for crystallisation conditions was carried out using MRC 2 well sitting drop 96 well plates (Molecular Dimensions). 80  $\mu$ l of each screen condition was dispensed into the reservoirs of the 96 well plate using a multichannel pipette. The standard set of crystallisation screens included JSCG, Proplex, MIDAS, Index, Morpheus, Structure and AmSO<sub>4</sub> screen. All screens except Index and AmSO<sub>4</sub> were obtained from Molecular Dimensions. Index and AmSO<sub>4</sub> screens were purchased from Hampton Research and Qiagen, respectively. 100 nl and 200 nl of 10 mg ml<sup>-1</sup> YonO was dispensed by the Mosquito robot (TTP Labtech) into each well of the MRC 96 well plate and mixed with 100 nl of reservoir solution. All subsequent screenings were carried out with purified YonO at a concentration of 10 mg ml<sup>-1</sup>. Once dispensed, 96 well plates were sealed and equilibrated overnight at 20 °C. Formation of crystals was visualised using a Leica microscope. Crystals of interest were harvested using a Hampton loop and were cryoprotected in saturated ammonium sulphate before storage in liquid nitrogen. Harvesting of crystals was performed by our collaborator, Dr Arnaud Baslé.

#### 4.4.2. Determination of Crystal Quality

All harvested crystals had their diffraction tested by synchrotron radiation on the IO4-1 beamline at Diamond Lightsource synchrotron in Oxfordshire by Dr Arnaud Baslé.

#### 4.4.3. Optimisation of Crystallisation Conditions

The conditions of the CoCl<sub>2</sub> custom 96 well plate optimisation screen can be seen in Figure 7-1b. The screen was mixed from concentrated stock solutions (1M HEPES pH 7/7.5, 1M MES pH 6/6.5, 3M (NH<sub>4</sub>) SO<sub>4</sub>, 2M NaCl) and dispensed into a deepwell block by an automated Biomek robot (Beckman – Coulter). The custom screen was then used as previously described (section 4.4.1).

#### 4.4.4. Gel filtration calibration

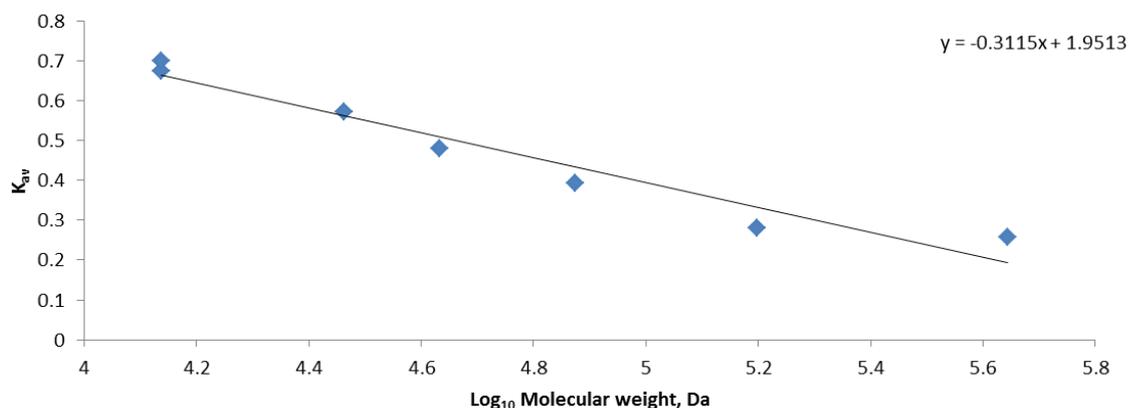
Gel filtration calibration of the Superdex 200 16/60 column was carried out as described in the High Molecular Weight Gel Filtration Calibration Kit (GE Healthcare). Standard proteins were loaded on the column and their elution volumes recorded (Figure 4-1a). The elution volume ( $V_E$ ), the void volume of the column ( $V_O$ ) and the total volume of the column ( $V_C$ ) were used to calculate  $K_{av}$  using the formula:  $K_{av} = \frac{V_E - V_O}{V_C - V_O} \cdot K_{av}$ . The values were plotted against  $\log_{10}$  of the corresponding molecular weight of the standard proteins to give a calibration curve.

To calculate approximate molecular weight of YonO eluting off Superdex 200 16/60 column at 70 ml and 60 ml,  $K_{av}$  values were determined and the equation of the calibration curve trend line was used to determine the  $\log_{10}$  molecular weight of eluted protein.

a)

Protein	Elution Volume	Void Volume $V_o$	Column Volume $V_c$	MW Da	$\log_{10}$ MW	K <sub>av</sub>
Ferritin	63.52	43.95	120.00	440000.00	5.64	0.26
conalbumin	73.75	43.95	120.00	75000.00	4.88	0.39
anhydrase	87.44	43.95	120.00	29000.00	4.46	0.57
RNaseA	95.10	43.95	120.00	13700.00	4.14	0.67
aldolase	65.20	43.95	120.00	158000.00	5.20	0.28
Ovalbumin	80.40	43.95	120.00	43000.00	4.63	0.48
RNaseA	97.17	43.95	120.00	13700.00	4.14	0.70

b)



c)

yonO Elution volume	$K_{av}$	$\log_{10}$ MW	MW Da
59.52	0.204733728	5.615968721	413017.7542
70.96	0.355161078	5.132279491	135606.1828

**Figure 4-1: Calibration of Superdex 200 gel filtration column. a)  $K_{av}$  values of Protein standards.** Protein standards were loaded on Superdex200 16/60 column and elution volumes were recorded. Elution volumes were used to calculate  $K_{av}$  values (see text for formula). **b) Gel filtration calibration curve.**  $K_{av}$  values of protein standards were plotted against the  $\log_{10}$  of their molecular weight. The equation of the trend line is shown. **c) Calculation of molecular weight of eluted YonO.** Gel filtration elution volume of YonO monomer and potential oligomer were determined and used to calculate  $K_{av}$  values. Using the equation of the calibration curve trend line, the  $\log_{10}$  molecular weight of eluted YonO was determined. From this, the molecular weights were obtained.

#### 4.4.5. Post crystallisation Optimisation of Crystal Diffraction

##### 4.4.5.1. HC1 Dehydration of Crystals

Unless stated otherwise, crystals for dehydration were generated by dispensing 300 nl and 600 nl of purified YonO into each well of an MRC 2 well plate and mixing with 300 nl of the corresponding reservoir solution. The  $\text{CoCl}_2$  and  $\text{AMSO}_4$  screens were used for generation large crystals ( $\geq 100 \mu\text{M}$ ) for dehydration. The plates were sealed and allowed to equilibrate at 20 °C overnight.

For dehydration of crystals at the Diamond Light Source, the cryostream on the IO4-1 beamline was removed and replaced by the HC1 dehumidifier. To equilibrate the HC1

dehumidifier, the screen condition for each crystal to be tested was dropped onto a Hampton loop and loaded onto the goniometer. The HC1 relative humidity was then increased or decreased to find the humidity which allowed for a constant drop size, at which point the humidifier was equilibrated. The drop size was monitored using the camera loaded onto the beamline and the HC1 control software.

Once the HC1 dehumidifier was equilibrated, a YonO crystal was harvested in a Hampton loop and loaded directly onto the goniometer. The crystal was then dehydrated by lowering the humidity for a given time, after which the diffraction was tested.

#### ***4.4.5.2. In situ screening of dehydration conditions***

*In situ* screening and testing of diffraction at Diamond Light Source was carried out by Dr Arnaud Baslé. The method below was based on that of Douangamath et al. (2013). Purified YonO was dispensed into a CrystalQuick X-ray plate (Molecular Dimensions) and mixed with reservoir solution at a ratio of 2:1. 200 nl and 140 nl of YonO was dispensed in the wells of CrystalQuick X-ray plate (Molecular Dimensions) by a Mosquito robot and mixed with 100 nl and 70 nl of reservoir solution, respectively. Condition G12 (100 mM MES pH 6.5, 1.8 M Ammonium sulphate) of the Ammonium Sulphate Screen (Qiagen) was used for the reservoir. Crystals were grown by overnight incubation at 20 °C. Dehydration was achieved through vapour diffusion with a 96 well dehydration screen. The dehydration screen was added to the reservoirs of the CrystalQuick plate after YonO crystals had formed and dehydration was allowed to proceed overnight at 20 °C. The CrystalQuick plate was loaded onto the IO4-1 beamline. Crystals from each dehydration condition were tested for increased diffraction.

### **4.5. *In vitro* Transcription Methods**

#### ***4.5.1. Assembled Elongation Complexes***

Assembled complexes were set up as previously described (Korzheva et al., 1998, Sidorenkov et al., 1998, Yuzenkova et al., 2010, Zenkin et al., 2006b). All reactions were set up in a final volume of 10 µl. All oligonucleotides were synthesised by IDT technologies, sequences of which are displayed in the corresponding Figures. Template

DNA strands were modified to bear a biotin tag at the 5' terminus. For each transcription reaction 1 pmol of template DNA strand was incubated with 0.5 pmol of corresponding RNA for 5 minutes at 45 °C in the presence of transcription buffer. Unless stated otherwise, MgCl<sub>2</sub> was present in the transcription buffer at a final concentration of 10 mM. Once formed, the DNA:RNA hybrids were allowed to cool at 4 °C for 20 minutes. 5 pmol of RNA polymerase (YonO, ORF6 RNAP or *E. coli* core RNAP) was added to the hybrid and incubated at 30 °C (ORF6 RNAP) or 37 °C (YonO and *E. coli* core RNAP) for 5 minutes. If required, a 5 time excess of non-template DNA strand was added (5 pmol) and the complexes were incubated for a further 5 minutes at 30 °C or 37 °C.

To visualise RNA generated during the transcription reaction, the RNA was <sup>32</sup>P labelled at either the 5' or 3' end. For 5' end labelling, the RNA was labelled with γ-[<sup>32</sup>P]-ATP (see section 4.5.2) by T4 polynucleotide kinase (Thermo Scientific) prior to assembly of elongation complexes. To label the RNA at the 3' end, assembled complexes were allowed to incorporate an α-[<sup>32</sup>P]-NTP for 5 minutes at 37 °C before complex immobilisation and washing.

If the assembled complexes were to be immobilised, at this point 5 µl of either streptavidin sepharose beads (GE Healthcare) were added and the assembled complexes were immobilised to the beads by vortexing at room temperature for 5 minutes. Prior to use, sepharose beads were equilibrated in transcription buffer. Streptavidin sepharose beads immobilised complexes by the 5' biotin tag present on the template DNA strand. Alternatively, complexes could be immobilised to Ni<sup>2+</sup> sepharose beads through the His tag present on the RNA polymerase. Once immobilised, unstable complexes and unassembled components were removed by washing the immobilised complexes with transcription buffer containing high salt (1M KCl) followed by washing in low salt transcription buffer (40mM KCl).

Transcription reactions were started by the addition of NTPs and incubation at 30 °C or 37 °C. The final concentration of NTPs for each experiment can be found in the corresponding Figure along with the incubation time. Transcription reactions were stopped by the addition of an equal volume of formamide containing stop buffer. Transcription reactions were resolved by 23% denaturing (8M urea) polyacrylamide gel

electrophoresis (PAGE). Gels were exposed and revealed by the Phosphorimaging system (GE Healthcare). ImageQuant software was used for any subsequent analysis.

For cleavage reactions, complexes were assembled as described but in the absence of  $MgCl_2$ . Reactions were started by the addition of  $MgCl_2$  and were resolved and analysed as described above. Pyrophosphorolysis was performed using immobilised complexes as above. Except reactions were carried out in transcription buffer pH 6.8 and initiated by the addition of 500  $\mu M$  pyrophosphate.

#### 4.5.2. 5' $^{32}P$ labelling of oligonucleotides

To  $^{32}P$  label oligonucleotides to be used in assembled elongation complexes, 150 pmol of RNA/ DNA oligonucleotide (sequence displayed in the corresponding Figure) was mixed with 5  $\mu l$  of  $\gamma$ - $^{32}P$ -ATP and 15 units of T4 Polynucleotide kinase (PNK) in reaction buffer A (Thermo scientific) with a final volume of 30  $\mu l$  (5  $\mu M$  oligonucleotide final concentration). The reaction was incubated for 1 hour at 37  $^{\circ}C$ . PNK was inactivated by heating at 65  $^{\circ}C$ . Oligonucleotides were cleaned up using a Microspin 6 chromatography column (Bio-rad) following the manufacturer's instructions.

To label DNA oligonucleotides used in primer extension, sequencing and to generate  $^{32}P$  labelled DNA used in footprinting experiments, 12.5 pmol of DNA oligonucleotide was labelled in a 25  $\mu l$  reaction to generate DNA oligonucleotide at a final concentration of 0.5  $\mu M$ . DNA was cleaned up as before (See above).

#### 4.5.3. Transcription Initiation

DNA for transcription initiation experiments was generated via PCR using *B. subtilis* genomic DNA as template (for primers used, see Table 4-1). Predicted start sites of transcription, as determined by analysis of transcript profiling (Nicolas et al., 2012), were amplified with an additional 100 bp of sequence either side of the predicted start site. Each transcription reaction was carried out in a final volume of 10  $\mu l$ . For each reaction, 1 pmol of template DNA was mixed with 5 pmol of YonO in transcription buffer. To label RNA produced during transcription, 0.15  $\mu M$   $\alpha$ - $^{32}P$ -UTP (unless stated otherwise) was used to start transcription along with 50  $\mu M$  UTP, 500  $\mu M$  ATP, GTP, CTP. If required, heparin was added simultaneously with NTPs to a final concentration of 5  $\mu g\ ml^{-1}$ . Reactions were incubated at 37  $^{\circ}C$  and were stopped after

30 minutes (unless stated otherwise) by the addition of equal volume of stop buffer. RNA was resolved using 15% PAGE and visualised as described previously.

*E. coli* RNAP transcription from the T7A1 promoter was used to generate RNA of a known size. The T7A1 promoter was present on a 328 bp DNA fragment that was biotinylated on the 5' terminus of the template strand. For each 20  $\mu$ l reaction, 1 pmol of core RNAP was mixed with 0.5 pmol T7A1 DNA, 5 pmol of  $\sigma^{70}$ , 50  $\mu$ M CAUC primer, 10  $\mu$ M ATP, 0.15  $\mu$ M  $\alpha$ -[ $^{32}$ P]-GTP and incubated at 37  $^{\circ}$ C. After 2 minutes of incubation, 10  $\mu$ M GTP was added and reactions were incubated for a further 2 minutes. When provided ATP and GTP, RNAP initiates transcription on the T7A1 promoter and synthesises an 11 nt RNA. Elongation complexes were immobilised to streptavidin beads and washed once in high salt transcription buffer followed by 8 washes in low salt transcription buffer. Transcription was resumed by providing the immobilised elongation complexes 25  $\mu$ M NTPs. Reactions were incubated for 120 or 300 seconds before being stopped by the addition of an equal volume of formamide stop buffer.

#### 4.5.4. Primer Extension

To generate RNA for primer extension, transcription initiation reactions (described in section 4.5.4) were scaled up from 10  $\mu$ l to 200  $\mu$ l.

2  $\mu$ l of DNase I was added to reactions for 20 minutes and incubated at 37  $^{\circ}$ C. RNA was phenol chloroform extracted and subsequently ethanol precipitated overnight. Precipitated RNA was resuspended in 30  $\mu$ l RNase-free water.

For primer extension, 2  $\mu$ l of RNA was mixed and incubated with 1 pmol of 5' labelled DNA primer for 10 minutes at 65  $^{\circ}$ C. Samples were cooled on ice for 5 minutes before the addition of the  $MgCl_2$  to a final concentration of 20 mM. To this, 5X first strand buffer, dNTPS (final concentration) and 200 units of reverse transcriptase Superscript II (Invitrogen) were added. Primer extension was allowed to occur for 20 minutes at 37  $^{\circ}$ C. Reactions were stopped by the addition of an equal volume of stop buffer and primer extension products were resolved by denaturing 6% PAGE and visualised using Phosphorimaging (GE Healthcare).

#### *4.5.6. Potassium Permanganate DNA Footprinting*

5' <sup>32</sup>P labelled DNA was generated through PCR. To specifically label the template or non-template DNA strand, a 5' <sup>32</sup>P labelled primer was used. End labelled DNA was purified using the Qiagen Gel Extraction kit as per the manufacturer's instructions. For each reaction, 5 pmol of YonO was mixed with 1 pmol of 5' <sup>32</sup>P labelled DNA in transcription buffer and incubated for 10 minutes at 37 °C. If required, heparin was added to a final concentration of 5 µg ml<sup>-1</sup> followed by 5 minutes of further incubation. DNA was modified by the addition of 5 mM of KMnO<sub>4</sub> (final concentration) at 37 °C. The reaction was allowed to proceed for 30 or 60 seconds, after which point 500 mM β-mercaptoethanol was added. 2 µl of 10 µg ml<sup>-1</sup> glycogen was added before the DNA was extracted using phenol:chloroform:isoamyl alcohol (25:24:1). DNA was ethanol precipitated and dissolved in 90 µl of water. To this, 10 µl of piperidine was added and incubated at 90 °C for 20 minutes. Piperidine was removed by chloroform and DNA was ethanol precipitated. Pelleted DNA was resuspended in 20 µl formamide containing stop buffer. Footprint products were resolved on a 6% polyacrylamide sequencing gel alongside sequencing reactions of the same DNA template.

#### *4.5.7. DNase I Footprinting*

1 pmol of 5' <sup>32</sup>P end labelled template DNA was incubated for 10 minutes with YonO at room temperature. The amount of YonO used for each reaction was varied, see corresponding Figure for details. For each reaction, 1 unit of DNase I (Roche) diluted in 1 X incubation buffer provided, was added. Unless stated otherwise, reactions were incubated for 2 minutes at room temperature. To stop the reactions, an equal amount of stop buffer was added. Cleaved DNA was resolved on 6 % sequencing PAGE gel.

#### *4.5.8. Sanger Sequencing of DNA*

For each DNA template sequenced, 4 reactions were set up. Each reaction was comprised of 2.5 mM MgCl<sub>2</sub>, GoTaq flexi buffer, 0.1 pmol template DNA, 0.1 pmol 5' <sup>32</sup>P labelled DNA primer, 0.5 units of GoTaq DNA polymerase. To each individual reaction, a dideoxynucleotide (ddNTP) termination mix was added. Four ddNTP mixes were used, one for each nucleotide. ddATP, ddGTP, ddCTP and ddTTP were present in the respective sequencing reactions at final concentrations of 0.6 mM, 80 µM, 0.3 mM and 0.9 mM, respectively. The three remaining dNTPs had final concentrations of 10

$\mu\text{M}$ . The reactions were loaded into the thermocycler and sequencing reactions were performed with the following programme:

Temperature ( $^{\circ}\text{C}$ )	Time (s)	Cycles
94	120	
94	30	30
55	30	
72	30	
72	120	
10	pause	

An equal volume (8  $\mu\text{l}$ ) of formamide containing stop buffer was added to each sequencing reaction. The products of the sequencing reactions were resolved on a 6% PAGE gel in denaturing conditions and visualised by phosphorimaging.

## Chapter 5: Preliminary Characterisation of YonO, a Putative RNA polymerase of SP $\beta$ prophage

### 5.1. Introduction

YonO was first identified as an RNAP through distant homology to the eukaryotic rdRNAP (Iyer and Aravind, 2012, Iyer et al., 2003). The homology included the aspartate triad which is essential for catalysis and conserved amongst all eukaryotic rdRNAP and multisubunit (Iyer et al., 2003, Sosunov et al., 2005). Whilst YonO and rdRNAP share additional conserved amino acids and motifs, the only similarity between YonO and multisubunit RNAP is the presence of the triad and two predicted DPBB folds. None of the additional domains that participate in transcription, including the trigger loop, appear to be conserved in YonO. Therefore YonO has been designated as a bacteriophage rdRNAP maintained in the *B. subtilis* genome that functions in an as yet undetermined stage of the SP $\beta$  life cycle (Iyer and Aravind, 2012, Iyer et al., 2003).

YonO is a hypothetical protein and all current understanding of its function and structure has been derived purely from bioinformatics analysis. It remains to be seen if YonO possesses nucleic acid polymerisation activity. Proof YonO is a functional enzyme would raise a wide range of questions, including how it accomplishes transcription despite an apparent lack of catalytic domains and the impact of an additional RNAP on the transcriptome of *B. subtilis*.

### 5.2. Aims

The aim of this work was to determine if YonO is a catalytically active enzyme capable of nucleotide polymerisation. If polymerising activity is shown, we will begin the characterisation of YonO.

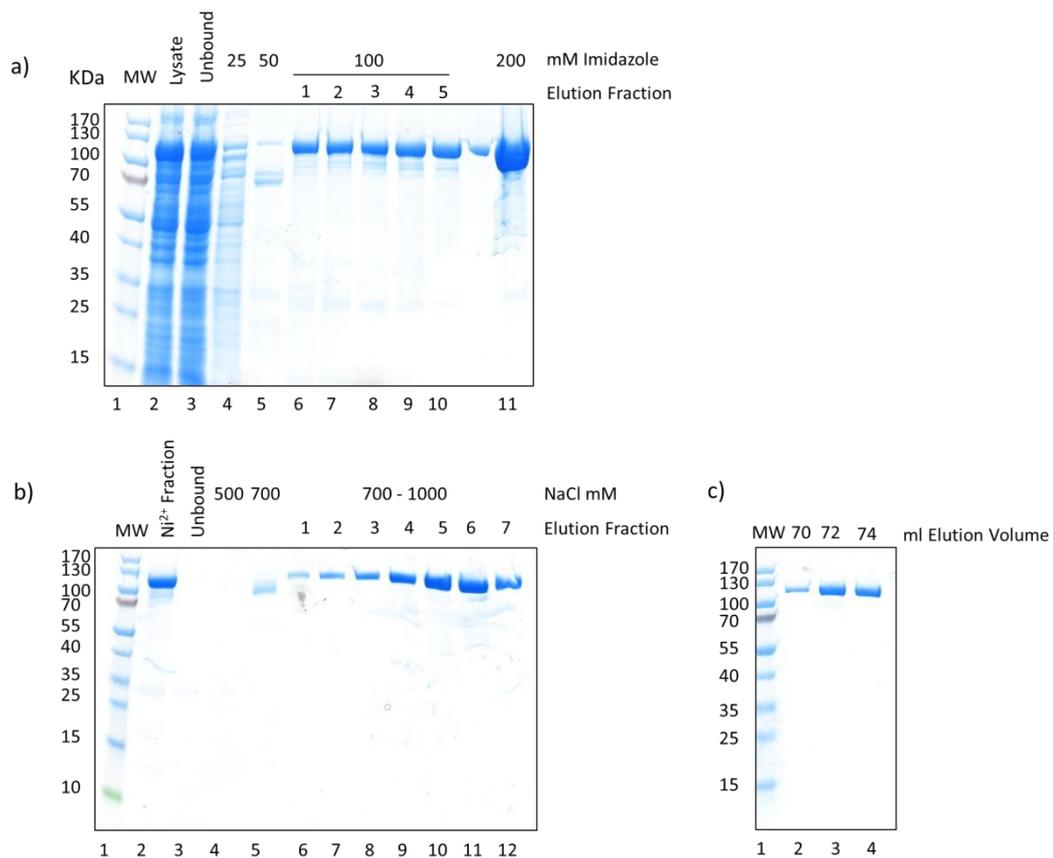
### 5.3. Expression and purification of YonO

In order to conduct *in vitro* characterisation of YonO, purified YonO was required. To this end, *yonO* was amplified from *B. subtilis* genomic DNA and cloned into the pET-28 expression vector. Cloning into pET-28 introduced an N-terminal His tag and allowed for IPTG induced over expression by T7 RNAP. The *E. coli* T7 expression strain carrying the pET-28: YonO construct was grown until mid-exponential phase and induced with 1 mM IPTG. Induced cells were grown for a further 3 hours at 37 °C before being

harvested. From visual inspection of the cell debris of lysed cells, it was clear expressed YonO protein had formed inclusion bodies. To prevent inclusion body formation, *E. coli* cells were shifted to 18 °C and grown overnight upon induction of YonO expression by IPTG. The pelleted cell debris of disrupted cells showed that a decrease in temperature for expression was successful in preventing the formation of inclusion bodies.

Expressed YonO bore a His-tag allowing for its purification from lysates by Ni<sup>2+</sup>-NTA chromatography (Figure 5-1a). To elute non-specific proteins bound to the column, the concentration of imidazole in the chromatography buffer was increased to 25 mM. The remaining protein bound to the Ni<sup>2+</sup>-NTA column, including YonO, was eluted by 100 mM and 200 mM imidazole. SDS PAGE analysis of the protein content of the 100 mM and 200 mM imidazole elution fractions revealed YonO eluted off the column at both concentrations. The majority of YonO was present in the 100 mM imidazole elution fraction. Therefore, the protein in this fraction was concentrated and purified further by heparin affinity chromatography (Figure 5-1b). Heparin is a DNA analogue and can be used for the affinity purification of DNA binding proteins such as RNAP. Once loaded onto the column, YonO bound tightly to the Heparin HiTrap column as 800 mM NaCl was required for elution. After Heparin affinity chromatography, the purity of YonO was determined by SDS PAGE (Figure 5-1b). This showed that after the two chromatography steps, YonO was purified to apparent homogeneity.

In addition to *in vitro* experiments, purified YonO was to be used for crystallisation screening. Gel filtration is a common final purification step for proteins designated for crystallisation as it increases the homogeneity of the protein sample by ensuring all protein is of an equal molecular weight and no oligomers are present. Therefore, YonO was loaded onto a Superdex 200 gel filtration column. The elution fractions were pooled and concentrated. Purified YonO was divided into two aliquots; one was dialysed against storage buffer and used for *in vitro* characterisation, whilst the other was used for crystallisation trials (see chapter 7).



**Figure 5-1: Purification of YonO. a) Ni<sup>2+</sup>-NTA chromatography.** The T7 express *E. coli* strain carrying the pET-28: *yonO* construct was induced with 1 mM IPTG and grown overnight at 18 °C. Cells were disrupted by sonication and lysates were loaded on the Ni<sup>2+</sup>-NTA column. The Column was washed with increasing concentrations of imidazole to elute proteins. Primarily, YonO was eluted by 100 mM imidazole and was collected in elution fractions 1 – 5. Protein content of elution fractions were analysed by SDS PAGE. Lanes 2 – 4 were loaded with 1 µl of protein. Lanes 5 – 10 were loaded with 10 µl of protein. MW refers to the molecular weight ladder. **b) Heparin affinity chromatography.** 100 mM imidazole elution fractions (Ni<sup>2+</sup> fraction) were pooled and loaded onto 5 ml Heparin HiTrap column. The column was washed with 500 mM and 700 mM NaCl YonO heparin elution buffer to remove non-specific proteins. The remaining proteins were eluted by increasing the NaCl concentration to 1M over a gradient. The protein content of elution fractions was visualised by SDS PAGE, which showed YonO was eluted by approximately 800 mM NaCl. All lanes were loaded with 10 µl of protein. **c) Superdex 200 gel filtration.** YonO in heparin elution fractions was concentrated to 10 mg ml<sup>-1</sup> and loaded onto the Superdex 200 16/60 column. 2 ml elution fractions were collected throughout gel filtration, those containing protein were analysed by SDS PAGE. YonO eluted after 70 ml of gel filtration buffer had flown through the column. 5 µl of protein was loaded in all lanes.

## 5.4. *In vitro* characterisation of YonO

### 5.4.1. Artificially Assembled Elongation Complexes

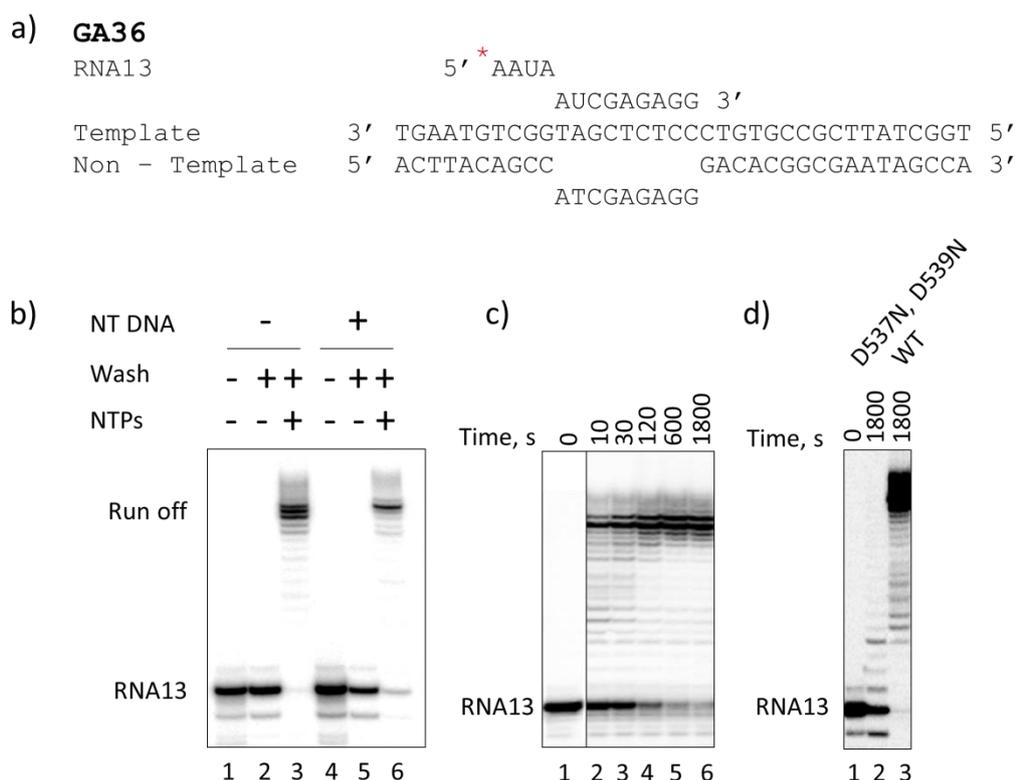
Frequently *in vitro* transcription experiments are performed using elongation complexes (EC) which were derived from transcription initiation at a well characterised promoter such as T7A1 promoter of the T7 coliphage. This technique requires prior identification of promoter elements and detailed knowledge of transcription initiation and so is not applicable for the investigation of a novel RNAP. These limitations can be bypassed by the artificial assembly of elongation complexes (Sidorenkov et al., 1998). Artificial elongation complexes are assembled from a single stranded RNA primer and DNA oligonucleotides (Figure 5-2). The RNA primer is complementary to template DNA strand and so forms an RNA:DNA hybrid. RNAP is able to bind this hybrid and extend the RNA primer upon the addition of NTPs and  $Mg^{2+}$ . To increase the stability and to fully resemble promoter derived elongation complexes, the non-template DNA strand fully complementary to the template strand is included in the assembly of elongation complexes. Artificial elongation complexes are indistinguishable from those derived from a promoter (Sidorenkov et al., 1998, Yuzenkova and Zenkin, 2010).

The standard GA36 scaffold (scaffold refers to the oligonucleotides bound by RNAP) used throughout this work is shown below. GA refers to the first two nucleotides to be incorporated upon provision of NTPs and 36 denotes the length of the template DNA strand. RNA13, so called due to its 13 nt length, hybridises with GA36 template DNA strand to give a hybrid of 9 bp. The 4 nt at the 5' end of RNA13 are not complementary to GA36 to mimic the disengaged transcript.



required for stability. When the non-template DNA strand is omitted during assembly, immobilised bacterial elongation complexes are destroyed by high salt conditions (Sidorenkov et al., 1998).

To determine if YonO possesses RNAP activity, as predicted by the conservation of the aspartate triad, YonO was incubated with the GA36 scaffold to assemble elongation complexes. In accordance with the length of the RNA primer (13 nucleotides), elongation complexes assembled on the GA36 scaffold will be referred to as EC13. The RNA13 component of the scaffold was 5' <sup>32</sup>P labelled to allow us to observe RNA polymerisation. The assembled complexes were immobilised to streptavidin beads and washed with either low ionic strength (40 mM KCl) or high ionic strength (1 M KCl) transcription buffer. The complexes were then provided with a high concentration of NTPs (500 μM) and incubated for 30 minutes at 37 °C.



**Figure 5-3: Determination of YonO RNAP activity. a) Sequence of GA36 scaffold.** The GA36 scaffold used to form YonO EC13. EC13 was detected through 5' <sup>32</sup>P labelling of RNA13 (indicated by red asterisk) prior to assembly. **b) Immobilisation of YonO elongation complexes.** Assembled complexes were immobilised onto streptavidin beads and washed with low salt (40 mM) transcription buffer. Subsequently, the complexes were incubated for 30 minutes at 37 °C in the presence of 500 μM NTPs. **c) Kinetics of YonO RNA polymerisation.** Elongation complexes formed in the absence of

the non-template DNA strand. Transcription was initiated as for Figure 5-3b but reactions were allowed to proceed for the indicated length of time. **d) The aspartate triad is essential for YonO RNAP activity.** EC13 was assembled in the absence of the non-template DNA strand using the YonO D537N D538N mutant and incubated for 1800 seconds at 37°C in the presence of 500 µM NTPs. Wildtype (WT) YonO EC13 were formed as a control. All of the above reactions were terminated by the addition of formamide containing stop buffer and resolved by 23% denaturing PAGE.

During the assembly of elongation complexes, RNAP is required to prevent loss of the RNA primer in the washing stages. In the above experiment (Figure 5-3b, lane 2), RNA13 was not lost during washing of the immobilised YonO elongation complexes with low salt transcription buffer. This shows that YonO is capable of binding the RNA:DNA hybrid to form a stable elongation complex, which is characteristic of an RNAP. Upon incubation with NTPs, full length (run off) transcripts were synthesised as a result of YonO elongating RNA13 to the end of the template DNA. The inclusion of the non-template DNA strand during complex assembly resulted in a small loss of RNA13 during washing (Figure 5-3b, compare lanes 2 and 5). However, this is also observed in the formation of bacterial elongation complexes and occurs due to the non-template DNA strand competing with RNA13 for the binding of the template DNA. If the non-template DNA strand does succeed in hybridising to the template DNA, the RNA13 is subsequently lost during the washing stages of complex assembly. The remaining elongation complexes which have successfully incorporated the non-template DNA strand exhibit much greater stability and resemble promoter derived complexes.

Interestingly, the washing of YonO EC13 comprised of only RNA13 hybridised to the template DNA with high salt did not lead to the loss of RNA13 (not shown). This indicates that the stability of YonO elongation complexes is not determined by the non-template DNA. This contrasts to bacterial RNAP, which does require the binding of the non-template DNA strand for the assembly of highly stable elongation complexes (Sidorenkov et al., 1998).

Previously, the immobilised YonO EC13 were allowed to transcribe for 30 minutes. As all RNA13 was fully extended during that time, this did not provide information regarding the speed of YonO polymerisation. Therefore, the extension of RNA13 by YonO EC13 was followed over time (Figure 5-3 c). After 10 seconds of incubation in the presence of 500 µM, YonO EC13 synthesised full length transcripts, revealing that

YonO, like other RNAPs, is a rapid and processive enzyme (Figure 5-3c, lane 2). Despite this efficient polymerisation, unextended RNA13 was present for approximately 2 minutes. This suggests YonO EC13 was initially paused upon assembly. The pausing of YonO ECs appears to be a general property, as it was observed with various other assembled elongation complexes. The synthesis of full length transcripts within 10 seconds suggests that once the pause was overcome, YonO was able to rapidly transcribe to the end of the template.

To determine if the predicted aspartate triad is catalytic, two residues (D537 and D539) of the YonO aspartate triad were mutated to asparagines by site directed mutagenesis. The mutated YonO was expressed in *E. coli* and purified via Ni<sup>2+</sup> - NTA chromatography. The substitution of two aspartates greatly diminished YonO transcription, as only 1 -4 nucleotides were incorporated after 30 minutes of incubation at 37 °C with a high concentration of NTPs (500 μM). Comparison to the activity of wildtype YonO (Figure 5-3d, lane 2 and 3) confirms that the aspartate triad is essential for the RNAP activity of YonO. The incorporation of 1 - 4 nucleotides by a mutated YonO is similar to *E. coli* RNAP; asparagine substitution of all three asp residues was required to completely abolish transcription (Sosunov et al., 2005) .

The above experiments confirm that YonO, despite only a distant relationship to the other known RNAPs, is a highly processive and rapid RNAP.

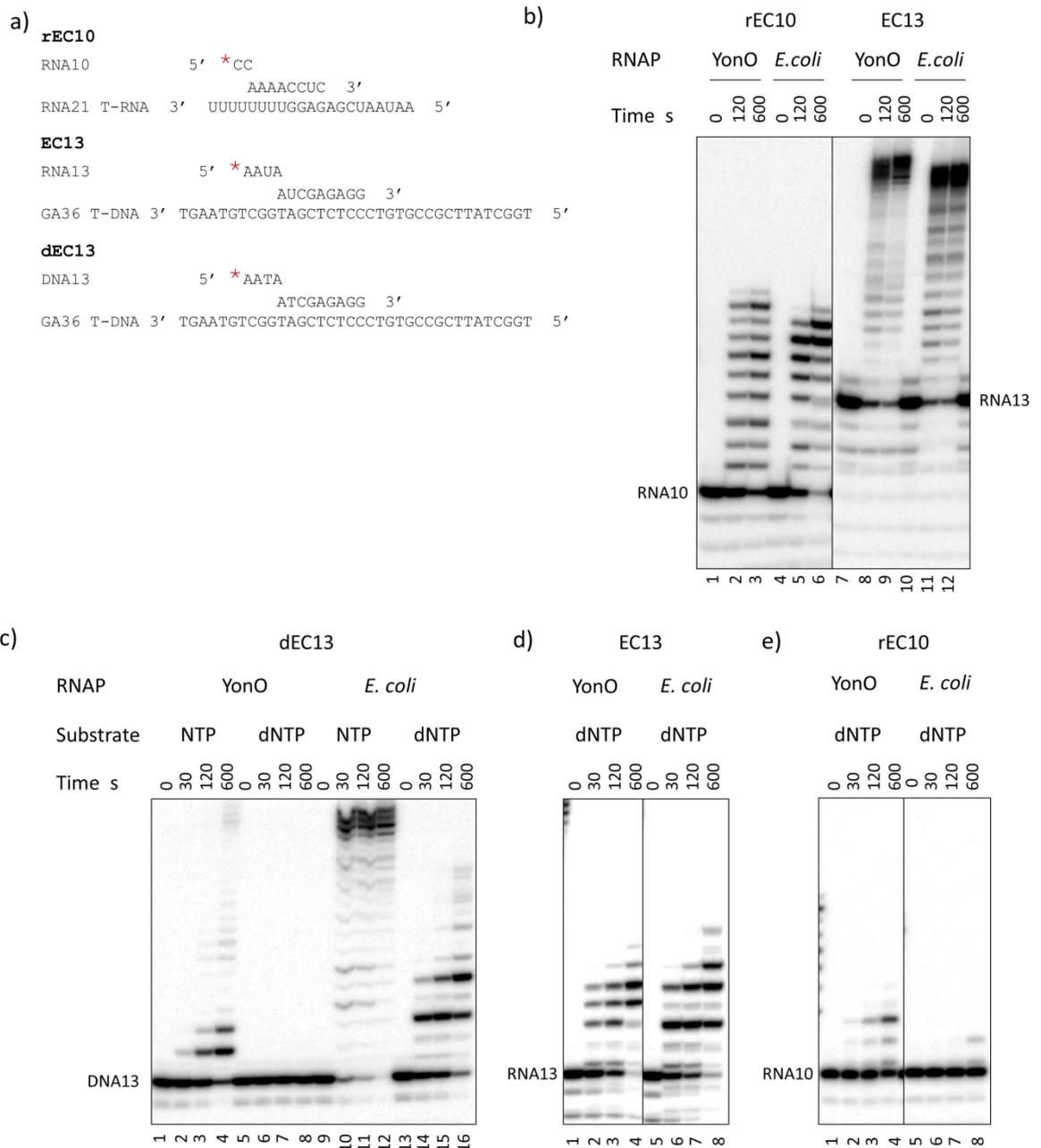
#### 5.4.3. *YonO is a strict DNA Dependant RNA Polymerase*

The previous experiment showed that YonO can accomplish transcription in a DNA dependant manner. This is contrary to the previously published bioinformatics analysis of YonO, which through amino acid sequence similarities predicted YonO functions as an RNA dependant RNAP (rdRNAP) (Krogh et al., 1996, Iyer et al., 2003, Iyer and Aravind, 2012). To investigate if RNA or DNA templates are preferred by YonO, a scaffold featuring an RNA primer (RNA10) hybridised to an RNA template (RNA21) was used to form YonO elongation complexes (referred to as rEC10) alongside the previously used EC13.

Upon incubation with NTPs, YonO rEC10 extended RNA10 to the end of the RNA21 template (lane 2 and 3, Figure 5-4). However, YonO was not processive on the RNA21 template. Only a minority of YonO rEC10 produced full length transcripts after 10

minutes of incubation and the ladder pattern of resolved transcripts indicates YonO paused at every position on the RNA template. This was in contrast to YonO EC13, which extended RNA13 in 2 minutes and rapidly transcribed to the end of the template with minimal pausing. *E. coli* RNAP assembled into rEC10 displayed comparable activity to YonO rEC10 (lane 5). However, *E. coli* rEC10 had increased processivity on the RNA template, with more complexes transcribing to reach the end of the template. Furthermore, *E. coli* rEC10 was quicker on an RNA template than YonO, with more complexes extended in 2 minutes (compare lanes 3 and 6). Examples of *E. coli* RNAP and RNAP II utilising an RNA template have been reported previously (Gildehaus et al., 2007, Lehmann et al., 2007, Wassarman and Saecker, 2006). The pronounced transcription on a DNA template, along with unprocessive transcription on an RNA template comparable to *E. coli* RNAP, strongly suggests YonO is a DNA dependant RNAP.

YonO had a clear preference for extending RNA using a DNA template. In addition to RNAP, DNA polymerases (DNAP) are able to extend RNA hybridised to a DNA template through incorporation of dNMPs. To confirm that YonO is not a DNAP, the ability of YonO EC13 and rEC10 to extend RNA through incorporation of dNMPs was investigated. As an additional experiment to determine the DNAP activity of YonO, complexes (dEC13) using a scaffold comprised of a DNA primer (DNA13) hybridised to the GA36 DNA template strand were assembled and provided with dNTPs or NTPs. When supplied with a high concentration of dNTPs, YonO dEC13 was completely inactive. It was, however, able to utilise NTPs to extend DNA13 but the extension was incredibly slow and unprocessive. YonO EC13 supplied with dNTPs was able to extend RNA13 but again, the extension was poor, with only 2 – 4 dNMPS being incorporated after 2 minutes of incubation. Increased incubation did not result in substantial further incorporation of dNMPs, suggesting YonO had stalled and was unable to transcribe further along the template. An inability to efficiently extend RNA or DNA through incorporation of dNMPs shows YonO does not possess DNAP activity.



**Figure 5-4: Investigating the RNA dependant RNAP and DNA polymerase activity of YonO.** **a) Scaffold sequences.** Sequences of the RNA21, GA36 and dGA36 scaffolds are shown. Elongation complexes and transcription was followed through 5' <sup>32</sup>P labelling of RNA/DNA primer (shown by red asterisk). **b) Determining RNA dependant activity of YonO.** YonO and *E. coli* core RNAP rEC10 and EC13 complexes were formed with RNA21 and GA36 scaffolds, respectively. Transcription was initiated by the addition of 500 μM NTPs and allowed to proceed for the indicated times. **c) Determining DNA polymerase activity of YonO.** YonO and *E. coli* dEC13 complexes were formed with the dGA36 scaffold which gave a DNA:DNA hybrid. Transcription was initiated by addition of 500 μM NTPs or dNTPs. **d) and e) RNA primed DNA polymerase activity of YonO.** RNA21 and GA36 scaffolds were assembled into EC10 and EC13 as before, except transcription was initiated by provision of 500 μM dNTPS. All reactions were terminated by the addition of stop buffer. Transcripts were resolved on 23 % PAGE gel and visualised through phosphorimaging.

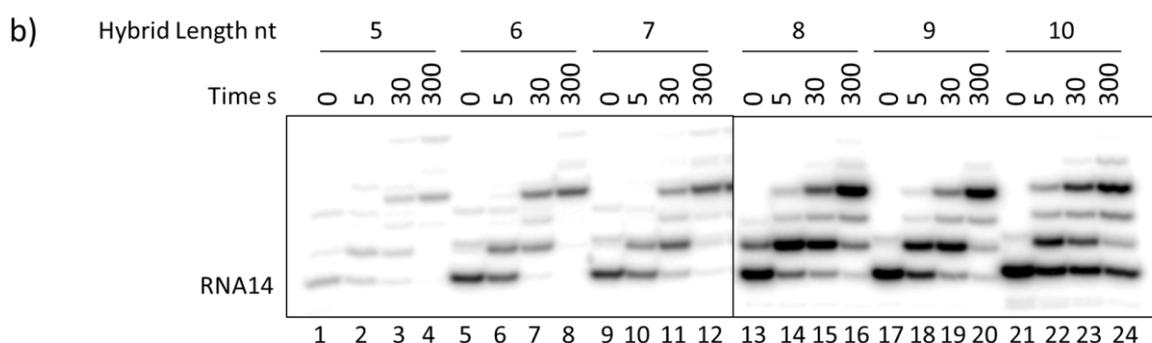
To compare YonO to a well-studied RNAP, *E. coli* dEC13 complexes were assembled. Unexpectedly, *E. coli* dEC13, when supplied with either NTPs or dNTPs was comparable to *E. coli* EC13. In the presence of NTPs, *E. coli* dEC13 fully extended DNA13. The reaction was quick, with all complexes extended in 2 minutes of incubation. *E. coli* dEC13 extended DNA13 by incorporating 3 to 5 dNMPs. Whilst dNMP incorporation by *E. coli* dEC13 was slow and unprocessive, it was comparable to dNMP incorporation by EC13. Together, these results suggest *E. coli* RNAP is able to utilise a DNA:DNA hybrid in a similar manner to an RNA: DNA hybrid. The rapid extension of DNA through incorporation of NMPs likely arises through the DNA:DNA hybrid mimicking the RNA:DNA hybrid sufficiently enough to be utilised by *E. coli* RNAP. Furthermore, through incorporation of NMPs, the DNA:DNA hybrid would transition to a RNA:DNA hybrid. Subsequent, downstream transcription would resemble canonical, DNA dependant transcription. *In vivo*, *E. coli* RNAP would not be recruited to a DNA:DNA hybrid reducing the significance of this observation. It does, however, suggest that YonO has increased stringency in hybrid binding compared to *E. coli* RNAP.

In conclusion, through altering the nature of nucleic acid scaffolds (RNA/DNA primers combined with RNA/DNA templates) and the substrate (NTPs/dNTPs), we have shown YonO is a DNA dependant RNAP and not an rdRNAP or DNAP. Rapid synthesis of full length transcripts by YonO was strictly dependant on RNA hybridised to template DNA hybrid and NTP substrates. Whilst YonO did elongate RNA in an RNA template dependant manner, it failed to do so in a rapid and processive manner characteristic of a nucleic acid polymerase.

#### 5.4.4. Determination of Optimum Hybrid Length

Throughout transcription, elongation complexes maintain a constant hybrid length. The hybrid length is a determinant of the processivity and stability of the elongation complex. Previously, structural and biochemical data were used to show a hybrid length of 9 bp and 8 bp were maintained by *E. coli* and T7 RNAP, respectively (Sidorenkov et al., 1998, Temiakov et al., 2002, Bochkareva et al., 2012, Vassylyev et al., 2007a). To investigate the hybrid length preferred by YonO elongation complexes,

alternative RNA primers were used in conjunction with the GA36 template DNA strand to give hybrids ranging from 5 to 10 bp in length (Figure 5-5a). As the template DNA remained constant, we were able to label all active elongation complexes through incorporation of  $\alpha$ -[<sup>32</sup>P]-GMP to make EC14 complexes. In addition to labelling the complexes, the efficiency of incorporating  $\alpha$ -[<sup>32</sup>P]-GMP would show the effect of varying the hybrid length. To observe the influence of hybrid length on elongation, complexes were further supplied with 100  $\mu$ M ATP, the next nucleotide to be incorporated according to the DNA template.



**Figure 5-5: Influence of hybrid length on YonO elongation complexes. a) Scaffold sequences.** The sequences of scaffolds assembled into elongation complexes are shown. The hybrid length is shown and the sequences of the hybrid are in red. **b) Influence of the hybrid length on YonO elongation complexes.** YonO EC13 were formed and labelled by incorporation of  $\alpha$ -[<sup>32</sup>P]-GTP. To observe extension, EC14 was incubated with 100  $\mu$ M ATP for the indicated time before reactions were stopped and transcripts resolved by 23% denaturing PAGE.

The greatest incorporation of  $\alpha$ -[<sup>32</sup>P]-GMP was seen in EC13 with a hybrid of 8 – 10 bp (lanes 13 – 24, Figure 5-5b). However, a 10 bp hybrid promoted pausing of the complexes as after 300 seconds, a significant percentage of complexes remained unextended when supplied with 100  $\mu$ M ATP. As only active elongation complexes were labelled through incorporation of  $\alpha$ -[<sup>32</sup>P]-GTP, the inability of EC14 with a 10 bp hybrid to extend may be indicative of pausing due to an overextended hybrid upon the incorporation of  $\alpha$ -[<sup>32</sup>P]-GTP.

A hybrid less than 8 bp led to reduced incorporation of  $\alpha$ -[<sup>32</sup>P]-GMP, with the least incorporation occurring in EC13 with a 5 bp hybrid. This suggests YonO elongation complexes prefer an 8-9 bp hybrid. Unexpectedly, incubation of EC14 with 100  $\mu$ M ATP showed YonO readily misincorporates. All complexes misincorporated ATP for CTP within 5 seconds of incubation to give EC16. EC16 rapidly extended RNA16 to become EC17 through incorporation of AMP, the next nucleotide to be incorporated according to the template sequence (refer to scaffold sequences in Figure 5-5a). The hybrid length appeared to influence misincorporation. When the hybrid length was reduced to 7 bp or less, additional misincorporation occurred during labelling of complexes (lanes 1, 5 and 9) and after 10 minutes incubations (lanes 4, 8 and 12). This additional misincorporation may be due to an alteration in the translocation state of the elongation complex. Whilst a reduced hybrid is not preferred, most likely due to its poor stability, the shortened hybrid may cause the complexes to favour a post-translocated state (Sidorenkov et al., 1998). In this state, the complexes could possess an increased propensity to both incorporate and misincorporate. As this is speculation, additional work is required to explore this possibility further.

Together, these results suggest YonO is similar to *E. coli* RNAP and T7 RNAP in that a hybrid length of 8-10 is optimal for transcription whilst suggesting YonO is an error prone RNAP.

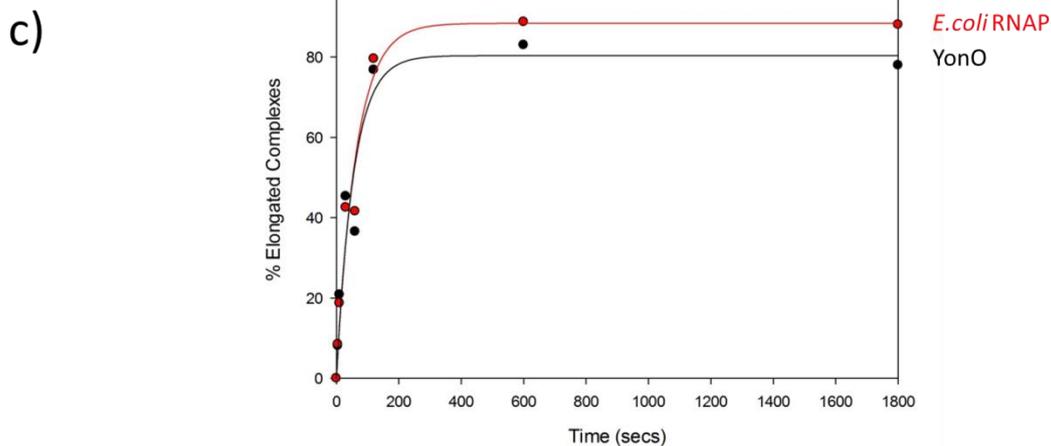
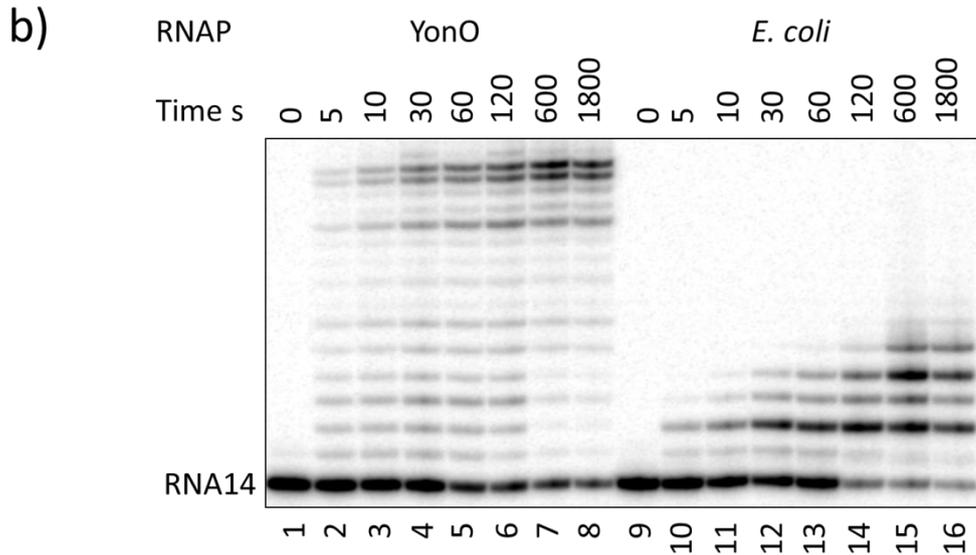
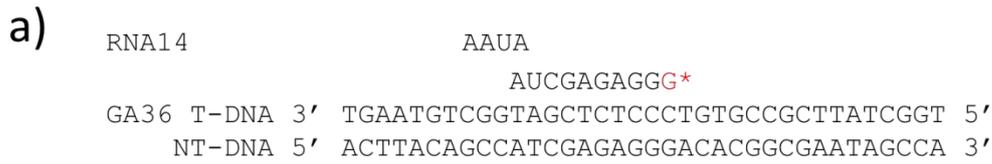
#### 5.4.5. Rate of Transcription

A prominent characteristic of RNAP is the rapid rate at which it accomplishes polymerisation. To see if YonO is comparable to a typical multisubunit RNAP in this regard, EC14 elongation complexes (3' labelled by incorporation of  $\alpha$ -[<sup>32</sup>P]-GTP by EC13) were allowed to transcribe in the presence of 1  $\mu$ M NTPs (Figure 5-6). In previous experiments, saturating concentrations of NTPs were used. At this

concentration, the rate of YonO and *E. coli* RNAP elongation is too high to make informative comparisons. A much reduced concentration of NTPs would drastically slow the rate of polymerisation.

As mentioned previously, it appears YonO pauses on the EC14 scaffold. This conclusion was arrived at due to the presence of unextended RNA14 at saturating NTP concentrations, despite active ECs transcribing to the end of the template within 5 seconds. As only active complexes are labelled through the incorporation of  $\alpha$ -[ $^{32}\text{P}$ ]-GTP, the unextended RNA14 suggests that upon incorporation of a single nucleotide, active YonO ECs pause. Upon the provision of all NTPs, the YonO ECs which overcome the pause rapidly transcribe to the end of the template with little pausing elsewhere, even in low NTP concentrations (1  $\mu\text{M}$ ). This was in contrast to *E. coli* RNAP EC, which at this concentration of NTPs had a tendency to pause after the extension of EC14. The differential pausing reflects differences between the two enzymes in recognising the template DNA sequence (Bochkareva et al., 2012).

The propensity for YonO to pause on artificially assembled complexes prevented accurate determination of the rate of polymerisation. However, calculation of the rate of polymerisation ( $K_{\text{obs}}$ ) of YonO ECs exhibiting pausing ( $K_{\text{obs}} = 0.018 \text{ s}^{-1}$ ) showed that YonO was as fast as non-paused *E. coli* RNAP ( $K_{\text{obs}} = 0.016 \text{ s}^{-1}$ ). As a fraction of the YonO ECs used for the calculation of  $K_{\text{obs}}$  were paused, the accurate rate of YonO polymerisation would be expected to be higher. Therefore, combined with YonO transcribing to the end of the template, this shows YonO is more processive and potentially faster than *E. coli* RNAP.



**Figure 5-6: YonO is a fast and processive RNAP. a) Scaffold sequences.** The sequence of the GA36 scaffold used to form the initial YonO and *E. coli* EC13 complexes is shown. G\* indicates the incorporation of  $\alpha$ - $^{32}\text{P}$ -GMP by EC13 to label ECs. **b) Comparison of YonO and *E. coli* elongation.** YonO and *E. coli* EC13 complexes were assembled and labelled by incubation with  $\alpha$ - $^{32}\text{P}$ -GTP to form EC14. EC14 were immobilised and washed to remove unincorporated radiolabelled nucleotides. YonO and *E. coli* EC14 complexes were incubated with 1  $\mu\text{M}$  NTPs for the indicated times at 37  $^{\circ}\text{C}$ . Reactions were stopped by addition of stop buffer and resolved by 23% PAGE. **c) Rate of YonO elongation.** The observed rates ( $K_{\text{obs}}$ ) of elongation were calculated by non-linear regression using Sigma Plot. The curve is the fitting of the kinetic data to a single exponential equation.

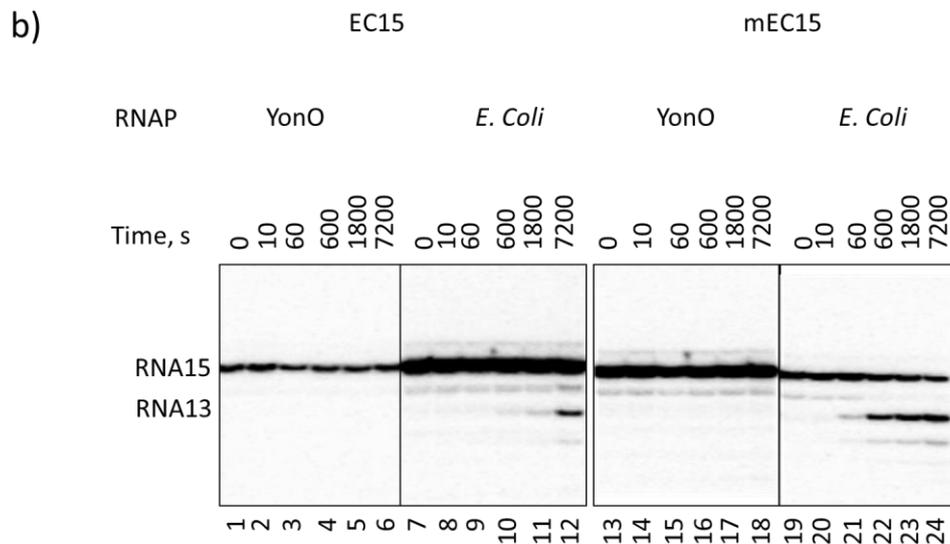


#### 5.4.7. Hydrolysis by YonO

In addition to synthesising RNA, multisubunit RNAPs have been shown to catalyse the hydrolysis of phosphodiester bonds at the 3' end of the RNA chain (Orlova et al., 1995, Zenkin et al., 2006b). This intrinsic activity contributes to the fidelity of transcription by allowing the removal of wrongly incorporated nucleotides (misincorporation).

Hydrolysis of RNA requires elongation complexes to be in a 1 bp backtracked state. The previously used EC13 are stabilised in the post-translocated state, preventing the backtracking of elongation complexes required for hydrolysis. By assembling elongation complexes (EC15) with an extended RNA primer (RNA15) (see Figure 5-8), RNAP is permitted to oscillate on the template DNA between pre-translocated, post-translocated and 1 bp back tracked states (Roghalian et al., 2011). In EC15, the 3' NMP of RNA is complementary to the corresponding position on the template strand DNA. However, a mismatch between the 3' NMP of RNA15 and the template DNA would stabilise the 1 bp backtracked state, as opposed to the pre and post-translocated states, and stimulate hydrolysis (Zenkin et al., 2006b). Therefore, in a parallel experiment, the DNA sequence of the template strand was altered to introduce an A:A mismatch upon hybridisation with RNA15 during mismatch elongation complex (mEC15) assembly.

To determine if YonO is able to perform hydrolysis, assembled EC15 and mEC15 were incubated in the presence of  $Mg^{2+}$ . After two hours of incubation, neither YonO EC15 nor mEC15 accomplished RNA hydrolysis. In comparison, *E. coli* EC15 and mEC15 demonstrate the stimulation of hydrolysis by a 3' mismatch, with hydrolysis occurring after 60 seconds by the mEC15. The ability of RNAP to hydrolyse RNA varies between organisms, with RNAP II of *S. cerevisiae* having very slow hydrolysis (Nielsen and Zenkin, 2013, Izban and Luse, 1992, Reines, 1992). Despite this, the 120 minutes incubation of EC15 and mEC15 was thought to be sufficient to detect any potential hydrolysis by YonO. Therefore, this experiment strongly suggests that YonO does not possess the ability to remove misincorporated nucleotides through hydrolysis.



**Figure 5-8: Determination of YonO hydrolysis activity. a) Sequences of oligonucleotides in EC15 and mEC15.** Elongation complexes were followed by 5' <sup>32</sup>P labelling of RNA15, shown by red asterisk. mEC15 3' end RNA A:A mismatch was introduced through alteration of DNA sequence. **b) Correct and mismatch elongation complexes to determine YonO hydrolysis activity.** EC15 and mEC15 complexes were immobilised and washed before incubation with 10 mM Mg<sup>2+</sup> for the indicated lengths of time. Reactions were stopped by the addition of formamide stop buffer and were resolved by 23% denaturing PAGE. Transcripts were visualised by phosphorimaging.

#### 5.4.8. Misincorporation by YonO

Maintaining high fidelity during transcription is essential for correct gene expression. Intrinsic hydrolysis by RNAP and the action of Gre proteins and their S factor analogues in eukaryotes and archaea corrects for the incorporation of the incorrect nucleotide. Prior to removal of misincorporated NMPs, the major determinant of fidelity is the discrimination of incorrect substrates such as non-complementary NTPs or dNTPs by RNAP during the nucleotide addition cycle (Yuzenkova et al., 2010). Previous experiments have suggested that YonO is prone to misincorporation. To investigate

this further, YonO and *E. coli* RNAP EC13 were assembled. Active elongation complexes were labelled through the incorporation of  $\alpha$ -[<sup>32</sup>P]-GTP to give EC14. To observe misincorporation, EC14 was provided with each of the 4 nucleotides (ATP, CTP, GTP and UTP).

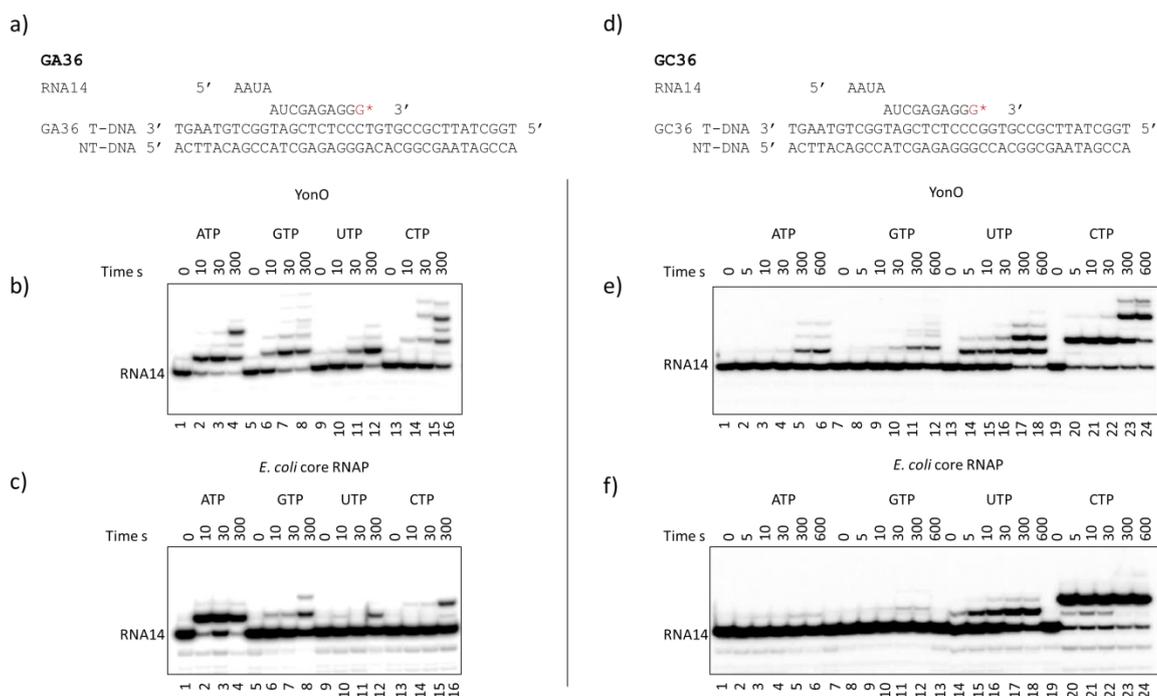
In EC14, the next correct nucleotide to be incorporated according to the template was AMP. As expected, YonO and *E. coli* EC14 rapidly incorporated AMP to produce RNA15 within 10 seconds, (lanes 1 -4, Figure 5-9 b and c). After 10 minutes of incubation, YonO misincorporated AMP to synthesise RNA16. RNA16 was rapidly extended to RNA17 through the correct incorporation of AMP (lane 4, Figure 5-9b). For the remaining nucleotides, YonO displayed increased misincorporation compared to *E. coli* EC14. Multiple misincorporation events occurred upon incubation with CTP (Figure 5-9b, lane 13 – 16). These events can be explained upon analysis of the template sequence. Following ATP, the next nucleotides to be incorporated are CMP, AMP and CMP. After misincorporation of CMP to produce EC15, all EC15 complexes correctly incorporate CMP, resulting in EC16. Misincorporation of CMP instead of AMP, followed by correct incorporation of CTP is repeated again, thus explaining the synthesis of RNA18. Misincorporation of GTP, UTP and CTP instead of ATP does occur by *E. coli* RNAP EC14. However, the reactions are slow with only a small fraction of complexes misincorporating after 10 minutes of incubation.

To see if misincorporation is a general characteristic of YonO, an alternative scaffold was used to generate a different misincorporation event. Elongation complexes were assembled with the GC36 scaffold and labelled as before through the incorporation of  $\alpha$ -[<sup>32</sup>P]-GTP, giving EC14. The GC36 scaffold was identical to GA36 scaffold used previously except EC15 is formed through the incorporation of CMP, rather than AMP (refer to Figure 5-9d for sequences). Upon incubation with the correct NTP (CTP), YonO quickly synthesised RNA16 within 5 seconds by incorporating 2 CMP nucleotides. After 5 minutes of incubation, RNA18 was synthesised through the misincorporation of CTP for ATP, followed by correct incorporation of CMP. When supplied with a non-complementary nucleotide (ATP, GTP and UTP), YonO EC14 misincorporated to synthesise RNA15 (ATP and GTP) and RNA16 (UTP). Out of the non-complementary nucleotides, UTP had the fastest rate of misincorporation, with RNA15 synthesised

within 5 seconds. Increased incubation led to further misincorporation of UTP, with all EC14 being converted to EC15 or EC16 after 5 minutes of incubation.

The extent of YonO misincorporation can be seen by comparison to the *E. coli* RNAP EC14 supplied with each nucleotide. As with YonO, *E. coli* EC14 correctly incorporated CTP to synthesise RNA16 but no further extension was observed throughout the time course of the experiment. When incubated with non-complementary nucleotides, *E. coli* EC14 only misincorporated UMP. Unlike YonO, misincorporation was limited to the extension of RNA14 by one UMP. Furthermore, misincorporation of UMP was slow as approximately half of EC14 remaining unextended after 5 minutes of incubation. The remaining nucleotides, ATP and GTP were misincorporated by *E. coli* RNAP EC14. However, the reactions were incredibly slow with RNA15 only being detected after 5 minutes of incubation.

In conclusion, the misincorporation of non-complementary NTPs by YonO elongation complexes shows YonO is more error prone than multisubunit RNAP. Comparing misincorporation on the GA36 and GC36 scaffolds, it is apparent YonO is less efficient at discriminating between differing purines and differing pyrimidines. On the GA36 scaffold, misincorporation of GTP, a purine, instead of ATP occurs at a faster rate than misincorporation of the pyrimidines, UTP or CTP. The same observation was made during misincorporation on the GC36 scaffold. The fastest rate of misincorporation was UTP rather than CTP. In this regard, the pattern of YonO misincorporation is comparable to *E. coli* EC14, albeit at much faster rates (Yuzenkova et al., 2010).



**Figure 5-9: Misincorporation by YonO elongation complexes. a) GA36 scaffold.**

Sequence of GA36 oligonucleotide scaffold used to assemble EC13. Complexes were labelled through incorporation of  $\alpha$ -[ $^{32}$ P]-GTP (shown in red) to form EC14. **b) Misincorporation by YonO GA36 EC14.** EC14 was incubated with 100  $\mu$ M of ATP, GTP, UTP or CTP at 37  $^{\circ}$ C for the indicated times. **c) Misincorporation by E. coli core RNAP GA36 EC14.** As for Figure 5-9b except complexes were assembled with E. coli core RNAP. **d) GC36 scaffold.** Sequence of GC36 oligonucleotide scaffold used to assemble EC13. Note the next NTP to be incorporated after GTP is CTP, not ATP as with GA36 scaffold. EC14 was formed as before. **e) and f) Misincorporation by YonO and E. coli GC36 EC14,** respectively. EC14 was incubated with 100  $\mu$ M ATP, GTP, UTP and CTP as before. All reactions were stopped after the indicated time by addition of stop solution. Transcripts were resolved on 23 % PAGE gel and visualised by phosphorimaging.

## 5.5. Discussion

YonO was successfully purified by utilising an N-terminal His tag. Despite greatly facilitating the purification of recombinant proteins, there is a chance that a His tag will have a detrimental effect on protein structure or function. These negative effects are dependent on the position of the tag (N or C terminal end) and the folding of the protein. If the tagged N-terminal or C-terminal domain participates in a key structural fold or is embedded in the active centre, the function of the protein may be compromised. However, compared to other purification tags, the His tag is small (2 KDa in the case of YonO) and so is less likely to interfere with protein folding or

function. As YonO displays proficient RNA polymerase activity with an N or C-terminal (not shown) His tag, it is unlikely that the function of YonO is impeded or altered by the presence of the His tag.

Prior to this work, YonO was a hypothetical protein only predicted to function as an RNAP through distant homology to rdRNAP and the multisubunit RNAP (Iyer et al., 2003). Through utilisation of artificially assembled elongation complexes, we have directly shown YonO possesses RNAP activity. The activity was dependant on chelation of  $Mg^{2+}$  ions revealing YonO, is like all nucleic acid polymerases that conduct catalysis through a general two metal mechanism (Sosunov et al., 2003, Steitz, 1998). The polymerising activity of YonO is comparable to that of *E. coli* core RNAP; it is both fast and processive. This was not expected, as the only functional module identifying YonO as an RNAP is the aspartate triad. No other RNAP functional domains have been identified in the YonO amino acid sequence. This includes the trigger loop, which plays a crucial role in all reactions catalysed by RNAP (Temiakov et al., 2005, Wang et al., 2006, Yuzenkova et al., 2010, Yuzenkova and Zenkin, 2010, Zhang et al., 2010).

Whilst the speed and processivity of YonO is similar to that of its multisubunit relatives, the determinants of elongation complex stability appear to differ. Unlike bacterial RNAP, YonO complexed with only the RNA and template DNA strand withstood high ionic conditions. In bacterial RNAPs, the hybrid and the presence of the downstream non-template DNA are major contributors to the stability of elongation complexes (Sidorenkov et al., 1998). The stability of elongation complexes is also influenced through interactions between the downstream DNA duplex and RNAP domains such as the clamp and jaw domains (Ederth et al., 2002, Chakraborty et al., 2012, Gnatt et al., 2001). According to amino acid sequence comparisons, these domains are not conserved in YonO, which may explain why Yono does not depend on the non-template DNA strand for stability. Therefore, this raises the question of how YonO is able to maintain tight binding of the hybrid.

The catalytic activity of YonO also raises questions with regards to the mechanisms employed by YonO for accomplishing transcription. These questions include how the substrates enter and are bound by the active centre, how YonO translocates along the template, and most interestingly: how YonO performs catalysis. Elongation is preceded

by initiation and followed by termination. Therefore, the ability of YonO to transcribe implies mechanisms of transcription initiation and termination exist, both of which remain completely undefined.

rdRNAPs operate in the post-transcription gene silencing (PTGS) pathway occurring across a wide range of eukaryotic organisms including humans (Maida et al., 2009). Despite this, our *in vitro* experiments shows processive transcription by YonO is DNA dependant and requires NTP substrates. This, along with SP $\beta$  being a DNA based bacteriophage, suggests YonO does not function to produce double stranded RNA necessary for PTGS. Supporting this notion is the absence of identifiable PTGS proteins encoded in the SP $\beta$  genome (Iyer et al., 2003). One eukaryotic rdRNAP related to YonO is QDE-1 of *Neurospora crassa*. Interestingly, QDE-1 has been shown to transcribe on single strand DNA templates (Lee et al., 2009, Aalto et al., 2010). This is in addition to the previously observed transcription on a single stranded RNA template. Therefore, closer similarity to rdRNAP over multisubunit RNAP does not preclude YonO from functioning as a DNA dependant RNAP.

The use of artificial elongation complexes has revealed additional characteristics of YonO. YonO appears not to be able to intrinsically catalyse the hydrolysis of phosphodiester bonds, even when stabilised in a 1 bp backtracked position. This is surprising, as in multisubunit RNAPs the same active centre catalyses both phosphodiester bond synthesis and hydrolysis. Intrinsic hydrolysis has been well detailed in bacterial RNAP and has been confirmed to be performed by RNAP II (Nielsen and Zenkin, 2013, Yuzenkova et al., 2010, Yuzenkova and Zenkin, 2010, Zenkin et al., 2006b). Hydrolysis activity is stimulated by misincorporation and has therefore been linked to RNAP proof reading and fidelity (Zenkin et al., 2006b). YonO was able to remove 3' NMP of RNA through pyrophosphorolysis, offering a potential mechanism of proofreading. However, pyrophosphorolysis as a means of removing misincorporated nucleotides is not probable. This notion is based on all RNAPs, whilst being able to catalyse pyrophosphorolysis, using intrinsic hydrolysis or auxiliary factors such as Gre factors (bacteria) and TFIIS (RNAP II) for proofreading.

YonO is analogous to T7 RNAP in that both enzymes do not possess hydrolysis activity but are capable of pyrophosphorolysis (Huang et al., 2000). Pyrophosphorolysis was

ruled out as a mechanism of proof reading by T7 RNAP. Addition of PPi to stalled elongation complexes did not result in the removal of an incorporated 3' dNTP and resumption of transcription. It was concluded fidelity of T7 RNAP transcription was governed solely by highly stringent substrate selection (Huang et al., 2000). SP $\beta$  and other DNA based bacteriophages that utilise their own RNAP for gene expression will have less tolerance to decreased fidelity of transcription. Increased errors during transcription would result in truncated or mutated proteins that would not function correctly. Ultimately, this would limit the bacteriophages ability to successfully replicate. This is in contrast to RNA based bacteriophages that utilise rdRNAPs for replication. In these circumstances, lower rates of fidelity could be tolerated. Lower rates of transcriptional fidelity can even be advantageous, as it results in increased variation in the genome and therefore increased rates of evolution.

Despite *in vitro* experiments presented here demonstrating that YonO is more error prone than other RNAPs, the rate of YonO incorporation was faster compared to misincorporation. This implies that at high substrate concentrations, YonO is able to maintain fidelity through substrate selectivity. By testing misincorporation of each nucleotide in place of ATP and CTP, it does appear YonO is able to distinguish well between purines and pyrimidines, with GTP being misincorporated instead of ATP at a faster rate than CTP or UTP. The *in vivo* impact, if any, of error prone YonO transcription needs to be investigated. Despite RNAP having intrinsic proof reading activity, transcription factors exist to perform transcript hydrolysis upon misincorporation (Roghanian et al., 2011). No bacteriophage Gre homologues have been reported. However, only 25% of the 187 proteins encoded on SP $\beta$  have been assigned putative functions. The coding of a protein either distantly homologous or functionally analogous to Gre-like factors by SP $\beta$  cannot be entirely dismissed.

The *in vitro* activity of YonO reveals the presence of a second, functional RNAP existing in the genome of the model Gram positive organism *Bacillus subtilis*. The possible impact YonO has upon the host transcriptome remains to be investigated. PSI BLAST revealed the existence of YonO homologues across a broad range of bacteriophages that infect other gram positive bacteria, including the medically and biotechnologically important *Clostridium* and *Lactococcus* species (Iyer et al., 2003). Our work confirming YonO has RNAP activity provides evidence that the YonO-like homologues present in a

wide range of gram positive bacteria may also be catalytically active enzymes. This makes YonO an archetype of a novel class of single subunit RNAP and further work will develop this potentially new field of transcription.

As YonO is capable of RNA polymerisation, our work supports the proposal that YonO is a descendant of an ancient enzyme that represents an intermediate stage during the evolution of the multisubunit RNAP (Iyer and Aravind, 2012, Iyer et al., 2003). Despite bearing closer resemblance to eukaryotic rdRNAP, YonO accomplishes transcription in a DNA dependant manner. Speculating, the YonO ancestor may have been the predominant transcription enzyme during the emergence of the DNA world.

## Chapter 6: Investigation of Biological Function of YonO

### 6.1. Introduction

The SP $\beta$  prophage resides within the genome of the widely studied wildtype (WT) *B. subtilis* strain 168. At very low frequencies (1 in 10<sup>6</sup> cells), the prophage can be spontaneously induced. This results in phage development and the eventual lysis of cells and release of SP $\beta$  phage particles (Warner et al., 1977). Induction can be achieved at a much greater frequency by DNA damage. Experimentally, DNA damage can be brought about by the addition of Mitomycin C, a compound which induces the SOS response by causing DNA crosslinking (Goranov et al., 2006).

Sequencing of the 134 KB SP $\beta$  prophage genome allowed for the prediction of 187 putative ORFs. Subsequent analysis of the orientation of the ORFs revealed that they are arranged into three Clusters, termed Cluster I, Cluster II and Cluster III (Lazarevic et al., 1999). The majority of the proteins encoded by the ORFs of SP $\beta$  are yet to be ascribed a function. The proteins which have been given a function, either through homology or experimental evidence, have been used to categorise the Clusters based on the temporal expression of the genes they contain.

Typically, bacteriophage operons are categorised according to the life cycle stage in which they are expressed. The bacteriophage encoded genes expressed at the beginning of an infection are referred to as early genes. The genes expressed at the midpoint and end of infection are middle and late genes, respectively. Cluster III of SP $\beta$  has been proposed to contain the prophage's early genes, with many encoded proteins having homology to other bacteriophage proteins known to be involved in DNA replication. Additionally, over half of the ORFs in Cluster III encode for very short proteins, a feature similar to that of the early gene operons of the lytic bacteriophage sk1, which infects the gram positive bacteria *Lactococcus lactis* (Chandry et al., 1997).

Cluster I has not been categorised based on the temporal expression of its ORFs. The 23 ORFs of Cluster I encode proteins that function in apparently unrelated processes including integration of the SP $\beta$  genome into *B. subtilis* chromosome and production of the sublancin 168 bacteriocin (Lazarevic et al., 1999). The continued maintenance of substantial 134 KB SP $\beta$  genome in *B. subtilis* is potentially due to the advantages it confers to lysogenic strains, including the synthesis of sublancin 168.

Genes expressed during the late stage of phage development encode lytic enzymes. In addition to SP $\beta$ , *B. subtilis* strain 168 is lysogenic for PBSX, a defective prophage. In PBSX, the lytic enzyme is an N-acetylmuramoyl-L-alanine amidase encoded by the *xlyA* gene (Longchamp et al., 1994, Wood et al., 1990). This gene was shown to be part of the late operon along with the gene *xhIB*, which is thought to encode a holin (Wood et al., 1990). Holins act to form pores in membranes, granting bacteriophage lytic enzymes access to the cell wall (Krogh et al., 1998, Wang et al., 2000). SP $\beta$  Cluster II contains the genes *blyA* and *bhIB* which have homology to *xlyA* and *xhIB* of the PBSX late operon, respectively. *In vitro*, BlyA protein was shown to have amidase activity (Regamey and Karamata, 1998). In addition to *blyA*, the *yomI* gene of Cluster II encodes CwIP. Studies have revealed that CwIP has DD endopeptidase and muramidase activities which are capable of degrading the *B. subtilis* cell wall. CwIP is also predicted to function as a bacteriophage tail-fibre protein (Sudiarta et al., 2010). The protein products of *yomR*, *yomQ* and *yomP* of Cluster II are also likely to act as tail-fibre proteins (Lazarevic et al., 1999). The presence of genes encoding tail-fibre proteins is also characteristic of late operons. PBSX follows this pattern, with the tail-fibre protein genes such as *xkdV* (XkdV shares homology to YomR) being within the same operon as *xlyA* and *xhIB*. The similarities shared between SP $\beta$  and PBSX strongly suggest that Cluster II is comprised of genes expressed at the late stage of SP $\beta$  development.

*yonO*, the gene coding for the YonO RNAP is located at the beginning of Cluster II. Therefore, it is likely that YonO is expressed at the late stage of SP $\beta$  development. Typically, bacteriophage encoded RNAPs are responsible for the transcription of the late genes. Therefore, it is likely the *yonO* is transcribed by the endogenous *B. subtilis* RNAP. Once expressed, YonO would carry out transcription of the remaining SP $\beta$  late genes. Despite SP $\beta$  being first identified over 30 years ago and YonO being hypothesised to be an RNAP over 10 years ago, no work has been published which directly investigates transcription in SP $\beta$ . As mentioned above, YonO has very little homology to the multisubunit RNAP and no homology to any previously described single subunit RNAPs such as T7 RNAP. However, we have shown that YonO is a DNA dependant RNAP. From this, many questions regarding the role YonO plays during SP $\beta$  development arise including which genes, if any, are transcribed by YonO. If YonO does

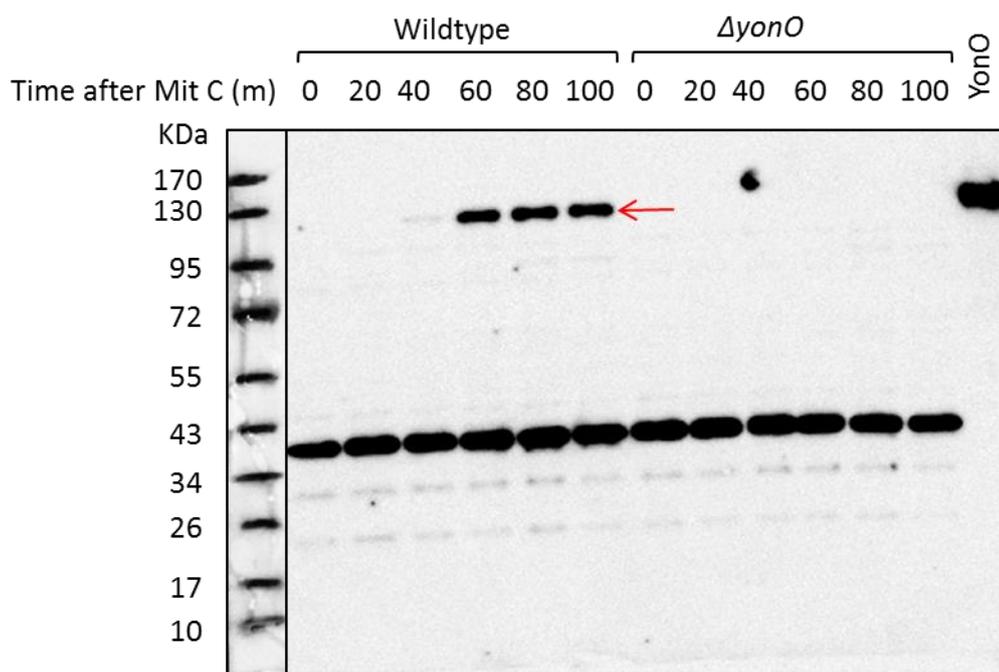
indeed transcribe Cluster II genes, how does it initiate and terminate transcription and how is it regulated? Below is our attempt to begin to answer these questions.

## **6.2. Aim**

The aim of this study was to investigate the role YonO plays in the development of SP $\beta$ . In particular, we wanted to address whether YonO is required for SP $\beta$  development and to determine whether YonO is indeed responsible for the transcription of Cluster II genes. Further to this, we aimed to determine if YonO is capable of transcription initiation and to elucidate potential promoters.

## **6.3. Expression of YonO during SP $\beta$ induction**

Previously, YonO had been classified as a hypothetical protein and the only evidence of expression came from microarray data published online in 2012 (Nicolas et al., 2012). The European-wide Bacillus systems biology collaboration, BaSysBio, investigated the response to environmental conditions at a transcriptomic level in *B. subtilis* (Nicolas et al., 2012). The results were made accessible online in the form of a searchable database. By searching this database, it was apparent that the greatest expression of *yonO* occurred between 45 and 90 minutes after treatment with Mitomycin C. Taking into account that SP $\beta$  is known to be induced by Mitomycin C, the expression of *yonO* in this condition is to be expected. As this data only shows expression at the level of transcription, we sought to confirm that YonO protein is present during SP $\beta$  induction by western blot.



**Figure 6-1: Expression of YonO during SP $\beta$  induction.** Wildtype and  $\Delta yonO$  cells were grown to mid – log phase and treated with  $0.5 \mu\text{g ml}^{-1}$  Mitomycin C. Cells were harvested every 20 minutes after treatment. Samples were disrupted by sonication and equal amounts of lysate ( $5 \mu\text{g}$  of protein) were resolved by SDS PAGE. YonO was detected by probing the PVDF membrane with anti – YonO polyclonal antibodies. Expression of YonO (indicated by red arrow) begins at 40 minutes, with maximum expression occurring after 60 minutes of induction. Additional bands are due to cross reactivity between the primary antibody and proteins present in *B. subtilis* lysate.

To allow visualisation of YonO expression by western blot, polyclonal  $\alpha$ -YonO antibodies were generated by providing purified YonO to the commercial company Eurogentec. YonO was inoculated into two rabbits over the course of 28 days, after which the serum was collected. From the provided serum, we purified polyclonal antibodies against YonO by affinity chromatography using cyanogen bromine activated sepharose resin covalently crosslinked to YonO.

For a negative control for YonO expression in induced WT cells, a  $\Delta yonO$  strain was generated. To do so, the coding region of *yonO* was replaced by homologous recombination with a neomycin resistance cassette. WT and  $\Delta yonO$  cells were treated with Mitomycin C and harvested every 20 minutes over a time course of 100 minutes. Probing with  $\alpha$ -YonO antibodies revealed that YonO was indeed present in the lysates of wild type cells treated with Mitomycin C. YonO expression begins approximately 40 minutes post induction, with maximum occurring between 40 and 60 minutes. YonO expression is maintained throughout the remainder of the time course. At

approximately 120 minutes after Mitomycin C treatment, cell lysis occurs. The observed appearance of YonO is in agreement with the microarray data of Nicolas et al, which shows expression of *yonO* gene occurs between 45 and 90 minutes after treatment with Mitomycin C (Nicolas et al., 2012). The additional bands present on the western blot are due to cross reactivity of the primary antibody with proteins present in the *B. subtilis* lysate. The bands are uniform across both wildtype and  $\Delta yonO$  strains and are present in all time points, revealing they are not YonO degradation products. The uniformity of the additional band allowed us to conclude the lysates were equally loaded.

During the above experiments, we did not observe a difference in the growth of wildtype and  $\Delta yonO$  strains (not shown). Furthermore, both wildtype and  $\Delta yonO$  strains lysed approximately 120 minutes after treatment with Mitomycin C. Both observations were to be expected. Deletion of SP $\beta$  and other prophage elements in the *B. subtilis* genome was shown not to have an impact on growth or physiology of *B. subtilis* cells (Westers et al., 2003). As discussed above, PBSX is also induced upon treatment of cells with Mitomycin C which results in the expression of lytic enzymes including the XlyA amidase (Longchamp et al., 1994). Therefore, the lysis of  $\Delta yonO$  cells cannot be used to determine if YonO is essential for SP $\beta$  development.

#### **6.4. Requirement of YonO for SP $\beta$ development**

When induced, PBSX is capable of causing host cell lysis (Seaman et al., 1964). This prevented us from using lysis of Mitomycin C treated cells to determine if YonO is required for SP $\beta$  induction. Therefore, to determine if YonO is essential for SP $\beta$  development and phage particle release, a strain sensitive to SP $\beta$  lysogenisation was used. Strain 1A100 is a derivative of the Su<sup>3+</sup> strain, an isolated strain spontaneously cured of SP $\beta$  and therefore susceptible to lysogenisation by SP $\beta$  particles (Zahler et al., 1977, Warner et al., 1977). When treated with Mitomycin C, wildtype *B. subtilis* cells lyse. As SP $\beta$  development is triggered by Mitomycin C, SP $\beta$  particles are present in the lysates of wildtype cells. When incubated with an SP $\beta$  cured strain such as 1A100, the SP $\beta$  particles lysogenise the cured cells resulting in small turbid plaques on a lawn of cured cells. The plaques are a result of inefficient lysogenisation of cells by the SP $\beta$  particles. The SP $\beta$  particles which do not lysogenise the cured cells undergo the lytic life cycle.

The lysate derived from lysis of WT cells was briefly incubated with A100A cells and plated. After overnight growth, small turbid plaques appeared in the lawn of 1A100 cells as a result of SP $\beta$  lysogenising the susceptible strain (Figure 6-2). In contrast, no plaques were observed when lysate from induced  $\Delta$ SP $\beta$  strain was incubated and subsequently plated with strain 1A100. When plated with strain 1A100, the lysate from  $\Delta$ yonO strain also did not produce any plaques. The inability of  $\Delta$ yonO strain to produce SP $\beta$  particles capable of lysogenising 1A100 strain indicates that YonO is essential for SP $\beta$  particle development which would be expected if YonO is a bacteriophage RNAP.



**Figure 6-2: YonO is essential for SP $\beta$  production *in vivo*.** Wildtype,  $\Delta$ SP $\beta$  and  $\Delta$ yonO strains were induced by Mitomycin C and subsequent lysates were tested for the presence of SP $\beta$  particles capable of forming plaques in a lawn of SP $\beta$  sensitive cells (Strain 1A100). Only lysates from the wildtype strain produced plaques, suggesting YonO is required for the production of SP $\beta$  in induced cells.

### 6.5. Identification of Genes Transcribed by YonO

*In vitro*, YonO displays efficient RNA polymerase activity. This, in combination with its expression during SP $\beta$  induction and apparent essentiality strongly suggests YonO is an SP $\beta$  specific RNAP functioning during the induction of the prophage.

Whilst YonO appears to be essential for SP $\beta$  development, the genes it transcribes remain unidentified. To address this, total transcriptome analysis (RNAseq) was performed on WT and  $\Delta$ yonO strains. Comparison of the WT and  $\Delta$ yonO transcriptomes during bacteriophage induction would reveal the genes under the control of YonO. Wildtype and  $\Delta$ yonO strains were grown to mid-log phase (0.5 OD<sub>600</sub>) and SP $\beta$  was induced with Mitomycin C. Western blot data revealed YonO is present in WT cells 60 minutes after induction. Additionally, SP $\beta$  induction was observed 60

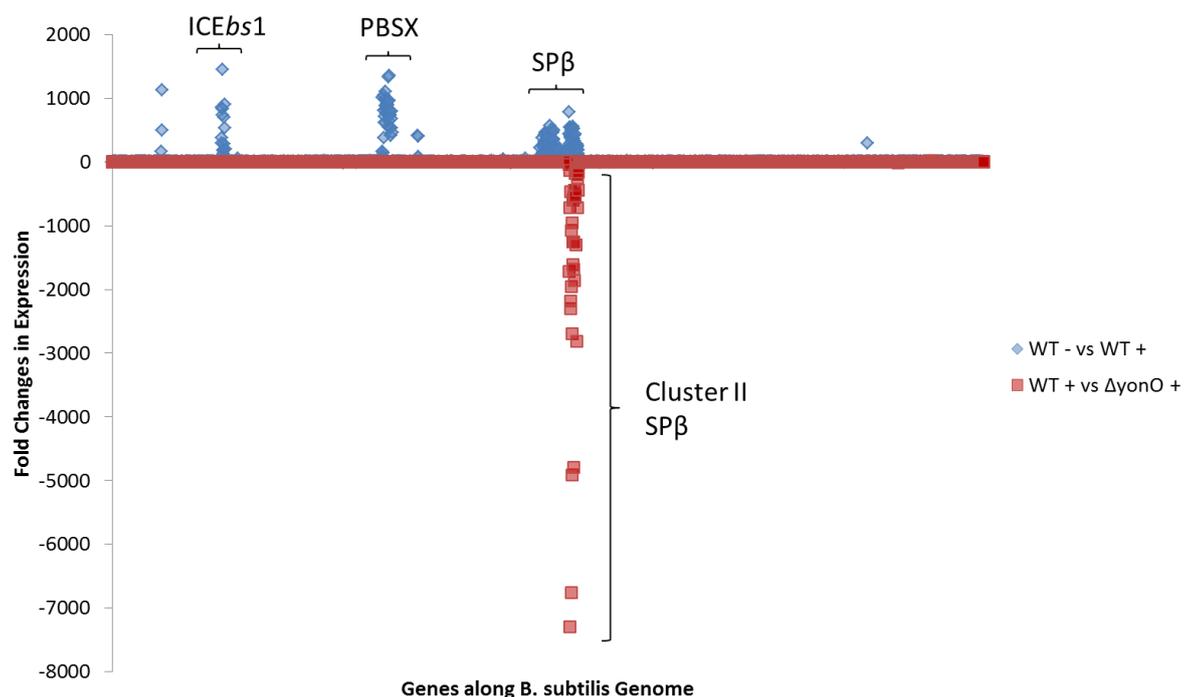
minutes after Mitomycin C treatment in previously published microarray data (Goranov et al., 2006). Therefore, cells were harvested 60 minutes after treatment with Mitomycin C. In parallel, wildtype and  $\Delta yonO$  cells without treatment were harvested along with biological replicates for all samples. Total RNA was extracted from the cells and subjected to next generation sequencing. The commercial company PrimBio performed all subsequent handling of total RNA and RNA sequencing. Total RNA contains rRNA which strongly skews RNAseq due to its abundance. Therefore, 16S and 23S rRNA was depleted prior to library preparation. Library preparation involved the ligation of adaptors and sample specific barcodes to the RNA followed by reverse transcription and purification of cDNA. Sequencing primers hybridise to the adaptor sequences whilst barcode sequences allow for multiplexing, the simultaneous sequencing of multiple samples. The sequencing reads were aligned to the *B. subtilis* reference genome and quality controlled via read metrics such as mapping quality and alignment score. For each annotated gene, a normalised expression value was calculated. The expression of each gene in the two different strains (wildtype and  $\Delta yonO$ ) and each condition (+/- Mitomycin C treatment) was compared, giving fold changes in expression. This data was provided by PrimBio.

Comparing the changes in gene expression of the wild type strain, with and without Mitomycin C treatment, the wide ranging impact of Mitomycin C could be seen (Figure 6-3). Primarily, Mitomycin C treatment induces the SOS response through DNA damage (Goranov et al., 2006). One aspect of the SOS response is the activation of the nucleotide excision repair machinery, comprised of UvrA, UvrB and UvrC (reviewed in Lenhart et al, 2012). Mitomycin C treatment led to an approximate 9 fold increase in the transcription of the *uvrA* and *uvrB* genes. A 4 fold increase in *uvrC* was observed. The pattern of lower induction of *uvrC* compared to *uvrA* and *uvrB* correlates with previous transcriptomic data (Au et al., 2005).

The prominent effect of Mitomycin C treatment is the induction of resident prophages PBSX, SP $\beta$  and the ICE $bs1$  mobile element through activation of the SOS response (Goranov et al., 2006, Auchtung et al., 2005). The RNAseq data correlated well with previously published data (Goranov et al., 2006, Seaman et al., 1964, Warner et al., 1977). Upon Mitomycin C treatment, expression of PBSX, SP $\beta$  and ICE $bs1$  genes was greatly increased as shown in Figure 6-3. Despite possessing homology to the late

operon of PBSX, the SKIN element was not induced by Mitomycin C. This observation is consistent with a previously reported absence of  $\beta$ -galactosidase activity in Mitomycin C treated cells with lacZ fusions to the SKIN late operon (Krogh et al., 1996). The 3 genes upstream of ICEbs1 displaying increased expression in response to Mitomycin C correspond to *ybfE*, *ybfG* and *ybfG*. The proteins encoded by these genes are uncharacterised but bioinformatics have shown that they contain peptidoglycan binding and hydrolase domains.

Looking closer at the induction of SP $\beta$  in response to Mitomycin C, RNAseq revealed that genes of Cluster II and III underwent increased expression whilst Cluster I did not. No Cluster I genes exhibited a significant fold change in expression when WT cells treated with and without Mitomycin C were compared. Referring to the BaSysBio database supported this observation (Nicolas et al., 2012). *SunT*, a representative gene of Cluster I, had equal expression levels in cells treated with Mitomycin C for 45 minutes and the negative control.



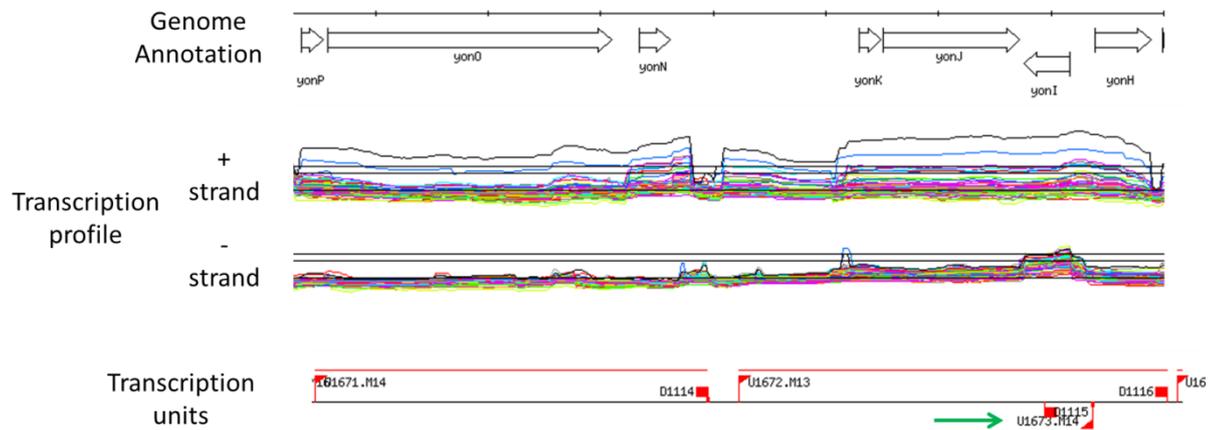
**Figure 6-3: Fold changes of expression across the *B. subtilis* genome in response to Mitomycin C.** The fold changes in expression of individual genes in WT cells, with and without Mitomycin C treatment (blue diamond) were plotted. The KEGG BSU accession number for each gene was used to plot along the X axis, showing the fold changes across the genome. The fold changes of expression between WT and  $\Delta yonO$  cells treated with Mitomycin C were plotted in parallel (red square). Negative values

correspond to a decrease in expression. The regions along the genome corresponding to PBSX, SP $\beta$  and ICEBs1 have been indicated.

To determine which genes are under the control of YonO, expression levels of wildtype and  $\Delta yonO$  cells treated with Mitomycin C were compared. There were no fold changes in Cluster I and III expression (Figure 6-3). The RNAseq revealed a large disparity between the wildtype and  $\Delta yonO$  strains with regards to the expression of Cluster II genes after Mitomycin C treatment, with comparisons of expression levels revealing large fold changes. The fold changes were negative; indicating Cluster II genes were not expressed in the induced  $\Delta yonO$  strain and suggesting that YonO is responsible for the transcription of Cluster II. Furthermore, the overall strength of YonO expression appears to be greater than that of the endogenous, host RNAP. This conclusion was arrived at by comparing the normalised expression values generated from the RNAseq data. Using the expression of *rpoC* (gene encoding for  $\beta'$  subunit) as an example of a gene highly expressed by the host RNAP, this gene has a normalised expression value of 0.56. This is in stark contrast to the expression value of the *yomI* gene of SP $\beta$  Cluster II, which has a normalised expression value of 9.9. It is appreciated the *rpoC* expression value was generated from cells treated with Mitomycin C. Cells treated by Mitomycin C are subjected to DNA damage and undergoing the SOS response. Therefore, additional work will be required to accurately compare the *in vivo* expression strength of YonO to other RNAPs such as the bacterial RNAP and T7 RNAP.

*yonO* is predicted to be the middle gene in an operon comprised of *yonP*, *yonO* and *yonN* (Figure 6-4) based on transcription profiles constructed from the high resolution BaSysBio microarray data (Nicolas et al., 2012). In the transcription profiles, sudden increases and decreases of transcription, referred to as upshifts and downshifts, were used to define transcription units. *yonP*, *yonO* and *yonN* were shown to be on the same transcription unit which indicates they belong to a single operon, the expression of which would be driven by a single promoter upstream of *yonN* (Figure 6-4). All three genes display greatly increased expression upon Mitomycin C treatment in the wildtype cells. However, in  $\Delta yonO$  cells treated with Mitomycin C, only *yonP* expression was comparable to wild type cells. There was no expression of *yonO* and *yonN* expression was reduced. This observation supports the prediction these three genes belong to an operon. In the  $\Delta yonO$  strain, the *yonO* coding sequence has been

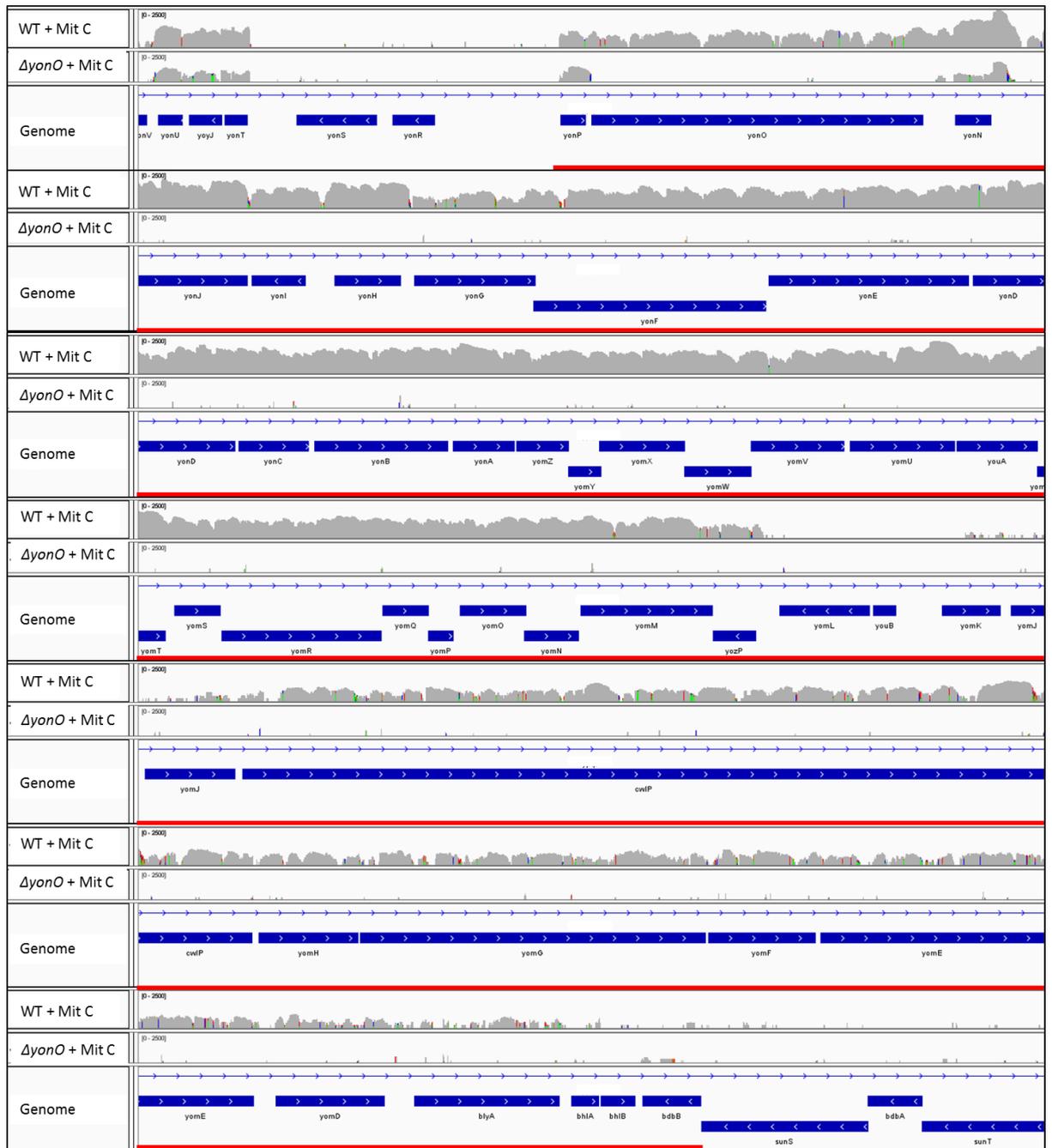
replaced with the kanamycin resistance cassette, which explains the complete absence of *yonO* expression in this strain. The kanamycin resistance cassette likely disrupts the transcription of the downstream *yonN* gene which would account for the reduction in *yonN* expression in the  $\Delta yonO$  strain. *yonO* belonging to an operon with *yonP*, in combination with the observation that *yonP* expression is not lessened in the  $\Delta yonO$  strain, suggests *yonO* is transcribed by the *B. subtilis* RNAP during SP $\beta$  development.



**Figure 6-4: Transcription profile of *yonO* and adjacent regions of SP $\beta$ .** Data retrieved from BaSysBio database. **Top:** *yonO* and the adjacent Cluster II genes are shown in the GenBank annotation of the wildtype *B. subtilis* genome. **Middle:** Transcription profiles generated from microarray data with a resolution of 22 bases. Profiles for both DNA strands are shown. The individual traces represent the transcription profiles for individual environmental conditions. **Bottom:** Transcription units were defined by sudden upshifts and downshifts of transcription, as observed in the transcription profiles. Upshifts and downshifts are depicted by red triangles and red squares, respectively. Transcription units above the genome (black line) are those on the plus strand. Transcription units on the – strand are below the genome. *yonP*, *yonO* and *yonN* exist on a single transcription unit started by upshift 1672. Downstream, *yonI* is present in a transcription unit on the minus strand with the corresponding upshift 1673 indicated by a green arrow. *yonI* is the sole ORF of the transcription unit and is therefore not expressed as part of any operon on either strand of DNA.

The alignment of wildtype and  $\Delta yonO$  RNAseq sequencing reads to the Cluster II region on the SP $\beta$  genome was used to generate a coverage map which can be seen in Figure 6-5. The coverage map revealed that the Cluster II genes downstream of *yonN* were not transcribed in Mitomycin C treated  $\Delta yonO$  cells. This is in contrast to induced wildtype cells, in which Cluster II genes were expressed. Therefore, the RNAseq indicates *YonO* is responsible for the transcription of Cluster II genes. Cluster II is suggested to contain late genes which are expressed towards the final stages of SP $\beta$  development, including lytic enzymes and bacteriophage structural proteins. Despite

being unrelated to any known examples, it appears YonO is similar to other bacteriophage encoded RNAPs in that it transcribes late genes.



**Figure 6-5: Coverage map of RNAseq reads aligned to the annotated SPβ Cluster II DNA.** The sequencing reads from wildtype (WT) and  $\Delta yonO$  cells treated with Mitomycin C were aligned to the annotated *B. subtilis* genome. The red underlining of annotated genes indicates those belonging to Cluster II. Differences between the reference sequence and read sequences are shown in red (thymine), blue (cytosine), yellow (guanine) and green (adenine). The number of reads was plotted onto a logarithmic scale.

## 6.6. Investigation of Transcription Initiation by YonO

### 6.6.1. Transcription initiation at the *yonI* predicted start site

One aspect common to all RNAPs is an ability to begin *de novo* synthesis of RNA on double stranded DNA. Whilst all multisubunit RNAPs require transcription factors to achieve promoter binding and DNA melting, single subunit RNAPs such as T7 RNAP can initiate transcription independently. We decided to determine if YonO is capable of independently initiating transcription on double stranded DNA. Any transcription initiation by YonO would raise many interesting, fundamental questions regarding the mechanism in which it accomplishes promoter binding and DNA melting.

RNAseq suggested transcription of *yonI* is controlled by YonO. Whilst most Cluster II genes appear to be transcribed by YonO, *yonI* is encoded on the minus strand and does not belong to any operons (Figure 6-4). This implies a YonO promoter could be upstream of the *yonI* coding sequence. According to BaSysBio transcription profile, upshift 1673 corresponds to a predicted transcription start site upstream of *yonI*. Therefore, 100 bp of DNA sequence upstream and downstream of the upshift 1673 position was used as a template for *in vitro* transcription to determine if YonO can initiate transcription.

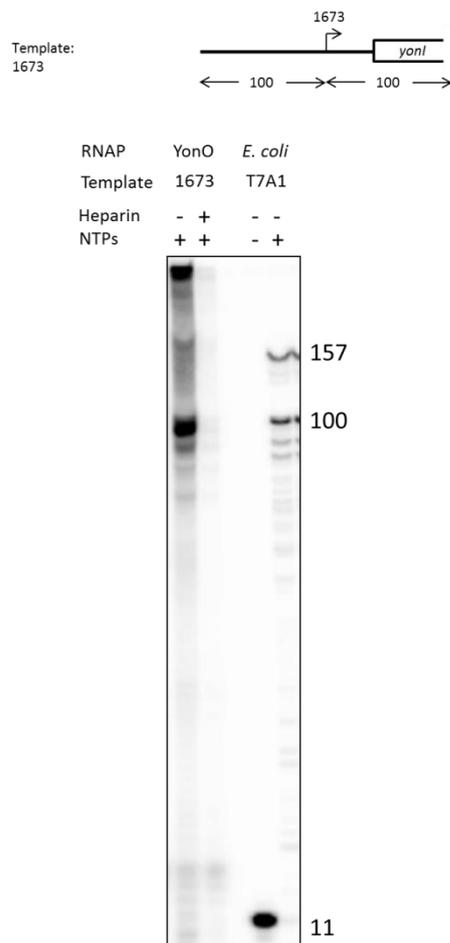
To see if YonO was able to start transcription on upshift 1673 template, YonO was incubated with the template DNA and NTPs at 37°C for 30 minutes. Synthesised transcripts were labelled through incorporation of  $\alpha$ -[<sup>32</sup>P]-UMP. YonO transcribed on the 1673 template to synthesise a transcript approximately 100 nt in length (Figure 6-6). The length of the transcript was determined through comparison to the well-studied tr2 terminator transcript produced by *E. coli* RNAP transcription initiated on the T7A1 template DNA, which gives transcripts 100 nt and 157 nt in length. The length of the transcripts indicates YonO is initiating transcription at the centre of the 1673 DNA template, possibly at the predicted *yonI* start site. *In vitro* transcription reactions were also performed in the presence of 5  $\mu\text{g ml}^{-1}$  heparin, which was added simultaneously with NTPs. Heparin competes with DNA to bind RNAP and remove complexes that result from non-specific interactions between RNAP and DNA. Furthermore, heparin binds to free RNAP and prevents multiround transcription. Despite high levels of potentially specific transcription, transcription on 1673 DNA by YonO was diminished in the presence of heparin. From this experiment it cannot be

determined if heparin destroyed unstable YonO complexes or restricted transcription to a single round.

The direction of transcription on the 1673 template could not be determined in the previous experiment and as such, transcription initiating in the vicinity of 1673 upshift may occur in the reverse direction, away from the *yonI* gene. To determine the direction of transcription, 50 bp of sequence was added to either end of the DNA template. An increase in the transcript length would reveal the direction of transcription. The addition of 50 bp to the downstream end of the template resulted in an increase in transcript length from 100 nt to approximately 150 nt (Figure 6-7). Varying the incubation times from 30 seconds to 30 minutes showed the transcripts accumulated slowly over time due to multiple rounds of transcription. As the transcripts synthesised by YonO in 30 seconds are comparable to those produced in 30 minutes of incubation in the presence of heparin, it suggests heparin reduces transcription by restricting YonO to a single round of transcription.

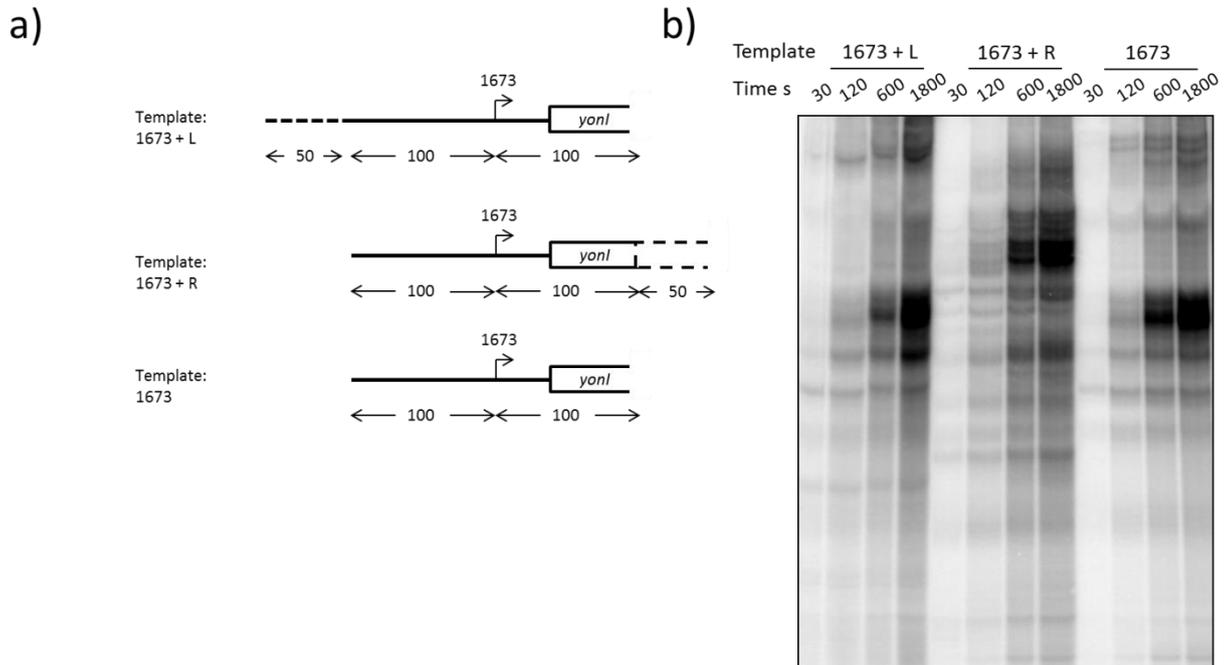
The run off transcripts produced by transcription initiation in the vicinity of the 1673 upshift appear to be heterogeneous (figure 6-7b). The differences in transcript length are mostly likely due to YonO using multiple transcription start sites at the *yonI* promoter. Alternatively, the heterogeneity may be due to minor variations in the overall length of the DNA template. These slight variations would have been introduced to the 5' ends of the DNA template during PCR amplification due to errors at the 5' end of the oligonucleotide primers. These errors would have been introduced during the 3' to 5' synthesis of the oligonucleotides.

The main conclusion to be drawn these two experiments is that YonO initiates transcription on double stranded DNA in a factor independent manner. Transcription factors are obligatory for multisubunit RNAPs to initiate transcription. To our knowledge, YonO is the first homologue of the multisubunit RNAP capable of transcription initiation without the need for accessory factors. This observation raises many questions, as initiation of transcription independent of other factors implies YonO has a previously undescribed method of recognising and melting duplex DNA. The experiments suggest YonO initiates transcription in the vicinity of the 1673 upshift, the predicted *yonI* transcription start site. Furthermore, YonO transcribes with a downstream direction towards the *yonI* coding sequence suggesting YonO is indeed responsible for transcription of *yonI*.



**Figure 6-6: Transcription initiation by YonO on upshift 1673 DNA.** DNA containing upshift 1673, the predicted *yonI* transcription start site, was used as a template for YonO transcription initiation. YonO was incubated with DNA before the addition of NTPs (500  $\mu$ M ATP, GTP, CTP, 50  $\mu$ M UTP and 0.15  $\mu$ M  $\alpha$ -[ $^{32}$ P]-UTP). Transcription reactions were incubated for 30 minutes at 37  $^{\circ}$ C and the synthesised transcripts were

resolved by 15% denaturing PAGE. 100 nt transcripts synthesised by *E. coli* RNAP elongation complexes terminating at the tr2 terminator on T7A1 template DNA were used to measure the length of transcripts synthesised by YonO.



**Figure 6-7: Direction of transcription on the upshift 1673 template DNA. a) Upshift 1673 DNA templates.** Upshift 1673 template DNA was extended on either side by 50 bp to give 1673 + L and 1673 + R templates. The lengths of DNA (in bp) are shown below each template, with the additional 50 bp indicated by the dashed line. **b) Direction of YonO transcription on the 1673 template.** YonO was incubated with each template and transcription was initiated by the addition of NTPs (500  $\mu$ M ATP, GTP, CTP, 50  $\mu$ M UTP and 0.15  $\mu$ M  $\alpha$ -[ $^{32}$ P]-UTP). Reactions were allowed to proceed for up to 30 minutes at 37  $^{\circ}$ C before the addition of formamide containing stop buffer. Transcripts were resolved by 15% denaturing PAGE.

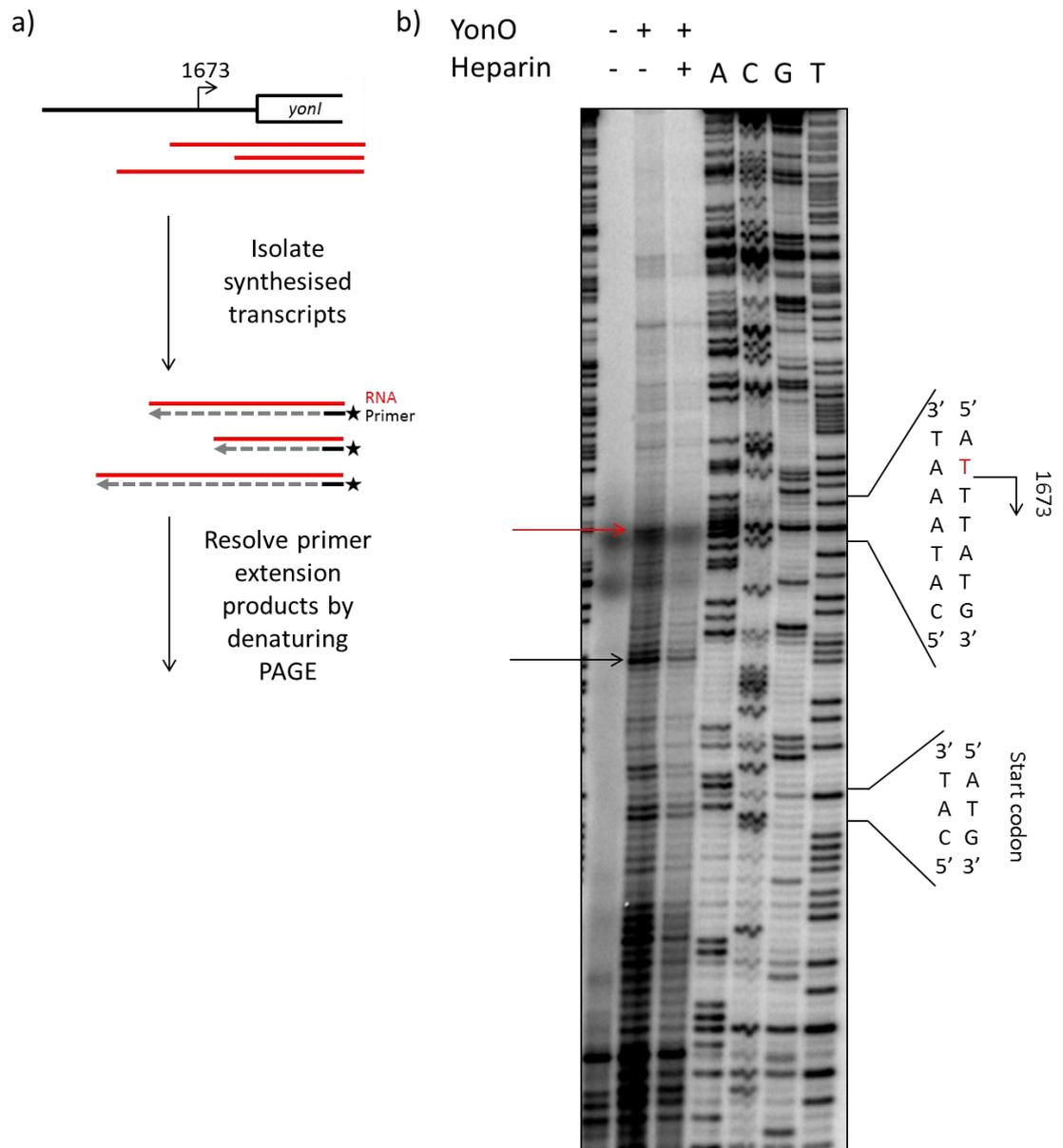
#### 6.6.2. Primer extension to confirm transcription initiation at the predicted *yonI* start site

To determine if YonO is initiation transcription at the predicted *yonI* start site, primer extension was carried out on transcripts synthesised from 1673 template DNA using 5'  $^{32}$ P labelled primer 1673 R. Primer 1673 R was complementary to the 3' end of transcripts produced by downstream run off transcription, as shown in Figure 6-7. In order to calculate the DNA sequence corresponding to the transcription start sites, dideoxy chain terminator sequencing of 1673 template DNA was also carried out using primer 1673 R. The site of the 1673 upshift was identified in the sequencing lanes. According to primer extension, transcripts were produced from the initiation of

transcription at the 1673 upshift (red arrow in Figure 6-8). However, there is not an over-abundance of primer extension products from the 1673 upshift, despite this transcript being the most synthesised in the *in vitro* transcription reaction. The primer extension suggests transcription also started approximately 20 bp downstream from upshift 1673 (shown by black arrow, Figure 6-8), possibly indicating multiple start sites are utilised for transcription of *yonI*. According to the BaSysBio transcription profile, there is only one transcription start site upstream of *yonI*. However, since the resolution of the tiled microarray is 22 bases, an additional start site in such close proximity to upshift 1673 may not have been identified. A radioactive signal migrating with upshift 1673 transcripts can be seen in all lanes. This signal is even present in the absence of YonO as so is likely due to a contaminant carried over from the synthesis or 5' <sup>32</sup>P labelling of the primer.

Looking at Figure 6-6, *In vitro* transcription on the 1673 template also produced short transcripts. Consistent with this, primer extension suggests there are additional transcription start sites on the 1673 template that are near to the start codon of *yonI*. The *yonI* start codon is shown in the sequencing of 1673 template DNA in Figure 6-8b.

In summary, the primer extension does suggest YonO initiates transcription at upshift 1673. However, the presence of additional bands indicative of transcription start sites downstream of upshift 1673 prevent us from concluding that upshift 1673 is a definite YonO transcription start site.



**Figure 6-8: Primer extension to determine YonO transcription start sites. a) Schematic of primer extension.** The 1673 DNA template shown, featuring the putative *yonI* transcription start site was used to generate transcripts (shown in red) for primer extension. Transcription reactions were also carried out in the presence of  $5 \mu\text{g ml}^{-1}$  heparin.  $5'$   $^{32}\text{P}$  labelled Primer 1673 R (shown in black) used for extension was complementary to all RNA produced by run off transcription downstream of the *yonI* promoter. **b) Primer extension of 1673 derived transcripts.** cDNA products of primer extension were resolved on a 6% denaturing PAGE gel alongside dideoxy terminated DNA sequencing of 1673 using primer 1673 R. The sequences of upshift 1673 and the start codon of *yonI* are shown. The red and black arrows indicate RNA synthesised by transcription starting at upshift 1673 and approximately 20 bp downstream, respectively.

### 6.6.3. $\text{KMnO}_4$ footprinting to observe DNA melting at the predicted *yonI* transcription start site

During initiation, RNAP holoenzyme binds to the promoter DNA to form a closed promoter complex. Through the action of  $\sigma^{70}$ , the promoter DNA is melted and the template strand is positioned into the active centre of RNAP. Melted DNA bound by RNAP is referred to as the open promoter complex. The formation of open promoter complexes can be observed by  $\text{KMnO}_4$  footprinting. In this technique, the thymine bases in the melted DNA are modified by  $\text{KMnO}_4$  and subsequently cleaved by treatment with piperidine. The pattern of cleaved DNA, resolved by denaturing PAGE, can be interpreted to determine the position of open promoter complexes on the DNA template.

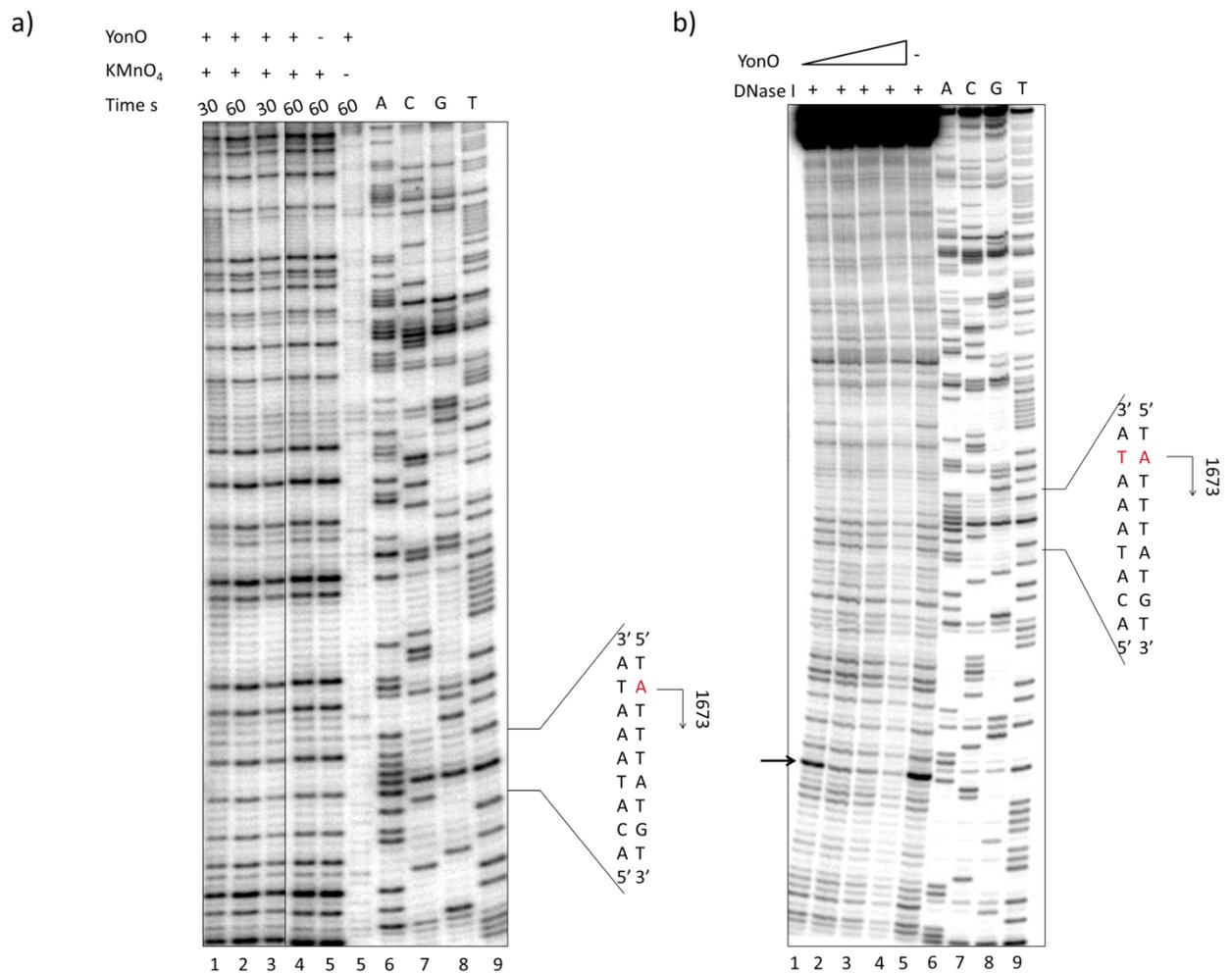
Therefore,  $\text{KMnO}_4$  footprinting of YonO incubated with the 1673 template was performed to determine if YonO is able to melt DNA and form a stable open promoter complex at the predicted *yonI* transcription start site. To detect the cleaved DNA, the 5' end of the 1673 DNA template strand was  $^{32}\text{P}$  labelled through PCR using 5'  $^{32}\text{P}$  labelled primer 1673 R.

Surprisingly, given the efficient transcription on the 1673 template, comparisons of  $\text{KMnO}_4$  footprints in the presence and absence of YonO revealed no DNA melting occurred (Figure 6-9a). Heparin was added to investigate the stability of any potential YonO open promoter complexes. Since open promoter complexes were not seen, heparin did not alter the footprint. The absence of evidence for DNA melting in the  $\text{KMnO}_4$  footprint implies that YonO open promoter complexes are highly unstable and transient. This characteristic was unexpected and is in contrast to the multisubunit RNAP which forms stable open promoter complexes that can be observed through  $\text{KMnO}_4$  footprinting (Nechaev et al., 2000). An exception to this is the RNAP of *T. aquaticus*.  $\text{KMnO}_4$  footprinting fails to detect *T. aquaticus* RNAP open promoter complexes suggesting the complexes are unstable (Andrey Kulbachinskiy, personal communication). It has also been proposed T7 RNAP open complexes are unstable, with evidence suggesting the initiating nucleotides are needed for stabilisation of open complexes (Villemain et al., 1997). Therefore, YonO may be similar to T7 RNAP and *T. aquaticus* RNAP in that its open promoter complexes are highly unstable and transient.

Even in the absence of YonO, the DNA template was modified by  $\text{KMnO}_4$ , resulting in cleavage by piperidine. This gave background signal, making the identification of DNA cleavage due to YonO open promoter complex formation difficult. To reduce this background in future experiments, the modification of melted DNA by  $\text{KMnO}_4$  could be probed by primer extension rather than piperidine cleavage. A 5'  $^{32}\text{P}$  primer annealed to the 3' end of the  $\text{KMnO}_4$  modified DNA strand can be extended by DNA polymerase. However, primer extension is halted as the DNA polymerase encounters the modified DNA. Upon resolving the primer extension products by denaturing PAGE, the site of open promoter complex formation can be determined.

#### *6.6.4. DNase I footprint to observe YonO closed promoter complex formation*

Since we were unable to detect open promoter complex formation, we investigated YonO's ability to bind the 1673 template and form closed promoter complexes by DNase I footprinting. Protein bound DNA is protected from endonucleolytic cleavage by DNase I. Resolving the cleavage products by denaturing PAGE gives a footprint pattern which reveals the position of protein binding on the DNA. To visualise the cleavage products, 1673 template DNA was  $^{32}\text{P}$  labelled at the 5' terminus of the non-template strand. YonO was incubated with 1673 template DNA at room temperature prior to the addition of DNase I. The incubation temperature can influence the binding of YonO to the template DNA; a reduced temperature (room temperature) would slow the disassociation of YonO from the template and increase detectable protection against DNase I. Titrating the concentration of YonO revealed a very weak protection of the DNA from endonucleolytic cleavage. A detectable reduction in DNA cleavage was only observed when YonO was at a 20 times excess over DNA (lane 4, Figure 6-8b). This shows that YonO is only capable of weakly binding to the double stranded DNA. From this experiment alone, it cannot be concluded that this weak binding is non-specific. However, given the protection from DNase I was not limited to a discrete region of the DNA template; this appears to be the case.



**Figure 6-9: DNA footprinting of YonO on upshift 1673 DNA. a) KMnO<sub>4</sub> footprint.** YonO was incubated with 1673 DNA 5' <sup>32</sup>P labelled on the template strand for 10 minutes at 37 °C. KMnO<sub>4</sub> was added and reactions were allowed to proceed for the indicated time before addition of β-mercaptoethanol and piperidine treatment. YonO and KmnO<sub>4</sub> were omitted for negative controls. Dideoxy DNA sequencing on the 1673 was performed using 5' <sup>32</sup>P labelled primer 1673 R. Reactions were resolved on 6 % denaturing PAGE gel and visualised by phosphorimaging. Sequence of upshift 1673 (red) on the template strand is shown. **b) DNase I footprint.** YonO was incubated with 1673 DNA (labelled as before) for 10 minutes at room temperature. YonO was mixed with 1673 DNA at a ratio of 1:1, 5:1, 10:1 and 20: 1. DNase I was added for 2 minutes at room temperature. Reactions were stopped by addition of stop buffer and resolved on 6 % denaturing PAGE gel. To identify sites of DNA protection, dideoxy DNA sequencing reactions were loaded. The arrow indicates a preferential DNase I cleavage site.

## 6.7. Discussion

Whilst showing YonO is a functional enzyme, *in vitro* experiments did not shed light on the involvement of YonO in SP $\beta$  induction. Previously, YonO was classified as a hypothetical protein and the only evidence for its expression was microarray data (Nicolas et al., 2012). Here we have directly shown that YonO is expressed during development of SP $\beta$ . Furthermore, our RNAseq data suggests YonO is responsible for transcription of the SP $\beta$  late genes. As it is completely unrelated to T7 RNAP, these results suggest that YonO is a novel type of bacteriophage RNAP.

Bacteriophages vary widely in their strategies used for gene expression. Many bacteriophages manipulate the host bacterium's RNAP for the expression of their early and middle genes before utilising their own RNAP to express late genes. The work presented here suggests that SP $\beta$  falls into this category of bacteriophages, with YonO being employed for the expression of the SP $\beta$  late genes. In this regard, SP $\beta$  is comparable to the highly studied T7 coliphage.

An alternative method of gene expression is the total hijacking of the host RNAP for the transcription of all bacteriophage genes. T4 coliphage is a well-studied example of a bacteriophage which does not encode its own RNAP. The biology of T4 bacteriophage is reviewed extensively in Miller et al., 2003. To enable the complete control of the host RNAP, T4 bacteriophage encodes a wide range of additional proteins which target the host RNAP to the bacteriophage promoters at the correct times during infection. For instance, to express its late genes, T4 encodes for its own  $\sigma$  factor. However, this alone is not sufficient and so T4 encodes many more proteins that contribute to the expression of late genes. These include proteins that mimic  $\sigma$  factors and bind the host RNAP to prevent the transcription of host genes and a DNA clamp which facilitates in the opening of late promoter DNA duplexes. This clamp itself is loaded onto DNA by a heterodimer whose proteins are also encoded by T4. All of this is in stark contrast to the T7 bacteriophage, which accomplishes late gene expression by simply encoding for its own RNAP. This simple and economical method of gene expression is one advantage conferred to bacteriophages by possessing their own RNAP, as the bacteriophage is not encumbered with genes coding for the host RNAP hijacking machinery.

Another significant advantage of encoding an RNAP is the bypassing of host RNAP regulation. Host RNAPs are subjected to a multitude of regulations and processes that impact on transcription. These include regulation of transcription initiation, elongation speeds, and termination by factors such as Rho. Even in the absence of these external regulations, multisubunit RNAPs are impeded by events such as pausing and backtracking that are brought about by the template DNA sequence (Bochkareva et al., 2012). As bacteriophage RNAPs are not affected by these regulations, they are free to rapidly transcribe long operons. Indeed, the T7 RNAP has a much greater rate of elongation than the multisubunit RNAP of its host (Tahirov et al., 2002). Previous experiments (shown in chapter 5) suggest YonO is also faster than multisubunit RNAPs, with a potentially lower  $K_m$  too. If this is indeed the case, this would provide an additional advantage, as YonO could still achieve rapid transcription in low NTP conditions.

The conclusion that YonO is essential for SP $\beta$  development and is required for Cluster II transcription is based on Cluster II not being comprised of a single operon. A single operon under the control of one promoter upstream of *yonP* would be disrupted by the replacement of *yonO* with a kanamycin cassette, explaining the complete absence of Cluster II gene expression in  $\Delta yonO$ . The transcription profile calculated from the BaSysBio microarray data does not support this. Rather, it suggests *yonO* is only in an operon with *yonP* and *yonN*. *yonI* of Cluster II is coded on the minus stranded and does not belong to any operon. The fold changes calculated from RNAseq data indicate *yonI* is only expressed in wildtype cells, not in  $\Delta yonO$ , in response to Mitomycin C. This further supports the conclusion YonO is required for the transcription of Cluster II genes. However, to allow us to confirm that YonO is definitely responsible for Cluster II expression, a scarless  $\Delta yonO$  strain will be constructed. In this strain, transcription of Cluster II would not be disrupted by the insertion of a kanamycin resistance cassette.

In addition to a scarless  $\Delta yonO$  mutant, the conclusion that YonO is essential for SP $\beta$  induction and transcription of Cluster II could also be verified by complementation experiments. In these experiments, a copy of the *yonO* gene under the control of an inducible promoter would be introduced into a  $\Delta yonO$  strain through homologous recombination elsewhere on the chromosome. If YonO is indeed responsible for Cluster II transcription, the simultaneous induction of YonO expression and SP $\beta$  in this

strain would result in the restoration of Cluster II transcription and SP $\beta$  particle production. The state of Cluster II transcription could be observed by analysis of the transcripts at either a single gene level (Northern blot or qPCR) or at the transcriptomic level (RNAseq). Furthermore, the production of SP $\beta$  particles could be assayed by incubating the lysates of the induced cells with an SP $\beta$  sensitive strain, as performed in section 6.4.

It was reasoned a YonO promoter would be upstream from the *yonI* coding sequence as *yonI* is independent from all other Cluster II genes and according to RNAseq, is transcribed by YonO. Therefore, the DNA surrounding upshift 1673, a predicted transcription start site was used as template DNA to determine if YonO is able to initiate transcription. Surprisingly, YonO was able to initiate transcription on double stranded DNA. Furthermore, YonO initiated transcription in the vicinity of the predicted *yonI* and transcribed downstream towards the *yonI* coding sequencing. Whilst supporting the notion that YonO initiates transcription at the predicted start site, primer extension also suggested the existence of multiple start sites downstream of upshift 1673. Therefore, further work is needed before it can be concluded that this is a genuine YonO transcription start site.

As YonO is capable of initiating transcription, future work will also involve determining the promoter sequence recognised by YonO. In addition to confirming that YonO is required for expression of Cluster II genes, future RNAseq experiments will contribute to identifying transcription start sites and promoter elements. The first nucleotide of a transcript retains its triphosphate moiety. Therefore, transcription start sites can be identified with RNAseq by selecting transcripts featuring 5' triphosphates during library construction (Sharma et al., 2010). Assuming YonO transcription start sites are identified, comparisons of the upstream DNA sequences may allow for the identification of promoter elements. As YonO has been observed initiating transcription at the predicted *yonI* transcription start site, this too can be utilised for the identification of promoter elements. An *in vitro* approach allows control over the transcription reaction and so the DNA sequence upstream of the *yonI* transcription start site could be modified. By observing the effects the modifications have on transcription, the *yonI* promoter may be identified. Furthermore, these alterations to the template DNA would allow us to delineate the specific sequence of the promoter.

Given its relationship to the multisubunit RNAP, YonO initiating transcription on double stranded DNA was not expected. As discussed previously, multisubunit RNAPs are dependent on transcription factors for initiation. In addition to recognising and melting promoter DNA, transcription factors are required for the binding of nucleotides, initial catalysis and promoter escape (Kulbachinskiy and Mustaev, 2006, Sainsbury et al., 2013). Therefore YonO independently initiating transcription raises many fundamental questions that will form the basis of future work. One of the major questions to be answered is how YonO, a 100 KDa protein, is able to independently initiate transcription whilst the 500 KDa multisubunit RNAP cannot. In being able to bind and melt duplex DNA independent of other factors, YonO is analogous to T7 RNAP. T7 RNAP utilises two structural motifs, the specificity loop and the intercalating hairpin to accomplish these tasks. Residues of the specificity loop recognise the T7 promoter DNA whilst the intercalating hairpin separates the DNA strands (Yin and Steitz, 2002, Cheetham and Steitz, 1999). It is possible two structural motifs analogous to the specificity loop and intercalating hairpin exist in YonO. Bioinformatics has predicted YonO contains a zinc ribbon (Iyer and Aravind, 2012). Zinc ribbons are utilised by many proteins that interact with nucleic acids. The N-terminal domain of the  $\beta'$  subunit contains a single zinc ribbon, whilst the corresponding domain in eukaryotes and archaea contains two (Iyer et al., 2003). Based on its position in the structure of the RNAP holoenzyme, it has been hypothesised that this zinc ribbon participates in promoter binding (Murakami et al., 2002a, Vassylyev et al., 2002). Therefore, it is plausible to suggest the zinc ribbon of YonO may also contribute to promoter binding.

Another potential similarity between YonO and T7 RNAP is an inability to form stable open promoter complexes. Despite DNase I footprinting suggesting YonO binds 1673 DNA, albeit weakly, KMnO<sub>4</sub> footprinting suggested YonO does not form stable open promoter complexes. *In vitro* studies suggest that whilst DNA melting by T7 RNAP is rapid, formation of open promoter complexes is thermodynamically unfavourable (Bandwar and Patel, 2001, Villemain et al., 1997). Furthermore, the transition from a closed to a stable open promoter complex by T7 RNAP is stabilised by the presence of the NTPs complementary to the +1 and +2 positions on the template DNA (Stano et al., 2002). Speculating, YonO may be similar to T7 RNAP in transcription initiation as rapid

DNA melting and a requirement for the initiating NTPs would account for our inability to detect DNA melting by YonO. Alternatively, a previously unidentified transcription factor may contribute to stabilisation of open conformation of DNA melted by YonO. Only 25% of proteins encoded on SP $\beta$  have homology to known proteins (Lazarevic et al., 1999). Therefore, SP $\beta$  encoding its own specific transcription factor cannot be ruled out.

## Chapter 7: Attempts to Determine the Structure of YonO

### 7.1. Introduction

Structural studies of RNAP have contributed greatly to the field of transcription. Since 1999, a great number of structures have been published, providing answers to many important biological questions regarding transcription. The first high resolution structure of the core multisubunit RNA polymerase from *Thermus aquaticus* was published in 1999 by Zhang et al. (1999). Previously, cryoelectron microscopy had revealed that RNAP has an overall shape similar to that of a crab claw (Darst et al., 1989). The high resolution structure, elucidated by X-ray crystallography, confirmed this to be true (Cramer et al., 2001, Zhang et al., 1999). Furthermore, the structure confirmed and explained a vast amount of previous biochemical and genetic data. One such example is the elucidation of the location of the active centre. In the structure, this motif mapped to the centre of the enzyme and was shown to be chelating a  $Mg^{2+}$  ion, thus revealing the location of the active centre within the enzyme and confirming that the aspartate triad motif is responsible for chelating Mg I (Zhang et al., 1999, Zaychikov et al., 1996). Furthermore, Mustaev et al utilised  $Fe^{2+}$  hydroxyl cleavage to identify 9 distinct sites of  $\beta'$  and  $\beta$  which contact Mg I (Mustaev et al., 1997). Mapping of the 9 sites showed they spanned across the primary amino acid sequence of the two subunits (Zhang et al., 1999). The crystal structure of RNAP revealed that the 9 sites are within 20 Å of Mg I in the ternary structure thus confirming the previous biochemical data.

The explanation of previous biochemical data is a single example of how the structure of RNAP has advanced the field of transcription. The structures of RNAP have provided insights into all aspects of transcription, including initiation, elongation, substrate binding, and mechanism of catalysis (for a recent review, see Martinez-Rucobo and Cramer (2013) ). Therefore, since YonO is very distantly related to the multisubunit RNAP and lacks most of the essential domains, we sought to elucidate its structure at an atomic level.

### 7.2. Aims

The aim of this work was to determine the structure of YonO by X-ray crystallography. To this end, a collaboration was established with Dr Arnaud Baslé of the Structural

Biology Lab (SBL) at Newcastle University. Unless stated otherwise, David Forrest performed all cloning, protein expression, purification and crystallisation. Dr Baslé provided expertise on all aspects of crystallisation, harvested crystals and performed data collection at the Diamond Light Source synchrotron.

### **7.3. Screening for Crystallisation Conditions**

Highly purified, concentrated protein is required for crystallisation screens. The three step purification of YonO bearing an N-terminal 6x His-tag by Ni<sup>2+</sup> - NTA chromatography, Heparin affinity chromatography and gel filtration resulted in YonO purified to near homogeneity (discussed previously in Chapter 5).

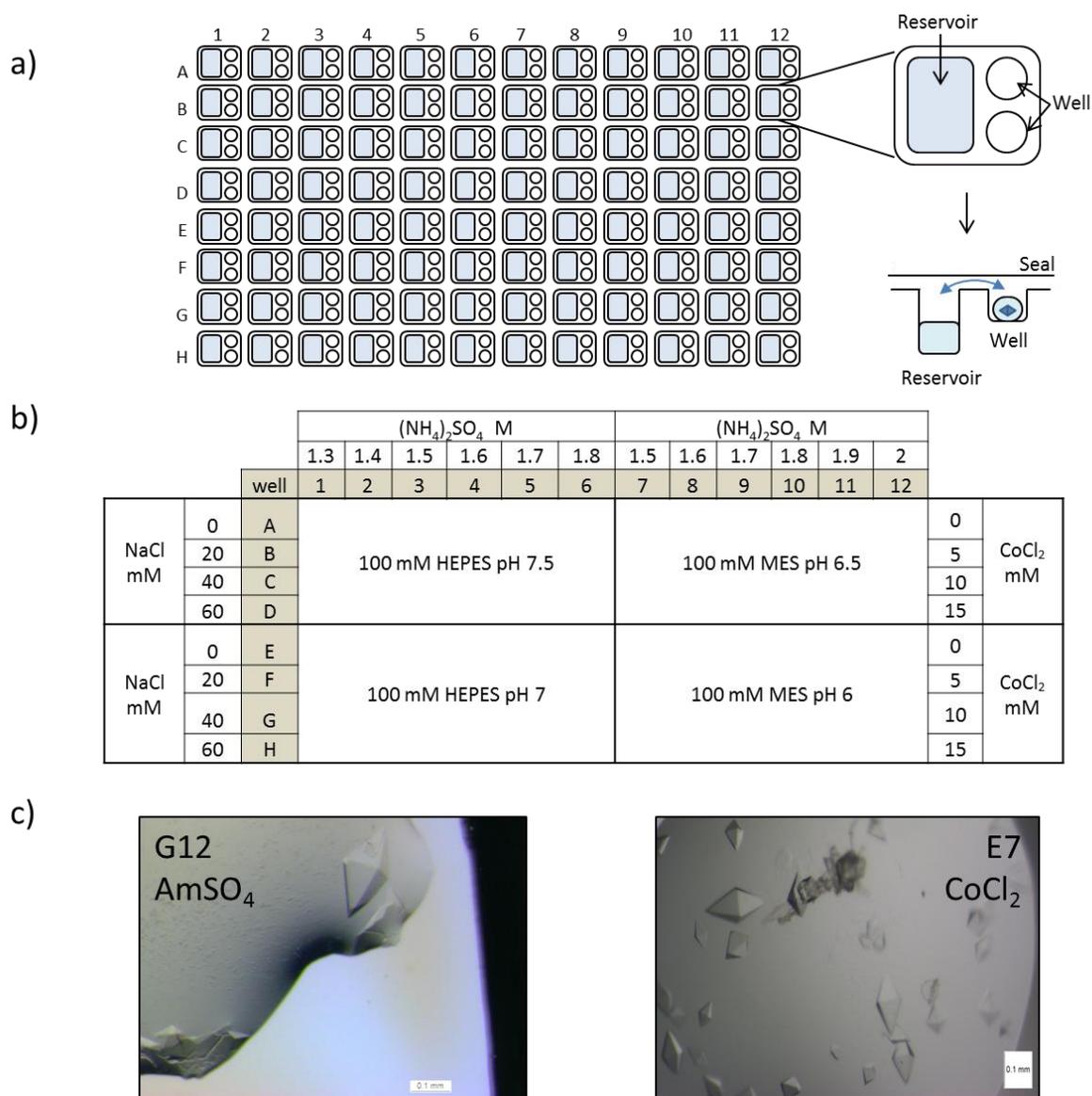
YonO was concentrated to 10 mg ml<sup>-1</sup> and trialled for crystallisation by equilibration against a standard set of commercial crystallisation screens. The standard set of screens included JSCG+, Proplex, MIDAS, Index, Morpheus, Structure screens. Details of these screens can be found in the material and methods. Each screen is comprised of 96 different crystallisation conditions, with each condition being dispensed into a reservoir of a MRC 2 well sitting drop 96 well plate (the layout of an MRC 2 well sitting drop plate is shown in Figure 7-1a). Adjacent to each reservoir on the MRC plate are two wells into which purified YonO was dispensed and mixed with the corresponding condition from the reservoir. Two wells allow variation in the ratio of protein and reservoir solution. In accordance with typical crystallisation screens, YonO was mixed with reservoir solution at a ratio of 1:1 and 2:1. The plates were sealed and equilibration was allowed to proceed overnight at 20 °C. Approximately 16 screening conditions were positive, giving bipyramidal crystals of varying sizes. To determine if the crystals diffracted to a sufficient resolution ( $\leq 3.5$  Å) for structure determination, crystals were harvested by Dr Baslé. Harvested crystals were cryoprotected in saturated (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and flash cooled in liquid nitrogen. Typically, crystals are cryopreserved at 100 K to reduce radiation damage during data collection. The crystals were tested for diffraction by loading onto the in-house X-ray generator. All crystals diffracted poorly, with the resolution limited to ~7-8 Å.

## 7.4. Optimisation of YonO Crystallisation

### 7.4.1. Optimisation of YonO screening condition

Optimisation of crystallisation conditions is routinely performed to improve the size and diffraction quality of crystals. To allow optimisation of YonO crystallisation conditions, the constituents of the 16 positive screening conditions were compared. It was determined that  $\text{CoCl}_2$ ,  $\text{NaCl}$  and  $(\text{NH}_4)_2\text{SO}_4$  encouraged the formation of crystals in pH ranging from 6.5 to 7.5.

Based on the initial crystallisation conditions, a custom optimisation screen ( $\text{CoCl}_2$  screen) was designed which varied the concentrations of the precipitant ( $(\text{NH}_4)_2\text{SO}_4$ ), salts ( $\text{NaCl}$  and  $\text{CoCl}_2$ ) and buffer pH (Figure 7-1b). The screen was mixed and dispensed into an MRC 2 well sitting drop 96 well plate by a Beckman BioMek NX2 automation workstation. After YonO was dispensed into the MRC plate and equilibrated overnight at 20 °C, screens were inspected manually. Whilst crystals formed in a range of conditions, a greater number of conditions in the lower right hand quadrant of the screen (conditions E6 – E12, F6 – F12, G6 – G12 and H6 – H12, Figure 7-1c) resulted in large bipyramidal crystals. This indicated that whilst the conditions for YonO crystallisation are flexible to a degree, pH 6 and the presence of  $\text{CoCl}_2$  promotes fewer nucleation events and therefore encourages the formation of larger crystals. In all positive screen conditions, the precipitant was  $(\text{NH}_4)_2\text{SO}_4$ . Therefore, the commercial  $\text{AmSO}_4$  screen was also trialled. As with the custom screen, overnight equilibration led to the formation of large bipyramidal crystals in a wide range of conditions.



**Figure 7-1: Optimisation of YonO crystallisation. a) MRC 2 well sitting drop 96 well plate.** Each position on the 96 well screen is comprised of a reservoir and two wells. Protein is dispensed into the wells and mixed with the reservoir solution. YonO was dispensed at a ratio of 1:1 (top well) and 2:1 (bottom well). The side view demonstrates vapour diffusion between the protein and reservoir. **b) CoCl<sub>2</sub> custom screen.** A custom screen was designed to optimise the conditions of YonO crystallisation based on the constituents of the reservoirs from the initial crystallisation. **c) Optimised YonO crystallisation.** YonO formed large bipyramidal crystals in many of the CoCl<sub>2</sub> and AmSO<sub>4</sub> screen conditions. Two examples are shown along with the corresponding crystallisation condition and screen.

Crystals from the CoCl<sub>2</sub> and AmSO<sub>4</sub> screens were harvested as previously described.

Crystals were transported to the Diamond Light Source synchrotron (DLS) for diffraction data collection. Synchrotron radiation is more intense than that produced by an in-house X-ray generator and can therefore provide higher resolution diffraction

data. Despite this, each of the crystals diffracted with a resolution of approximately 7-8 Å.

#### *7.4.2. Discovery and crystallisation of a possible YonO tetramer*

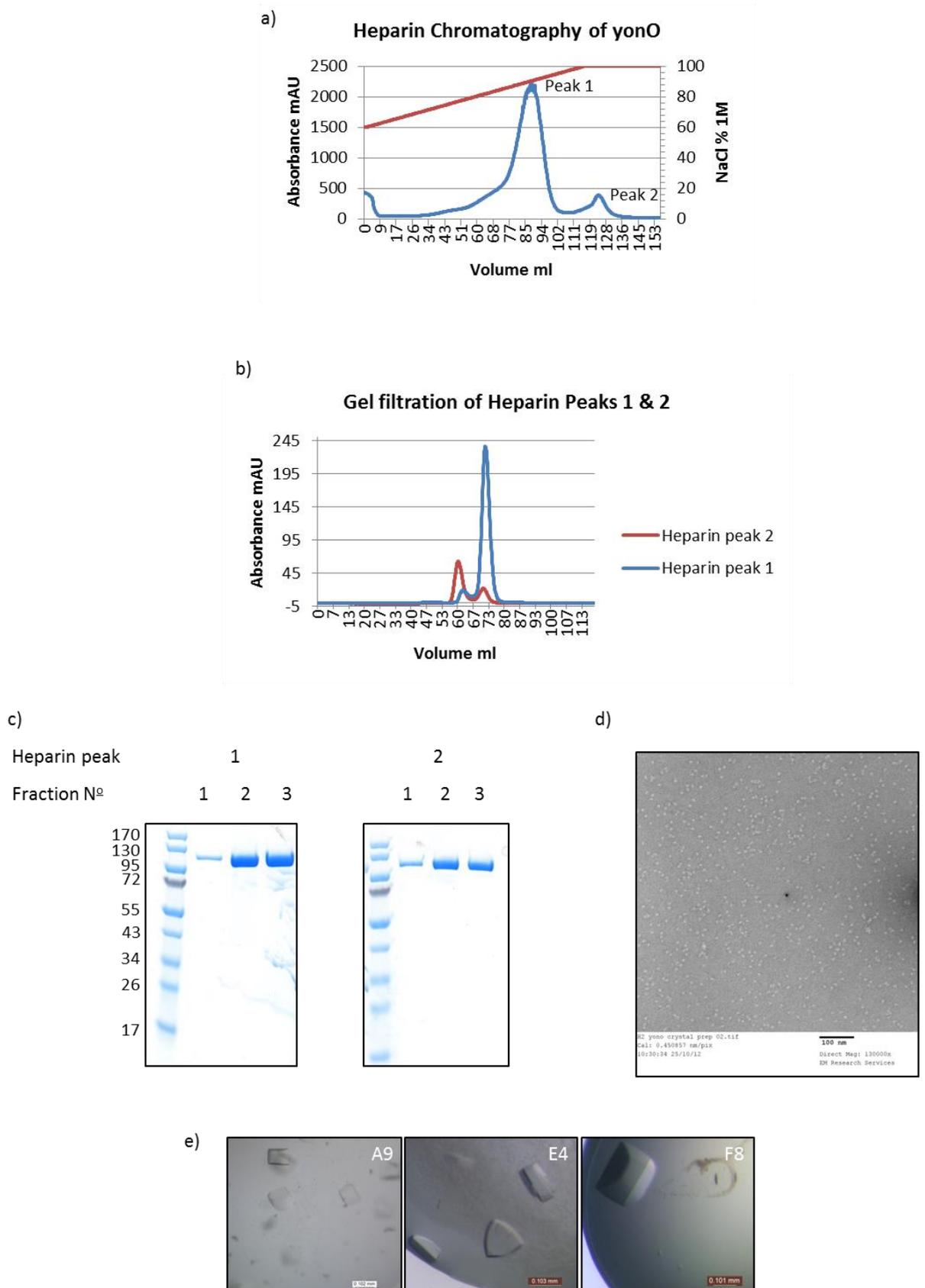
During Ni<sup>2+</sup>-NTA chromatography, YonO eluted from the column at two concentrations of imidazole (100 mM and 200 mM). The majority of YonO eluted at 100 mM imidazole and so this elution fraction was chosen for further purification by Heparin affinity and gel filtration. However, the YonO eluting at 200 mM imidazole during Ni<sup>2+</sup>-NTA chromatography was also further purified using heparin affinity chromatography. Interestingly, loaded YonO eluted off the Heparin Hi-trap column at two concentrations of NaCl. As before, YonO eluted at 800 mM NaCl. However, SDS PAGE revealed YonO also eluted off the column at 1M NaCl (Figure 7-2a). Based on the heparin chromatography chromatogram, the 800 mM and 1M NaCl elution fractions are referred to as peak 1 and peak 2, respectively. This presence of peak 2 was specific to the YonO eluted by 200 mM imidazole during Ni<sup>2+</sup>-NTA chromatography. Peak 2 suggests a minority of YonO exists in a conformation with increased affinity for heparin. Oligomerisation of YonO would explain an increased affinity. As gel filtration resolves proteins according to their molecular weight and shape, YonO samples from peaks 1 and 2 were independently loaded onto a Superdex 200 16/60 gel filtration column. As in previous gel filtration, YonO in peak 1 eluted off the column in an elution volume of 70 ml (Figure 7-2b). Peak 2 on the other hand, eluted off the column quicker with an elution volume of 60 ml. This showed YonO in peak 2 had an increased molecular mass. Since peak 2 only contained purified YonO (Figure 7-2c), the increase in molecular weight suggests YonO is capable of oligomerisation.

To investigate the prospect of oligomerisation further, the molecular weight of peak 2 YonO was calculated by gel filtration. To allow the calculation of molecular weight, the Superdex 200 column was calibrated using the elution volumes of protein standards with known molecular weights. The elution volumes of the standard proteins were used to calculate the K<sub>av</sub> constant, which represents the proportion of resin pores available to a given protein. The K<sub>av</sub> constants of the standard proteins were plotted against the log of their molecular weight to generate a calibration curve. From the calibration curve, it was calculated that YonO eluting off the gel filtration at 60 ml had a molecular weight of 413 KDa. YonO has a predicted molecular weight of 98 KDa (100

KDa with an N-terminal His tag), suggesting that YonO from peak 2 behaves as a tetramer.

The smaller form of YonO eluting at 70 ml (peak 1) had a calculated molecular weight of 135 KDa. The predicted molecular weight of YonO is 98 KDa. Therefore, the calculation of molecular weight suggests that the abundant form of YonO, the form used previously for crystallisation and *in vitro* studies, is monomeric. Negative stained transmission electron microscopy was used to confirm that YonO in peak 1 exists as a single, monomeric protein. At 130 000 x magnification, no aggregation or oligomers were observed (Figure 7-2d). The discrepancy between calculated and predicted molecular weights of monomer YonO can be explained by the physical shape of the YonO protein. The passage of a protein through a gel filtration column is influenced by its shape. Accurate calculation of a proteins molecular weight is dependent on the protein being globular, as the standard proteins used for the calibration are of a globular nature. Therefore YonO not being globular would lower the accuracy of the calculated molecular weight. Furthermore, the calibration curve is not linear and any small variations can reduce the accuracy of the calculated molecular weight. The N-terminal His tag only contributes approximately 2 KDa to the overall mass of YonO and so it is unlikely this influenced the estimation of the molecular mass of YonO and its oligomer. These limitations restricted us to only being able to estimate the molecular weight of YonO and its oligomerised state by gel filtration.

It was thought the YonO tetramer may form crystals with an alternative morphology to the bipryimadal crystal seen previously. One approach to obtaining higher resolution diffraction data is to alter the screening conditions to find those that allow for the formation of different crystal morphologies. In differing morphologies, the spatial organisation of the protein within the crystal is likely to be different and diffraction may be improved. Therefore, the tetramer was trialled for crystallisation by overnight equilibration against the standard set of screens, the custom CoCl<sub>2</sub> screen and the AmSO<sub>4</sub> screen. As with the YonO monomer, many conditions of the CoCl<sub>2</sub> screen allowed for the formation of crystals (Figure 7-2e). However, the tetramer crystals exhibited a range of morphologies. Crystals from the screens were harvested by our collaborator, Dr Arnaud Baslé, and were subjected to high energy synchrotron radiation. None of the tetramer crystals gave X-ray diffraction less than 7 Å.



**Figure 7-2: Purification and crystallisation of YonO tetramer.** a) **Heparin affinity chromatography of YonO.** YonO partially purified by Ni<sup>2+</sup>-NT chromatography was applied to a heparin column. Blue trace is absorbance of eluate at 280 nm. Red trace is NaCl concentration as a % of 1M. YonO was eluted from the column at ~80% (800 mM)

NaCl (peak 1) and at 100% (1M) NaCl (peak 2). **b) Gel filtration of YonO heparin peaks 1 and 2.** Heparin peaks 1 (blue trace) and 2 (red trace) were individually applied to a Superdex 200 column. Chromatography traces were overlaid. Peak 1 and 2 eluted at ~70 ml and ~60 ml, respectively. **c) SDS-PAGE comparison of heparin peaks 1 and 2.** **d) Electron microscopy of purified YonO.** Purified YonO was adsorbed on a carbon support and negatively stained using uranyl acetate. Protein was viewed at 130 000 x magnification. Microscopy was performed by Dr Robin Harris. **e) Crystals formed by YonO from heparin peak 2.** After gel filtration, YonO eluted in heparin peak 2 was concentrated and trialled for crystallisation. Crystals shown are from equilibration against the custom CoCl<sub>2</sub> screen. Specific screen condition are displayed in the top right.

## 7.5. Optimisation of YonO Crystal Diffraction

Many crystallography project fail due to an inability to find the correct conditions which promote protein crystallisation. In this case however, despite obtaining large ( $\geq 100 \mu\text{M}$ ) crystals, we were unable to obtain high resolution X-ray diffraction data. There are many techniques which have been developed to address this problem, a number of which were attempted to increase the resolution of diffraction obtained from YonO crystals.

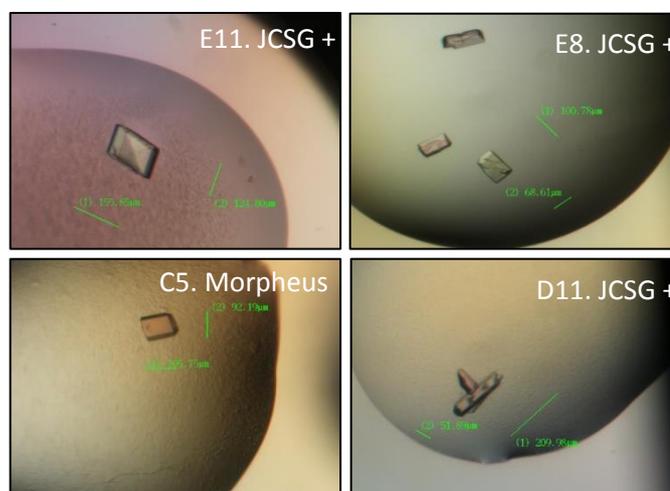
### 7.5.1. Microseeding to Produce Alternative Crystal Morphologies

The formation of a crystal occurs in two separate phases. The first step, nucleation, is the event in which, as the conditions become supersaturating (due to vapour diffusion against salt and a precipitant), proteins come out of solution by interacting to form a small well-ordered aggregate called a nuclei that has the potential to develop into a crystal. For the formation of crystal, additional protein is incorporated into the nuclei in a well-ordered manner. This growth is considered the second phase of crystal formation. The two phases of crystal growth are often encouraged by differing conditions. Nucleation requires a supersaturating state to bring about nucleation events. This is in contrast to growth, which requires a reduced saturating state to encourage protein to incorporate into a pre-existing crystal in a slow ordered manner, rather than forming new nuclei. Conditions which are too saturated result in unordered precipitation or unusable microcrystals.

Often the separation of nucleation and crystal growth phases can be beneficial. One condition may promote nucleation and not growth and vice versa. The technique of microseeding can address this (reviewed in Bergfors, 2003). Crystals can be grown in a previously trialled condition, harvested and physically broken. The broken pieces of

crystal (referred to as microseeds) can act as preformed nuclei in new crystallisation screens, onto which fresh protein can be incorporated to form new crystals (Bergfors, 2003, Ireton and Stoddard, 2004, D'arcy et al., 2007). The use of microseeds can bypass the nucleation stage and potentially result in crystals with different properties such as morphology and diffraction.

To this end, monomeric YonO was purified and screened against the CoCl<sub>2</sub> screen to produce large crystals. Crystals were crushed in the well of the 96 well plate, suspended in the corresponding reservoir solution and broken into fragments by vigorous vortexing with a seed bead. Crystallisation screens were set up as before except that microseeds were added to the wells of each screen in addition to freshly purified YonO. After overnight equilibration at 20 °C, crystals were observed in many of the conditions. Strikingly, crystals of many different morphologies formed (Figure 7-3). These were harvested and tested for diffraction using the in house x-ray generator. Salt crystals produce a very characteristic diffraction pattern; the diffraction resolution is high and the diffraction spots are far apart. All of the microseed crystals harvested gave diffraction patterns characteristic of salt.

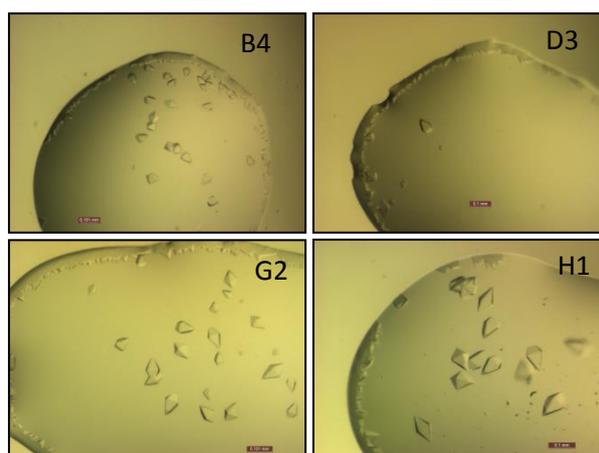


**Figure 7-3: Microseeding of crystallisation screens.** Microseeds were added to the wells of crystallisation screens in addition to fresh YonO protein and reservoir solution. Equilibration was performed overnight at 20 °C. Above, four examples of resultant crystals are shown. The condition and screen name are shown along with the scale bar.

### 7.5.2. Additive Screening

Crystallisation can be aided by the presence of an additive in the reservoir solution. The term additive refers to any compound added which may alter the crystallisation process of a given protein, with examples being  $\text{ZnCl}_2$  (salts), glucose (carbohydrates) and ethanol (organic compounds). An additive may alter crystallisation by stabilising the protein, reducing movement of mobile domains or by influencing protein-protein interactions. It is hoped that these alterations increase the order within the crystal which would increase the resolution of the collected diffraction data. An example of this is the addition of spermidine to the reservoir solution during crystallisation of the distant YonO homologue QDE-1 (Laurila et al., 2005).

For additive screening, a single condition known to produce crystals of YonO was chosen. This condition, G9 of the custom  $\text{CoCl}_2$  screen (see Figure 7-1b for  $\text{CoCl}_2$  composition) was dispensed into every reservoir of a MRC 2 well 96 well plate. Each reservoir was supplemented with an additive from the 96 well additive screen. The screen was then set up as previously described and equilibrated overnight. Many crystals formed across the screen (Figure 7-4). However, all crystals observed were bipyramidal and the additives resulted in a decrease of crystal size and an increase in nucleation events. The largest crystals were harvested (Figure 7-4) and tested at synchrotron for diffraction. None of the additives had a positive effect on diffraction resolution, with diffraction being limited to 7 – 8 Å.



**Figure 7-4: Additive screening for improved crystal diffraction.** Condition G9 of the custom  $\text{CoCl}_2$  screen (100 mM MES pH 6, 10 mM  $\text{CoCl}_2$  and 1.7 M  $(\text{NH}_4)_2\text{SO}_4$ ) was used as the reservoir solution in a 96 well screen. To each reservoir, an additive was added prior to overnight equilibration with YonO. Small bipyramidal crystals formed in

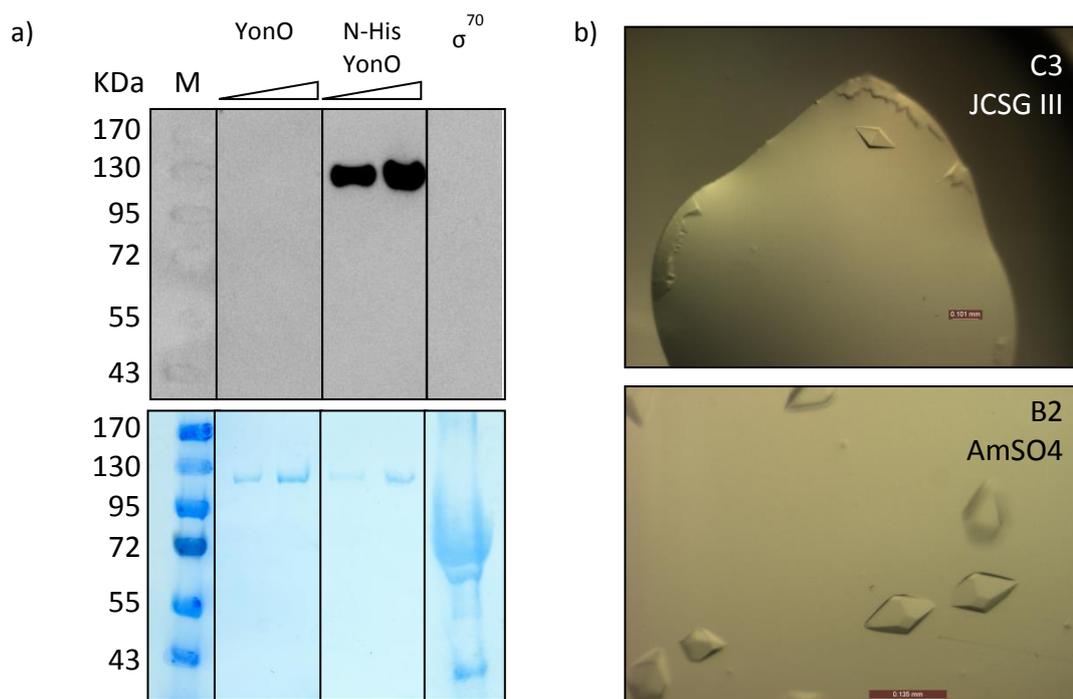
many conditions. Shown in this Figure are the largest crystals formed in the presence of an additive in conditions B4 (G9 + 100mM KCl), D3 (G9 + 10 mM spermine tetrahydrochloride), G2 (G9 + 3% hexanediol) and H11(G9 + 4% 2,2,2-trifluoroethanol).

### *7.5.3. Removal of N-terminal His tag*

Out of the purification tags available, the hexahistidine tag (His-tag) is most frequently used for protein purification. In particular, it is favoured for the purification of proteins destined for crystallisation screens. This is due to its small size, 1 KDa, which rarely impedes on the structure of a protein or its crystallisation (Smyth et al., 2003).

Nevertheless, it is accepted that the presence of a His-tag can occasionally be detrimental to crystallisation as their flexibility allows them to interfere with protein – protein interactions (Dale et al., 2003). Therefore, there is a possibility that the N-terminal His-tag present on YonO prevented the ordered crystal packing required to form diffracting crystals.

To address this, YonO was cloned into the expression vector pET-21, with a TAA stop codon preventing the inclusion of a C-terminal His-tag from the vector. Despite the lack of a tag, YonO was purified to apparent homogeneity through the use of a heparin column, for which YonO has a very high affinity, with approximately 800 mM NaCl being required for elution of YonO from heparin. Western blot with anti-His tag primary antibodies was used to confirm the purified YonO was tag-free (Figure 7-5a)

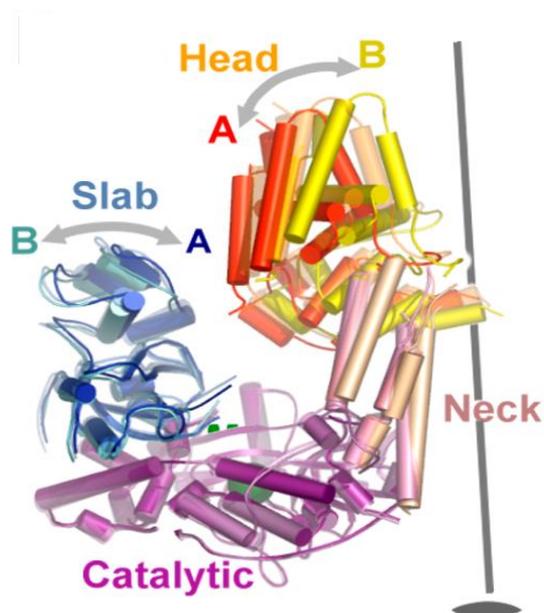


**Figure 7-5: Crystallisation of tag free YonO. a) Western blot to confirm absence of His-tag on YonO.** Tag free YonO was expressed and purified by heparin affinity chromatography and gel filtration. 5 and 10 pmol of purified YonO, either tag free or with an N-His tag were resolved by SDS – PAGE and transferred onto PVDF membrane. Membrane was probed using anti-His-tag antibodies and visualised using HRP-conjugated anti-mouse antibodies. **b) Tag free YonO forms bipyrarnidal crystals.** Purified tag free YonO was trialled for crystallisation by overnight equilibration with the standard set of crystallisation screens. Drop sizes of 100 nl was used with ratios of 1:1 and 2:1 (protein: reservoir). Shown are two representative conditions (displayed in top right) which gave bipyrarnidal crystals.

Tag-free YonO was trialled for crystallisation by equilibration against the standard set of screens (see section 7.3.1), the AmSO<sub>4</sub> screen and CoCl<sub>2</sub> screen. As before, bipyrarnidal crystals formed after overnight equilibration (Figure 7-5). The diffraction of the crystals was limited to 7 Å. Despite not increasing the diffraction, tag-free YonO crystals did allow us to conclude that the His-tag is not responsible for the poor diffraction observed throughout the crystallisation trials. Tag free YonO was purified to near 100% purity without Ni<sup>2+</sup>-NTA chromatography. Additionally, the absence of the His-tag did not have a detrimental effect on the diffraction quality of crystals. Therefore, all subsequent crystallisation trials were conducted using tag free YonO.

#### 7.5.4. Use of Artificial Transcription Scaffolds to Stabilise YonO RNAP

Despite optimisation, microseeding and additive screening, the conditions permissive for the growth of crystals with high resolution diffraction remained elusive. A possible explanation for this is intrinsic flexibility of YonO during crystallisation. RNAPs contain mobile domains. In the multisubunit RNAP, the clamp domain swings down to grip the downstream DNA (Chakraborty et al., 2012, Cramer et al., 2000, Cramer et al., 2001). The notion that YonO contains similarly mobile domains is plausible, as it was observed that QDE-1, a distant homologue of YonO, is able to adopt open and closed conformations through movement of the head and slab domains (Salgado et al., 2006)(Figure 7-6).



**Figure 7-6: Open and closed conformation of QDE-1. Taken from Salgado et al, 2006.**

The side view of a QDE-1 monomer is shown. The slab and head domains of two monomers belonging to the same homodimer were seen to be positioned at different angles. This is represented by double headed arrows between monomer A and monomer B. In an independent, second dimer structure, the head and slab domains are positioned at the same angle. Taken together this suggests the head and slab domains have a degree of flexibility which may correspond to the opening and closing of QDE-1.

Whilst the mobility of RNAP domains was not problematic for the growth of well ordered, high quality crystals previously, it cannot be ruled out that potential intrinsic

flexibility of YonO is detrimental to crystal formation. To address this potential reason for poor crystal diffraction, we sought to crystallise YonO bound to a nucleic acid scaffold. It was hoped that upon binding to the scaffold, YonO would adopt a stable, rigid conformation which would allow for increased order within the crystal.

For the determination of the *Thermus thermophilus* elongation complex structure, RNAP was assembled with a nucleic acid scaffold prior to crystallisation (Vassilyev et al., 2007a). This scaffold (denoted as scaffold 1) was obtained and used for the assembly of YonO elongation complexes (YonO EC) (scaffold 1, Figure 7-7). To confirm that YonO ECs were both stable and active, assembled complexes were immobilised by the 5' biotin tagged template DNA strand to streptavidin beads and washed with transcription buffer containing low or high concentrations of salt. Only stable elongation complexes are resistant to high ionic strength (Sidorenkov et al., 1998). Upon washing the immobilised complexes with high salt transcription buffer, non-template DNA (NTDNA14) and RNA (RNA16) was lost, showing the YonO EC were unstable (Figure 7-7, compare lane 1 against lane 2 and lane 6 against lane 7). The assembled complexes were not destabilised by low salt concentrations, as RNA16 and NTDNA14 were not removed after washing immobilised complexes with 40 mM KCl. Furthermore, complexes extended RNA16 to the end of the template in the presence of NTPs (Figure 7-7, compare lane 4 against lane 5 and lane 9 against 10,). However, not all YonO elongation complexes extended RNA16. Scaffold 1 was originally designed for form stable *T. thermophilus* elongation complexes that do not oscillate between the post and pre-translocated states, which would interfere with crystallisation. Therefore, the inability of all YonO complexes to extend RNA16 in the presence of NTPs can be explained by increased stability of complexes due to the sequence of scaffold 1 oligonucleotides (Vassilyev et al., 2007a).

Scaffold 1 was comprised of the minimum nucleic acid strands required to form a stable elongation complex. The short length of the scaffold was designed to reduce the likelihood that either RNA or DNA extrudes from the surface of RNAP and interferes with crystal contacts during crystallisation (Kashkina et al., 2006). Whilst bacterial RNAP is able to form stable elongation complexes on this scaffold, YonO did not. The non-template DNA strand is known to have a stabilising effect on elongation complexes (Sidorenkov et al., 1998). In scaffold 1, the non-template DNA strand

(NTDNA14) was short, with the strand only annealing to the downstream region of the template DNA strand (as seen in Figure 7-7). It was reasoned that a longer non-template strand may increase the stability of YonO elongation complexes (referred to as scaffold 2 which is shown in Figure 7-7). The non-template strand was extended at the 5' end by 9 nucleotides (NTDNA23). The extended sequence was non-complementary to the template strand to mimic the transcription bubble (Yin and Steitz, 2002). Scaffold 2 YonO ECs were assembled, immobilised and washed with transcription buffer containing high or low concentrations of salt. The stability of the complexes was improved by the presence of the longer non-template DNA strand as both RNA13 and NTDNA23 were present after washing scaffold 2 complexes with high salt buffer (Figure 7-7, compare lanes 11 and 17 to lanes 1 and 7,). The increased stability of scaffold 2 EC is further demonstrated by comparisons of RNA13 extension by scaffold 1 and 2 complexes radiolabelled at the 5' end of the non-template strand (Figure 7-7, compare lane 18 to 8,). In these reactions, RNA13 extended by YonO was labelled through the incorporation of  $\alpha$ -[ $^{32}\text{P}$ ]-AMP during elongation. Scaffold 2 resulted in increased amounts of extended RNA13 which suggests more RNA13 was retained in assembled elongation complexes after washing with high salt buffer.

One complication with the crystallisation of YonO is the requirement of 500 mM NaCl to maintain solubility of YonO at a concentration of 10 mg ml<sup>-1</sup> (not discussed or shown). When determining the stability of scaffold 1 and 2 elongation complexes, complexes were formed in the presence of 40 mM KCl prior to being washed by 500 mM NaCl transcription buffer. This differs to the formation of elongation complexes intended for crystallisation screens, which would be formed in the presence of the 500 mM NaCl necessary for YonO solubility. Elongation complexes are only resistant to high ionic strength after assembly, making it possible that elongation complexes may not form in 500 mM NaCl prior to crystallisation trials (Sidorenkov et al., 1998). Nevertheless, since scaffold 2 complexes showed increased stability at high salt concentrations, scaffold 2 was incubated with YonO prior to overnight equilibration against the standard set of crystallisation screens (see materials and methods). The screens resulted in bipyramidal crystals, all of which, when harvested and subjected to synchrotron radiation, diffracted to 7 – 8 Å. It is likely that the salt concentration prevented elongation complex formation as predicted and that the crystals did not

include the nucleic acid scaffold. YonO appears to not be like other RNAPs, which require a much reduced concentration of salt to remain soluble at  $10 \text{ mg ml}^{-1}$  allowing EC formation at the high protein concentrations required for crystallisation (Gnatt et al., 2001, Kashkina et al., 2006, Vassylyev et al., 2007a).

To attempt to reduce the salt concentration required for YonO solubility,  $10 \text{ mg ml}^{-1}$  YonO in  $500 \text{ mM NaCl}$  buffer was incubated with scaffold 2 before undergoing serial dialysis against buffers containing decreasing concentrations of salt. It was hoped the presence of a nucleic acid scaffold would alter the solubility of YonO and allow high protein concentrations in reduced salt concentrations. This was not the case, as with each dialysis against a lower salt concentration, precipitation occurred.

For the crystallisation of RNAP II elongation complexes, RNAP II was allowed to form elongation complexes, either through incubation with a scaffold or by transcribing on duplex DNA with a 3' single stranded tail. RNAP II started transcription on the tail which bypassed the need for initiation factors. The sequence of the downstream DNA permitted synthesis of a 14 nucleotide RNA without incorporation of UMP. When provided all NTPs except UTP, all active RNAP II complexes paused upon reaching the first Adenosine in the template DNA ensuring all RNAP II complexes were homogenous. (Kettenberger et al., 2004, Gnatt et al., 2001). In both methods, elongation complexes were purified and concentrated prior to crystallisation screens. Whilst not attempted due to time constraints, one of these approaches may allow for YonO elongation complexes to be maintained in high salt conditions. After initial assembly in low salt and low protein concentrations, elongation complexes would be purified by reverse heparin chromatography or gel filtration. Subsequently YonO elongation complexes would be concentrated to  $10 \text{ mg ml}^{-1}$  and trialled for crystallisation.



#### *7.5.5. Lysine methylation*

In addition to flexible domains, the surface entropy of a protein also influences its ability to assemble into well-ordered crystals. Proteins with increased surface entropy display greater variation on the protein surface. This added variation adds disorder to the crystal as the proteins ability to form well-ordered, repeating assemblies is reduced. Surface lysines have high entropy and are known to reduce the available contacts between proteins during crystal formation (Derewenda and Vekilov, 2006). Analysis of crystallisation data confirmed this by showing lysine as the least favoured amino acid for crystal contacts (Dasgupta et al., 1997). If surface lysines are known, point mutation can be carried out to remove these residues and improve crystallisation (Lawson et al., 1991, Derewenda and Vekilov, 2006, Walter et al., 2006). However, if a protein is undergoing crystallisation trials for the first time, it is unlikely surface lysines have been identified. Methylation of surface lysines offers an alternative way of reducing entropy and improving crystallisation and diffraction (Walter et al., 2006). In addition to reducing surface entropy, methylation increases the available crystal contacts by introducing the possibility of hydrophobic interactions between proteins (Walter et al., 2006).

Methylation was performed to see if reducing surface entropy increased diffraction of YonO crystals. Purified YonO was incubated with formaldehyde and dimethylamine-borane overnight and subsequently purified by gel filtration chromatography. Methylated YonO did not elute from the Superdex 200 column, indicating that YonO precipitated during gel filtration. Another outcome of methylating lysines is a reduction of solubility through the removal of the polar side chain and introducing hydrophobicity. Often, this can be beneficial as it can increase crystallisation of highly soluble proteins. Inspection of the amino acid composition of YonO revealed lysine to be the most abundant. It is unknown how many of these lysines are located on the surface of the protein. However, it has been suggested lysines are predominantly located to the protein surface (Derewenda and Vekilov, 2006) thus explaining the drastic loss of YonO solubility.

#### *7.5.6. Twinned Date Set and Molecular Replacement*

Typically, crystals are harvested and suspended in a cryoprotectant such as saturated  $(\text{NH}_4)_2\text{SO}_4$  before being flash cooled in liquid nitrogen. Flash cooling in liquid nitrogen

has multiple functions: to preserve the arrangement of solvent within the crystal, to prevent the formation of ice and to protect from X-ray radiation damage. However, cases where the act of flash cooling has caused a decrease in order and in diffraction resolution due to differential freezing of the protein and solvent have been reported (Heras and Martin, 2005, Juers and Matthews, 2001, Kriminski et al., 2002).

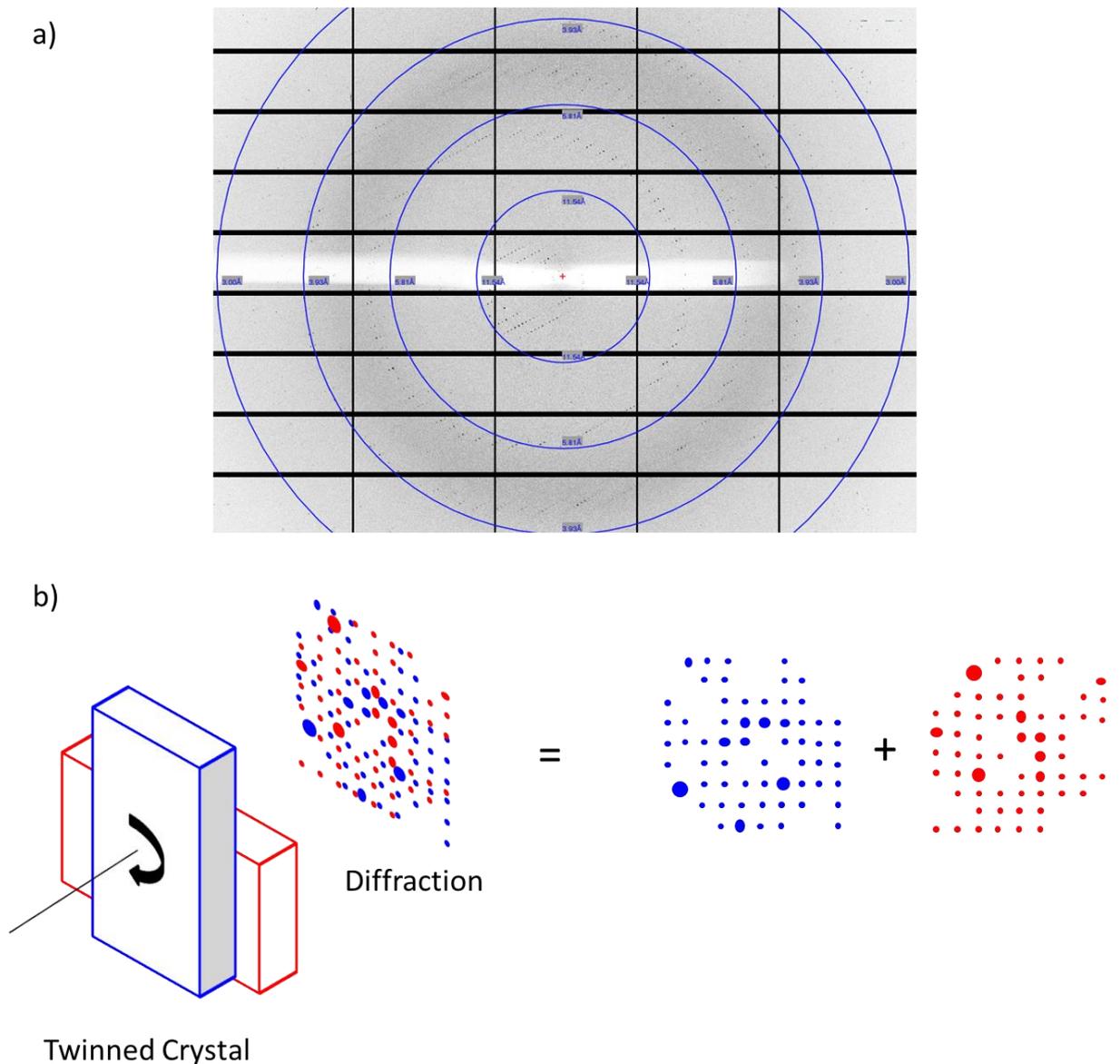
To see if crystal disorder introduced by flash cooling is responsible for the poor diffraction of YonO crystals, diffraction was attempted at room temperature (20 °C). A crystal, grown from tag free YonO equilibrated with condition C2 of JCSG II screen, was harvested and suspended in paratone-N oil. Previously, crystals were suspended in saturated (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. However, at room temperature (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> salt crystals are likely to form leading us to use paratone-N oil. Immediately after harvesting, the crystal was tested for diffraction using the in house X-ray generator. The cryo nozzle was removed to allow diffraction data to be collected at room temperature. Room temperature data collection did not lead to increased diffraction, with the crystal diffracting to 8 Å. Despite this, the crystal was stored in liquid nitrogen to be tested for diffraction using synchrotron radiation at Diamond Light source at a later date, as the increased energy of synchrotron radiation can improve diffraction. When exposed to synchrotron radiation on the IO4-I beamline, the crystal diffracted to 3 Å (Figure 7-8a). From this crystal, two complete data sets were collected, data set 1 and 2. Data set 1 and 2 had resolutions of 3 Å and 3.5 Å, respectively. The dataset statistics can be seen in Table 7-

Dataset	<u>1</u>			<u>2</u>		
Space group	P <sub>4</sub> 1 2 2			C 2 2 2 <sub>1</sub>		
Twin	44%			20%		
Data Collection	Overall	Inner Shell	Outer Shell	Overall	Inner Shell	Outer Shell
Resolution Range Å	2.95 -	8.85 -	2.95 - 3.13	3.5 -	9.90 -	
R <sub>merge</sub>	82.60	82.60		88.23	88.23	3.50 - 3.74
Total No observations	0.058	0.022	0.569	0.142	0.061	0.636
Total unique observations	122366	4312	19829	159447	6913	28302
Mean(I)/Standard Deviation (I)	20789	904	3274	24132	1169	4273
Completeness	15.9	32.4	2.7	10.8	17.3	5.2
Multiplicity	100	99.9	100	100	100	100
	5.9	4.8	6.1	6.6	5.9	6.6
Unit cell parameters						
a, b, c Å	87.96	87.96	240.4	124.8	124.8	241
α, β, γ °	90	90	90	90	90	90

1.

**Table 7-1: Data statistics for YonO Data sets 1 and 2.**

Crystals are ordered repetitions of protein in three dimension space. The most basic building block of the crystal is the asymmetric unit. In the simplest example, the asymmetric unit contains one whole protein molecule. Often the asymmetric unit can contain multiple protein molecules, or a fraction of a protein molecule. The asymmetric unit is reflected, repeated and rotated in space to generate the unit cell. Each dataset has a space group. The space group describes the rotations and the symmetry operators required to generate the unit cell from the asymmetric unit. The unit cell is repeated in three dimensional space to build up the crystal. Data sets 1 and 2 were given space groups  $P4_1 2 2$  and  $C2 2 2_1$ , respectively. The Matthews coefficient predicts the number of protein molecules in each asymmetric unit by calculating the protein content (Matthews, 1968). The coefficient predicted two YonO molecules in each asymmetric unit. The presence of multiple protein molecules in an asymmetric unit does not necessarily imply biological dimerization and can be a result of the crystal contacts between protein molecules.



**Figure 7-8: High resolution diffraction from a twinned YonO crystal. a) Diffraction pattern of YonO data set 2.** An example of diffraction from the YonO bipyramidal JCSG II C2. The diffraction image is taken from data set 2. The blue rings demarcate the diffraction resolutions. **b) Crystallographic Twinning.** Adapted from Rupp, 2010. Twinning involves the formation of different, overlapping domains in the same crystal. In this simple example, two domains have formed (red and blue) related with a  $90^\circ$  rotation. Upon X-ray diffraction, two overlapping diffraction patterns are generated. For specific types of twinned data, depending on the operating space group, it is possible to resolve the two diffraction patterns using a mathematical twin operator. However, for the YonO data set 1 and 2, with space groups  $P_4 1 22$  and  $C2 2 2_1$ , respectively, it was not possible.

Upon analysis of data sets 1 and 2, it was revealed that both datasets were twinned. In a twinned crystal, the protein has arranged into two different overlapping lattices (Figure 7-8b). Upon X-ray diffraction, the two different crystal organisations results in two overlapping diffraction patterns. Two overlapping diffraction patterns can prevent

reliable determination of the space group and result in complications in downstream data processing. Dataset 1 had a twin value of 44%, meaning approximately half of the crystal was ordered differently. This is classed as a perfect twin. Data set 2 displayed less twinning, having a twin value of 22%. For certain space groups it is possible to deconvolve the two overlapping diffractions using a twin operator, which describes the relationship between the two overlapping crystal patterns. However, for the specific space groups formed by YonO, no such twin law exists.

Despite twinning reducing the accuracy of the data set, molecular replacement was attempted using data set 2 to solve the structure of YonO. Briefly, molecular replacement involves using data from a previously solved structure of a homologous protein to help determine the structure of a second protein. The data of the known structure is rotated and translated in three dimensional space to find a fit to the data of the protein under investigation. The nearest homologue with a known structure is QDE-1. Molecular replacement using the program Phaser with the QDE-1 structure did not result in a fit (Mccoy et al., 2007). A less stringent molecular replacement approach is to use individual structural domains of QDE-1 to find a fit as this would take into account any differences in domain orientation and positioning between YonO and QDE-1. QDE-1 was divided into three domains: head and neck, catalytic, and slab domains. Molecular replacement was then repeated with each domain, but again, no solutions were found. Molecular replacement requires amino acid sequence identity of 30 % or above. The homology between YonO and QDE-1 is poor. Two iterations of pattern based PSI-BLAST are required to identify homology between YonO and QDE-1. This substantial lack of sequence homology explains the inability of molecular replacement to solve the structure of YonO.

#### **7.6. Dehydration of YonO crystals to Reproduce High Resolution Data Set**

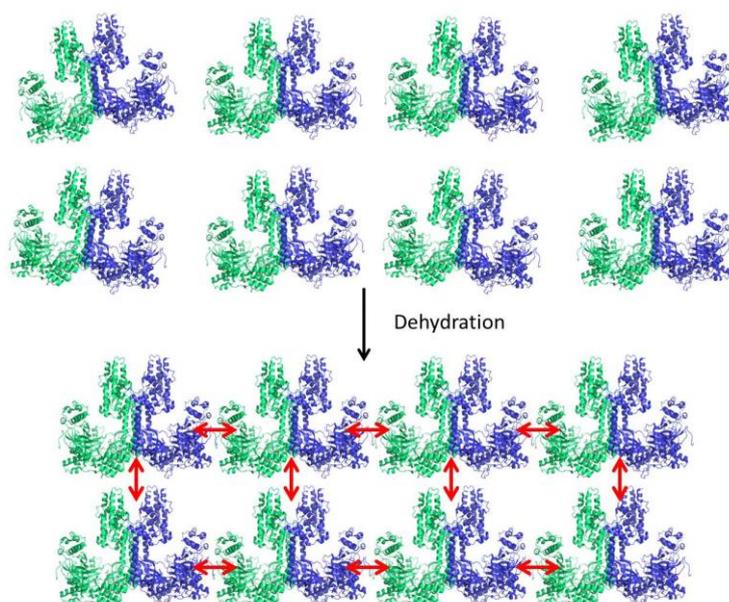
Twinning is thought to be stochastic; therefore we sought to reproduce the conditions which gave a crystal diffracting to 3 Å. A strategy for the consistent, reproduction of diffracting crystals would be a significant step forward in determining the structure of YonO. Data sets 1 and 2 were obtained from a crystal formed through equilibration against C2 condition of the JCSG II screen. The crystal was preserved in paratone-N oil and tested for diffraction on the in house x-ray generator at room temperature before being flash cooled and transported to synchrotron. It was suggested that exposure to

air prior to cryoprotecting in paratone-N oil resulted in the dehydration of the crystal. Dehydration, the removal of solvent from within a crystal is a well employed method for increasing the diffraction quality of crystals (Heras and Martin, 2005)

#### *7.6.1. Dehydration by HCl Dehumidifier*

The optimisation techniques attempted previously focus on altering the crystallisation of YonO to improve diffraction. A second, alternative approach is to improve diffraction through post-crystallisation techniques (Heras and Martin, 2005). One of the most investigated post-crystallisation technique is crystal dehydration.

Approximately 50% of a protein crystal is solvent; therefore solvent plays an important role in the ordering of protein. If a crystal displays poor diffraction, it is possible to reduce the solvent content by controlled exposure to air or vapour diffusion against increased concentrations of a desiccant such as a salt. A potential outcome of reducing the solvent content is an increase in crystal packing and order in the crystal (shown schematically in Figure 7-9). This in turn can have a positive impact on the quality of diffraction. To reproduce the crystallisation process, crystals were harvested and suspended in paratone-N oil before being exposed to the air for a defined length of time. The time of air exposure was not known and so it was varied from 30 – 165 minutes, with a crystal being flash cooled every 15 minutes to stop dehydration. The crystals were tested for diffraction at DLS but as before, all crystals only diffracted to  $\sim 7\text{\AA}$ .



**Figure 7-9: Effect of dehydration on crystal packing. Adapted from (Awad et al., 2013).** Shown is a crystal simplified to two dimensions comprising of a protein dimer. Dimer subunits are coloured green and blue. For this example, the protein dimer is QDE-1 (accession: 2J7N (Salgado et al., 2006)). In a hydrated crystal, solvent exists (white space in-between dimers) in between repeating protein units. Dehydration of crystals removes this solvent, increasing the crystal packing and order within the crystal (shown by red arrows).

Accurate and reproducible dehydration can be achieved through the use of the HC1 humidity control instrument at Diamond Light Source synchrotron (Sanchez-Weatherby et al., 2009). HC1 features a nozzle which can be mounted into the synchrotron beamline apparatus. Once mounted, HC1 projects an airstream over the crystal. The relative humidity of the airstream can be controlled, allowing reproducible dehydration. HC1 was installed onto the IO4-I beamline in place of the cryostream. Prior to dehydrating a crystal, the HC1 was equilibrated to find the relative humidity of the reservoir solution used in crystallisation. At the relative humidity of the reservoir solution, the hydration state of the reservoir solution drop remained constant. This relative humidity value was then used as the starting point for dehydration experiments. The crystal which gave high resolution diffraction grew when YonO was equilibrated against condition C2 of the JCSG II screen. Since the initial diffraction from this crystal was poor, it was reasoned the specific composition of this condition was not responsible for the increased diffraction. Therefore, whilst crystals grown against C2 JCSG II were used for HC1 experiments, crystals formed using the  $\text{CoCl}_2$  and  $\text{AmSO}_4$

screens were also dehydrated. Overnight equilibration against the  $\text{CoCl}_2$  and  $\text{AmSO}_4$  screens consistently produced large  $\geq 100 \mu\text{m}$  bipyramidal YonO crystals.

Initially, crystals from the  $\text{CoCl}_2$  screen were dehydrated. A crystal from A10 condition of the  $\text{CoCl}_2$  screen was mounted into the beamline and the diffraction was tested at 96 %, the equilibrated relative humidity of the reservoir solution. The crystal gave diffraction in the region of 7 – 8 Å. The HC1 control allows the relative humidity to be lowered with a defined gradient. Initially, the humidity was lowered in large increments with steep gradients. The relative humidity was lowered to 86 % with a gradient of  $2 \text{ \% min}^{-1}$ . The diffraction of the crystal remained constant at 7 – 8 Å showing the dehydration did not improve diffraction. Further reductions of humidity did not increase diffraction. At 45 % relative humidity the diffraction was completely lost, with crystals being completely dehydrated.

The humidity at which the diffraction can be improved may have a narrow range that was missed by lowering the relative humidity in large increments. Therefore, crystals from the C2 JCSG II condition were dehydrated with small reductions in humidity. The reduction of humidity from 92.5 % to 90 % did not increase the diffraction of YonO crystals.

An additional variable is the length of time a crystal is exposed to dehydrating conditions. Previously, diffraction was tested upon reaching the desired relative humidity. To determine if prolonged exposure to a given relative humidity can increase diffraction, crystals were exposed to a constant relative humidity before being tested for diffraction. Crystals from the A8 condition of the  $\text{AmSO}_4$  plate were tested for diffraction at 93 % relative humidity. The crystal diffracted to 7 – 8 Å. However, after 5 minutes at a constant relative humidity of 93 %, the diffraction was completely lost indicating crystals deteriorated over time.

For HC1 dehydration, crystals are not suspended in cryoprotectant as data collection was carried out at room temperature. However, the YonO crystal that previously diffracted to 3 Å was preserved in paratone-N oil. Therefore, diffraction was tested on crystals from the B4 condition of the  $\text{AmSO}_4$  screen suspended in paratone-N oil. Resuspending the crystal in paratone-N oil prior to testing the X-ray diffraction increased the life span of the crystals, as 7- 8 Å diffraction was not lost over time. The

crystal from which 3 Å data was collected was stored in liquid nitrogen prior to data collection at synchrotron. Therefore the B4 AmSO<sub>4</sub> crystals were removed from the beamline, stored in liquid nitrogen and retested for diffraction when the HC1 was removed from the synchrotron and replaced with the standard cryostream. Despite replicating the conditions which resulted in 3 Å data collection, the above efforts did not result in the collection of high resolution data from B4 AmSO<sub>4</sub> crystals.

#### 7.6.2. Dehydration through vapour diffusion and *In situ* data collection

Controlled dehydration of crystals can also be achieved through vapour diffusion against a reservoir solution with an increased concentration of precipitant such as NaCl or polyethylene glycol (Schick and Jurnak, 1994). The increased precipitant acts as a desiccant, withdrawing the water from the crystal. Alternatively, new salts previously not present in the reservoir solution can be added to function as desiccants. The dehydration is reproducible as the concentration of the desiccants can be controlled along with the time and temperature of dehydration.

Dehydration by the HC1 was time consuming as each crystal was mounted manually. *In situ* plate screening allows for high throughput testing of crystals at the synchrotron. *In situ* plate screening involves the direct loading of the 96 well crystallisation plate onto the synchrotron beamline. Multiple crystals in each condition can be rapidly tested for high quality diffraction without harvesting or cryoprotection, both of which can be damaging to the crystal. Recently, a 96 well plate based dehydration screen combined with *in situ* plate screening was described (Douangamath et al., 2013). The screen allowed for the simultaneous testing of 96 different salt concentrations to determine the optimum concentration for dehydration. Briefly, a 96 well plate screen was set up with only one reservoir solution known to allow for the formation of crystals. Once crystals had formed, a dehydration screen was added to the reservoir solution. The diffraction properties of the dehydrated crystals were then determined. The equipment required for the *in situ* dehydration screening is available at Diamond Light Source. Therefore, our collaborator (Dr Arnaud Baslé) performed the screen using purified tag-free YonO equilibrated against condition G11 from the AmSO<sub>4</sub> screen. Analysis of previous crystallisation screens identified this condition as consistently producing large ( $\geq 100$  μM) bipyramidal crystals. The dehydration screen was comprised of four desiccants: ethylene glycol, polyethylene glycol 200 (PEG 200), LiCl

and NaCl (Wheeler et al., 2012). Across the screen ethylene glycol and PEG200 was increased from 0 to 50%, covering relative humidities of 100 - 80 % (ethylene glycol) and 100 - 90% (PEG200). Both LiCl and NaCl were increased from 0 to 2.5 M, covering relative humidities of 100 – 95 %. Details of the screen can be found in materials and methods. The screen was added to the reservoirs and dehydration was allowed to proceed overnight. The plate was loaded onto the IO4-I beamline and diffraction was tested. For each desiccant trialed a dehydration point was determined. The dehydration point is a desiccant concentration. Above this concentration, the crystal becomes too dehydrated and diffraction is lost. Below the dehydration point, the desiccant does not have an effect on the hydration state of the crystal. Ethylene glycol and PEG200 had a dehydration points between 20 – 22.5 % and 17.5 – 20 %, respectively. LiCl and NaCl had dehydration point of 2.25 M. None of the relative humidities across the dehydration screen increased the diffraction of the crystals.

The dehydration thought to be responsible for increased X-ray diffraction of the YonO crystal was not controlled. Dehydration through air exposure is hard to reproduce (Heras and Martin, 2005) as many variables are in operation during air exposure including length of exposure, humidity and temperature. Despite using two alternative methods of controlled dehydration, we were unable to replicate the formation of crystals with high diffraction.

### **7.7. Crystallisation of the YonO homologue VM RNAP**

Homologous proteins are a potential solution to proteins which form poorly diffracting crystals. It is highly probable two closely related proteins are structurally very similar; meaning the structure of a second protein homologous to YonO could provide insights into the structure. Any structural variation between YonO and a homologue would likely be restricted to the surface of the protein. This is beneficial as the surface of the protein strongly influences the crystallisation process. A homologue of YonO may make different surface contacts leading to improved crystal packing and diffraction. BLASTp revealed the closest homologue to YonO is a hypothetical protein from *Bacillus vallismortis*, with the two proteins having 91% identity. The YonO homologue protein will be referred to as VM RNAP (Vallismortis RNAP). Analysis of the hypothetical proteins encoded alongside VM RNAP revealed the presence of phage like proteins. It is unknown if the bacteriophage is a functional. *B. vallismortis* is very closely related to

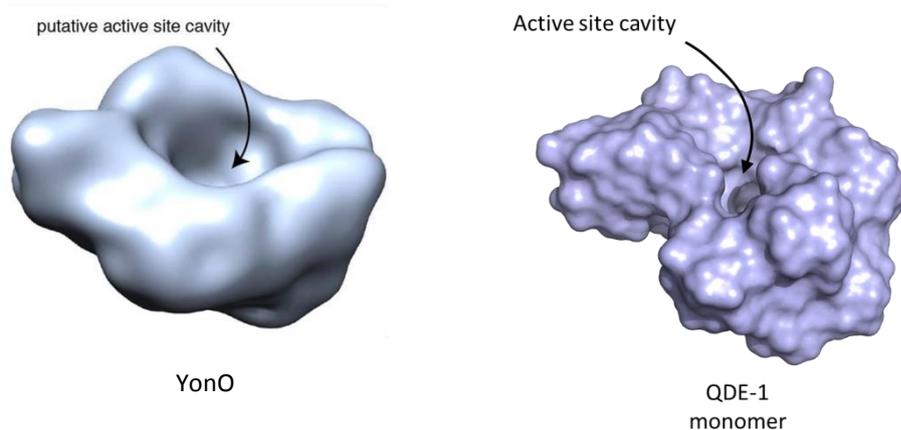
*B. subtilis* (Roberts et al., 1996). The two species can only be distinguished through their cell wall fatty acid composition and DNA sequence. The *B. vallismortis* strain was obtained from the Bacillus Genetic Stock Centre (BGSC) and the gene encoding VM RNAP was cloned into the pET-28 expression vector. VM RNAP was expressed and successfully purified to apparent homogeneity using the purification strategy developed for YonO, demonstrating the close similarities between YonO and VM RNAP. VM RNAP was trialled for crystallisation using the full collection of available commercial screens. The CoCl<sub>2</sub> screen, the custom screen developed for the optimisation of YonO crystallisation was also trialled. The crystallisation screens were left to equilibrate overnight. Equilibration against four conditions (E4, F4, G4 and H4) of the PACT screen and single condition (B4) of the JCSG+ screen resulted in VM RNAP forming small bipyramidal crystals. The crystals were harvested and tested for diffraction at Diamond Light Source. As with YonO, the diffraction was limited to 7 – 8 Å. The purification, crystal morphology and diffraction resolution of VM RNAP were extremely similar to YonO. This strongly suggests that VM RNAP maybe too close (91% identity) to YonO to crystallise differently. Alternative homologues with reduced sequence identity to YonO exist in other fermicutes. The reduced similarity may allow for different crystallisation and improved diffraction.

## **7.8. SAXS**

Whilst YonO readily crystallised, the organisation of the protein within the crystal did not permit diffraction of X-rays to a sufficient resolution. Alternative methods for determining the structure of a protein require solubilised protein rather than crystals. However, these techniques are not suitable for YonO; YonO is too large and too small for the determination of its structure by NMR and cryo-EM, respectively. Whilst it cannot provide a high resolution structure of YonO, small angle x-ray scattering (SAXS) would provide the molecular envelope revealing the overall shape of YonO. Purified YonO was supplied to our collaborator, Dr Abel Garcia-Pino of the VIB institute in Belgium who performed SAXS analysis.

The SAXS analysis revealed YonO potentially has a ring like shape, with a central cavity. We have tentatively concluded that the cavity is the pathway in which downstream DNA enters the enzyme. In keeping with other RNAPs, it is probable the active centre is located at the bottom of the cavity at the centre of the enzyme. In agreement with a

lack of amino acid sequence homology between YonO and RNAP, the domains of RNAP compared to pincers of a crab claw appear to be absent in YonO. The molecular envelope confirmed that YonO predominantly exists as a monomer. This is in contrast to QDE-1, the closest homologue of YonO with a determined structure, which was shown to be a dimer (Salgado et al., 2006). At this low resolution, it cannot be concluded if the overall shape of YonO is different from that of a QDE-1 monomer (compared in Figure 7-10).



**Figure 7-10: SAXS analysis of YonO.** SAXS analysis of YonO was performed by Dr Abel Garcia-pino. The resultant molecular envelope of YonO suggests it has a ring like structure with a central cavity which is hypothesised to be an entry channel for downstream and the location of the active site. However, the molecular envelope is a very low resolution structure and therefore it is possible the ring like structure is an artefact. The surface representation of QDE-1 is shown for comparison. QDE-1 does not have a ring like structure and the cavity is much narrower, reflecting the increased molecular weight of QDE-1 (160 KDa) compared to YonO (98 KDa).

## 7.9. Discussion

Highly pure and concentrated YonO was obtained and trialled for crystallisation.

Through optimisation with a customised crystallisation screen and the AmSO<sub>4</sub> screen, large bipyramidal crystals were reproducibly obtained. A sufficient diffraction resolution for structural determination is 3.5 Å and below. However, the diffraction of YonO crystals was limited to 7-8 Å. Low resolution diffraction often arises due to poor packing and order of the protein within the crystal lattice during crystallisation. Many techniques are available to increase the diffraction of crystals. Initially microseeding, additive screening and removal of the His tag were attempted to improve YonO

crystallisation and find alternative crystal morphologies. Only bipyramidal crystals diffracting to 7- 8 Å formed during our attempts to improve diffraction. During chromatography, we noticed a small percentage of YonO exists in a tetramer state. As it may have differing crystallisation properties including increased diffraction, the YonO tetramer was trialled for crystallisation. Despite forming crystals with a different morphology, diffraction of tetramer crystals was also limited to 7- 8 Å. Although not investigated, the YonO tetramer was not considered to be biologically relevant due to the small percentage of YonO protein existing in this state.

A single crystal did diffract to 3 Å, a sufficient resolution for structure determination. Two data sets were obtained from this crystal, both of which displayed twinning. The structure of QDE-1, the nearest homologue to YonO was used with data set 2 to attempt to solve the structure via molecular replacement but this was unsuccessful. Molecular replacement requires stronger homology than that between QDE-1 and YonO, explaining its failure to solve the structure. It was reasoned that the increased diffraction was due to dehydration of the crystal as a result of uncontrolled exposure to air during the testing of diffraction at room temperature. Attempts to repeat the growth of diffracting crystals with controlled dehydration were unsuccessful. It was estimated the crystal was exposed to air for 40 minutes. During that time, a number of variables were acting upon the crystal including exposure to the air, humidity and temperature. Reproduction of all the conditions which led to an increase in diffraction would be incredibly difficult. It could be suggested that the growth of a highly diffracting crystal during crystallisation was stochastic. The initial diffraction of the twinned crystal was poor when tested on the in-house X-ray generator, arguing against this notion.

High resolution diffraction of a crystal is dependent on the highly ordered organisation of the protein within the crystal lattice. Therefore, the inability of YonO to form well diffracting crystals can be attributed to disorder within the crystal. The organisation of a protein in a crystal is governed by the surface of the protein and any intrinsic flexibility. High surface entropy of a protein is detrimental to crystallisation as it introduces variation into the protein population (Dale et al., 2003). This variation reduces the proteins ability to form well-ordered assemblies within the crystal. The reduction of surface entropy is one of many strategies to improve diffraction and

crystallisation (Derewenda and Vekilov, 2006). Surface lysine residues are a source of entropy and are not favoured for crystal contacts (Dasgupta et al., 1997). The methylation of YonO surface lysines to reduce entropy resulted in precipitation, this is explained by the introduction of hydrophobicity on the most abundant amino acid in YonO (Walter et al., 2006).

Flexible domains of a protein can decrease ordered assembly of proteins in a crystal (Dale et al., 2003). A heterogeneous protein population, due to alternative conformations as a result of flexibility, will form crystals with increased disorder. Such crystals will diffract poorly. RNAP does have intrinsic flexibility, with the large clamp domain closing to make contacts with downstream DNA (Cramer et al., 2001). It is plausible to think YonO features mobile domains. Supporting this is the observation of different conformations in the structure of the YonO homologue QDE-1 (Salgado et al., 2006) Attempts to stabilise YonO in a single conformation through binding a nucleic acid scaffold failed due to high salt concentrations preventing YonO from forming stable elongation complexes. A rudimentary approach to reducing protein flexibility prior to crystallisation is limited *in situ* proteolysis (Dong et al., 2007, Wernimont and Edwards, 2009). At trace concentrations, proteases such as trypsin can cleave flexible domains and loops on the surface of the protein. The digested protein is then more amenable to ordered crystallisation, as potential sources of variation have been removed. However, a major limitation of this technique is the potential removal of domains required for structural integrity or flexible domains of interest and so this approach was not used.

To bypass crystallisation, the stage preventing the determination of the high resolution structure of YonO, SAXS analysis was performed. SAXS provided a low resolution image of YonO showing a ring like structure with a cavity designated as a putative DNA channel and active site. T7 RNAP and multisubunit RNAPs have structures resembling a right hand and crab claw, respectively. The shape of YonO as revealed by SAXS does not suggest that it has a structure similar to either of these polymerases. Additionally, SAXS revealed YonO exists in solution as a monomer unlike its distant homologue, QDE-1, which was shown to be a homodimer. The QDE-1 monomer has a claw structure, with the main channel running in-between the prominent head and slab domains. SAX data for QDE-1 is not available however; a surface representation of the

crystal structure suggests it is possible that at a low resolution QDE-1 may resemble a ring. Therefore, the ring-like structure of the YonO molecular envelope may be due to an inability of SAXS to resolve YonO domains analogous to the head and slab domains of QDE-1. Alternatively, YonO may indeed have a ring-like shape indicating a previously undescribed RNAP structure.

In this study, we have been unable to reproducibly obtain crystals that diffract at a suitable resolution for structure determination. A crystal structure of YonO would allow us to investigate the possibility of a ring-like shape as suggested by SAXS. Furthermore, it would shed light on the mechanisms employed by YonO to accomplish transcription and facilitate future *in vitro* characterisation by allowing structure based mutagenesis of YonO. Therefore, future work will include continued efforts to obtain the structure of YonO by X-ray crystallography.

An emerging way to address the problems arising from a protein's intrinsic flexibility is the use of the heavy chain only antibody variable fragments (VHH fragments) derived from camelid species (Griffin and Lawson, 2011). VHH fragments have aided in determination of structures by binding proteins and stabilising their conformations, thereby reducing protein heterogeneity and promoting increased order within the crystal lattice (Loris et al., 2003). The failure of proteins to make stable crystal contacts during crystallisation can also result in poorly diffracting crystals. This too can be addressed with VHH fragments, which are able to form stable contacts amongst themselves (Loris et al., 2003). Therefore, VHH antibodies against YonO will be obtained and trialled in crystallisation screens by our collaborator Dr Abel Garcia-Pino.

## Chapter 8: ORF6 RNA Polymerase of the *Kluyveromyces lactis* Killer System

### 8.1: Introduction

As discussed in section 3, the inability of the nuclear transcription machinery to recognise and express genes present on the *K. lactis* killer system plasmid, pGKL1, was the first evidence of a cytoplasmic transcription system specific for the *K. lactis* killer system (Romanos and Boyd, 1988, Stam et al., 1986). Further evidence came from heterologous expression of an antibiotic marker cloned in frame downstream of UCS1 promoter on pGKL1 (Tanguy-Rougeau et al., 1990).

The amino acid sequence of the 115 KDa protein encoded by ORF6 on pGKL2 was shown to be homologous to the  $\beta'$  and  $\beta$  subunit of RNAP (Ruprich-Robert and Thuriaux, 2010, Wilson and Meacock, 1988). Therefore, the ORF6 protein was identified as the RNAP central to the cytoplasmic transcription system specific for the *K. lactis* killer plasmids. Further analysis revealed approximately half of the functional motifs conserved across RNAP in all domains of life were absent in ORF6 RNAP (Ruprich-Robert and Thuriaux, 2010). Therefore, in addition to being of interest due to its cytoplasmic location and functioning in a previously uncharacterised transcription system, ORF6 RNAP may represent a novel class of RNAP exhibiting a minimal architecture. In addition to ORF6, ORF7 of pGKL2 is presumed to act as a subunit of the ORF6 RNAP. ORF7 encodes a 15 KDa protein with amino acid sequence homology with RNAP. In contrast to ORF6 RNAP, ORF7p sequence similarity is restricted to two regions at the C-Terminal end of  $\beta'$  (Schaffrath et al., 1997, Schaffrath et al., 1995b).

Both ORF6 RNAP and ORF7 have been shown to be essential for maintenance of pGKL1 and pGKL2. Attempts to delete ORF6 from pGKL2 by homologous recombination resulted in a mixed population of wildtype pGKL2 and  $\Delta$ ORF6 hybrid plasmids. Selection of the hybrid plasmid did not lead to a loss of the wildtype plasmid, showing ORF6 RNAP is essential (Schaffrath et al., 1995). Efforts to delete ORF7 resulted in a similar outcome (Schaffrath et al., 1997). Targeted deletion by homologous recombination led to integration of an auxotrophic marker into the chromosome, rather than the ORF7 locus on pGKL2. It was concluded the inability to replace ORF7 reflects its essentiality for plasmid maintenance.

The promoters of the ORF6 RNAP cytoplasmic transcription system have been identified. Short sequences termed upstream conserved sequences (UCS) have been located upstream each predicted ORF on pGKL1 and pGKL2. UCSs are sufficient to drive transcription of heterologous genes inserted into pGKL1 and pGKL2, thus confirming their role of promoters. However, no experimental evidence has been published showing a direct link between ORF6 RNAP and transcription from UCS promoters.

Generally speaking, the cytoplasmic transcription system remains poorly defined. Aside from ORF6 RNAP and ORF7p, ORF3p and ORF4p have been identified as the 5' methylguanoside capping enzyme and a putative ATP dependant helicase, respectively. Both ORF3p and ORF4p have been suggested to function in transcription. ORF1p, ORF8p, ORF9p and ORF11p are completely unlike any known proteins and so functions have not been assigned. The involvement of these proteins in transcription cannot be ruled out. An *in vitro* approach would allow investigation into the specific functions of the individual transcription proteins of pGKL2. To date, an *in vitro* approach has only been used to prove 5' triphosphatase and guanylyltransferase activities of ORF3p. ORF6 RNAP, the enzyme central to the transcription system, remains completely uninvestigated; its RNAP activity has not been confirmed by *in vitro* transcription assays. In addition to confirming ORF6 RNAP catalytic activity, *in vitro* experiments could contribute to the characterisation of all aspects of the transcription system, including the mechanism by which ORF6 RNAP performs RNA synthesis.

In the previous chapters, we have shown that YonO, despite apparently lacking every conserved RNAP domain except the aspartate triad, is a DNA dependant RNA polymerase. As ORF6 RNAP represents an intermediate between YonO and the multisubunit RNAP, we wanted to determine if ORF6 RNAP also has RNA polymerising activity.

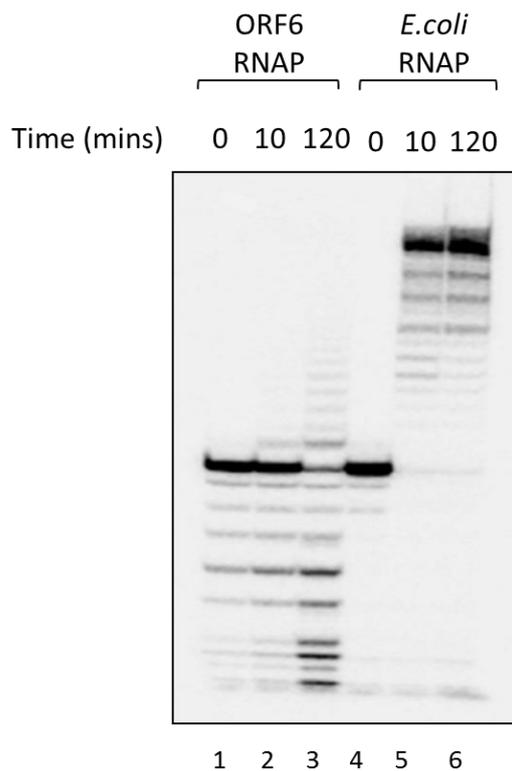
## **8.2. Aim**

The aim of this work was to begin the characterisation of ORF6 RNAP using an *in vitro* approach.

### 8.3. Previous Attempts to Express ORF6 RNAP

In order to carry out *in vitro* characterisation, purified ORF6 RNAP is required. Our previous attempts at expressing ORF6 RNAP in *S. cerevisiae* were unsuccessful (MRes thesis, 2011). We concluded the AT rich DNA sequence of ORF6 prevented expression by the nuclear transcription machinery. AT rich sequences have detrimental effects on transcription, possibly causing premature termination through coincidentally resembling termination signals and forming weak A:U hybrids.

As there was no expression in yeast, next we attempted expression in *E. coli* (MRes thesis, 2011). Initially, ORF6 was cloned into the pET-21 expression vector to allow IPTG inducible overexpression of ORF6 RNAP bearing a C-terminal His tag. Expression from pET-21 is T7 RNAP based and whilst this RNAP was able to transcribe the AT rich DNA sequence of ORF6, apparently all expressed protein formed inclusion bodies. To address this problem, ORF6 RNAP was purified from isolated inclusion bodies by Ni<sup>2+</sup>-NTA chromatography under denaturing conditions (8M urea). However, subsequent attempts to refold ORF6 RNAP were unsuccessful. It was hypothesised a small percentage of ORF6 RNAP was soluble but not purified due to the C-terminal His tag being embedded within the protein. Denaturing Ni<sup>2+</sup>-NTA chromatography on the soluble fraction of the lysate of induced cells confirmed this to be the case. Therefore, ORF6 was cloned into the pET-28 expression vector to give ORF6 RNAP an N-terminal His tag. This allowed the purification of a small quantity of soluble ORF6 RNAP. The ORF6 RNAP purified from *E. coli* was shown to have limited activity *in vitro*. Using promoter independent elongation complexes, ORF6 RNAP only extended RNA13 by 2 – 3 nucleotides after 120 minutes of incubation. It was hypothesised that the low activity may be due to improper expression of a eukaryotic protein in *E. coli*.



**Figure 8-1: The RNA polymerisation activity of ORF6 RNAP purified from *E. coli*.** Taken from MRes Thesis, 2011. Soluble ORF6 RNAP purified from *E. coli* was tested for activity by assembling elongation complexes using the GA36 scaffold. Upon addition of 500  $\mu$ M NTPs, complexes were incubated for 10 and 120 minutes. Reaction products were resolved by 23% denaturing PAGE and visualised using phosphorimaging.

#### 8.4: Expression of Codon Optimised ORF6 in *Saccharomyces cerevisiae*

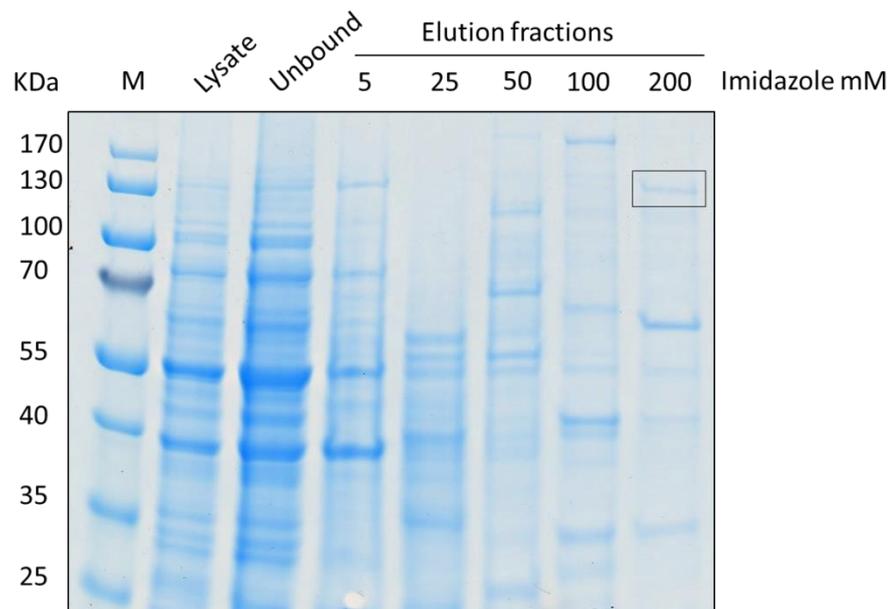
It was decided to re-attempt expression of ORF6 RNAP in *S. cerevisiae*, as it was suggested improper expression in *E. coli* led to a loss of ORF6 RNAP activity. However, to overcome the difficulty of AT rich DNA it was decided that the DNA sequence of ORF6 should be optimised for expression in *S. cerevisiae*. To optimise the DNA sequence for expression in yeast, the wildtype DNA sequence was supplied to the commercial company Eurogentec. The codon bias was altered to match that of *S. cerevisiae*, which in turn, reduced the AT content of ORF6 from 74% to 65%. The optimised ORF6 gene (from here on referred to as ORF6<sup>Ct</sup>) was synthesised and supplied cloned into pUC57, a commonly used cloning vector.

Previously, ORF6 RNAP was purified via Ni<sup>2+</sup>-NTA chromatography with an N-Terminal His tag. For expression of ORF6 RNAP from ORF6<sup>Ct</sup> in *S. cerevisiae*, ORF6<sup>Ct</sup> was cloned

into the pYES2/NT expression vector. pYES2/NT is a commercial 2 $\mu$  circular plasmid which introduces an N-terminal His tag and allows for P<sub>GAL1</sub> controlled overexpression. P<sub>GAL1</sub> is a strong promoter induced in response to activation of galactose metabolism. The pYES2/NT:ORF6<sup>C+</sup> construct was transformed into the commercially provided *S. cerevisiae* strain INVSc1.

To express ORF6 RNAP, INVSc1 carrying the pYES2/NT:ORF6<sup>C+</sup> construct was grown in synthetic complete media minus uracil (SC – URA). Growing cells in media lacking uracil allowed for selection of transformants as maintenance of the pYES2/NT:ORF6<sup>C+</sup> construct, which contained the URA3 selective marker, complemented INVSc1 uracil auxotrophy. The media was supplemented with glucose. Whilst glucose represses the P<sub>GAL1</sub> promoter, it allows for rapid growth and accumulation of biomass prior to induction of expression. Cells were harvested and washed before growth overnight in SC – URA medium supplemented with raffinose and galactose. Unlike glucose, the presence of raffinose does not inhibit galactose metabolism and induction of the P<sub>GAL1</sub> promoter and therefore provided an additional carbon source. For preliminary expression and purification of ORF6 RNAP, 6 litres of culture were harvested.

For purification of ORF6 RNAP, harvested cells were disrupted by French press and lysates were loaded onto a Ni<sup>2+</sup> - NTA column. Protein was eluted from the column with increasing concentrations of imidazole. Each elution fraction was analysed by SDS – PAGE to ascertain if ORF6 RNAP was present. ORF6 RNAP has a calculated molecular weight of 115 KDa. A protein of this molecular weight was present in the 200 mM imidazole elution fraction. Peptide mass fingerprinting confirmed this protein to be ORF6 RNAP. The reduction of AT content and codon optimisation was successful in allowing expression of ORF6 in *S. cerevisiae*. However, the yield of ORF6 RNAP was low. Poor heterologous protein expression is a major limitation of using *S. cerevisiae*. To address this, the purification was repeated and scaled up to 20 litres of yeast culture (Figure 8-2). Whilst this did increase the total amount of purified protein, compared to typical protein expression in *E. coli*, expression of ORF6 RNAP by *S. cerevisiae* remained low. Despite eluting in 200 mM imidazole, ORF6 RNAP was only partially purified. *S. cerevisiae* proteins are naturally rich in histidine residues, accounting for the contaminants also eluting at 200 mM imidazole.



**Figure 8-2: Ni<sup>2+</sup> - NTA Purification of ORF6 RNAP expressed from ORF6<sup>C+</sup> in *S. cerevisiae*.** *S. cerevisiae* strain INVSc1 carrying pYES2/NT:ORF6<sup>C+</sup> was grown overnight in SC – URA medium supplemented with raffinose and galactose. Harvested cells were disrupted by French – press and lysates were applied to a Ni<sup>2+</sup>-NTA column. Unbound protein was removed by washing with 5 mM imidazole elution buffer. The column was then washed with increasing concentrations of imidazole. All purification fractions were subjected to SDS-PAGE. A protein with ~ 115 KDa molecular weight (indicated by the box), that of ORF6 RNAP, was eluted by 200 mM imidazole fraction. Peptide mass fingerprint confirmed this protein to be ORF6 RNAP.

## 8.5. *In vitro* characterisation of ORF6 activity

### 8.5.1. Initial determination of ORF6 RNAP activity

As demonstrated by YonO, artificially assembled elongation complexes can be used to assay the polymerisation activity of a putative RNAP. Therefore, ORF6 RNAP was assembled into elongation complexes and tested for activity.

The ORF6 RNAP elongation complexes were assembled with the standard GA36 scaffold used throughout this work. ORF6 RNAP ECs were incubated with a high concentration (500 μM) of NTPs to see any potential RNAP activity. ORF6 RNAP exhibited poor polymerisation activity as after 30 minutes of incubation, RNA13 was only extended by 1 – 2 nucleotides. Furthermore, only a fraction of RNA13 was extended, indicating the ORF6 RNAP complexes were either paused or very slow. Previously, it was hypothesised the poor activity of ORF6 RNAP was due to improper



transcription. The pGKL1 and pGKL2 DNA thought to be transcribed by ORF6 RNAP is very AT rich, suggesting ORF6 RNAP is optimised for the transcription of AT rich DNA sequences. To investigate this possibility, we formed ORF6 RNAP elongation complexes with two alternative scaffolds: T-5S and GA36-8U (shown in Figure 8-4a). The length of the hybrid was varied by forming elongation complexes with Scaffold T-5S, which gives a 7 bp RNA:DNA hybrid. Template GA36-8U was hybridised to RNA13 to form the GA36-8U scaffold. The downstream sequence of the GA36-8U template is a poly-A tract, allowing us to see if ORF6 RNAP has a preference for AT rich DNA sequences. For comparison, *E. coli* core RNAP was assembled into elongation complexes with the same scaffolds. All elongation complexes were visualised by 5' <sup>32</sup>P labelling of the RNA oligonucleotide. As before, transcription reactions were performed in a high concentration of nucleotides (500 μM) with a long incubation time (30 minutes). Regardless of the scaffold, ORF6 RNAP only extended RNA13 by 1 – 2 nucleotides (Figure 8-4b). This suggests the hybrid length and downstream sequence of the template DNA are not contributing to the poor ORF6 RNAP activity. As expected, all RNA13 was extended by GA36 *E. coli* elongation complexes. The poly-A tract immediately downstream in the GA36-8U scaffold resembles a transcription termination signal and is known to cause RNAP to pause and backtrack thus explaining the low processivity of *E. coli* RNAP on this scaffold (Gusarov and Nudler, 1999, Nielsen et al., 2013).

All of our previous investigations into ORF6 RNAP activity have been with nucleic acid scaffolds featuring a DNA template, given its significant homology to the multisubunit RNAP. Despite accomplishing canonical gene expression in a DNA dependent manner, bacterial RNAP and RNAP II have been shown to possess RNA dependant activity (Lehmann et al., 2007). To investigate the possibility that ORF6 RNAP is strictly dependent on an RNA template, thus explaining the inactivity on the GA36 scaffold; we assembled elongation complexes with the RNA21 scaffold (Figure 8-4c). The complexes were incubated for 60 minutes in the presence of NTPs. In contrast to YonO and *E. coli* RNAP, ORF6 RNAP only extended the RNA10 primer by 1 -2 nucleotides during a 60 minute incubation. When incubated for that length of time with the GA36 scaffold, ORF6 RNAP extended the RNA beyond 1 -2 nucleotides suggesting a



GA36 and GA36-8U were used to form ORF6 RNAP elongation complexes. *E. coli* core RNAP elongation complexes using the same scaffolds were assembled for comparison. Non-template DNA was not included and the complexes were not immobilised. Transcription was initiated by the addition of 500  $\mu$ M NTPs and 10 mM  $Mg^{2+}$  and allowed to proceed for 30 minutes at 30 °C. Reactions were stopped by the addition of formamide containing stop buffer and were resolved on 23 % PAGE gel in denaturing conditions. Gels were visualised by phosphorimaging. **c) ORF6 RNAP does not utilise an RNA template.** ORF6 RNAP was incubated with the RNA21 scaffold to allow assembly of elongation complexes. YonO and *E. coli* RNAP complexes were also assembled to compare enzymatic activities. Transcription reactions were started by the addition of 500  $\mu$ M NTPs and incubated at 30 °C for 60 minutes.

### 8.5.3. Immobilisation of ORF6 RNAP Elongation complexes

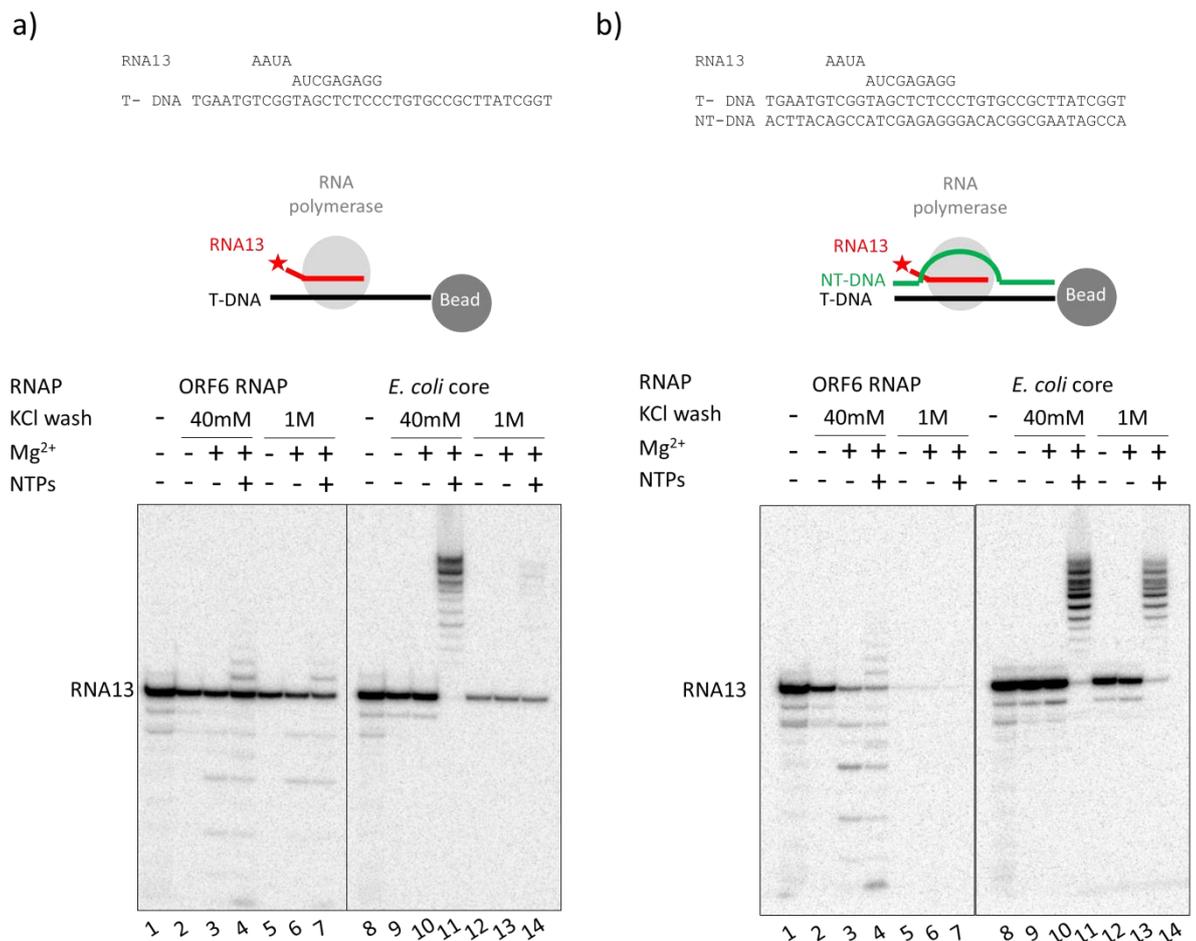
The ability of RNAP to tightly bind the hybrid increases the processivity of elongation complexes. Therefore, a possible explanation for the poor catalytic activity of ORF6 RNAP is an inability to stably bind the hybrid. The previous experiments were performed in solution, which does not directly show the stability of assembled complexes. To address this, ORF6 RNAP elongation complexes were assembled with the GA36 scaffold and immobilised to streptavidin sepharose beads through the 5' biotin tag of the template DNA (Figure 8-5a). Immobilisation can be used to assess if RNAP is binding the hybrid as unbound RNA13 and unstable complexes are removed by washing with transcription buffer (Sidorenkov et al., 1998). The stringency of this washing step can be improved by increasing the ionic strength of the transcription buffer. Initially, immobilised ORF6 RNAP elongation complexes were washed with low salt transcription buffer (40 mM KCl). Comparison of unwashed and washed elongation complexes (lane 1 and 2 of Figure 8-5a) shows that immobilised ORF6 RNAP elongation complexes withstood washing, suggesting ORF6 RNAP is capable of binding to the hybrid. This rules out weak hybrid binding as an explanation for the low processivity of ORF6 RNAP. To assess the stability of ORF6 RNAP elongation complexes, the immobilised complexes were also washed with transcription buffer containing 1 M KCl. RNA13 was present after washing, indicating the immobilised ORF6 RNAP complexes were stable. This was surprising given the non-template DNA strand was not included in the assembled complexes. Whilst not essential for transcription by artificial elongation complexes, the non-template DNA strand has been shown to be required for the assembly of salt stable *E. coli* elongation complexes (Sidorenkov et al., 1998). The non-template strand was included in assembled ORF6 RNAP elongation complexes to investigate its contribution to the stability and activity of ORF6 RNAP elongation

complexes. Immobilised complexes were washed with either high (1M KCl) or low ionic strength (40 mM KCl) transcription buffer. Surprisingly, the non-template strand had a destabilising effect on the ORF6 RNAP complexes, as all complexes disassembled in the presence of 1M KCl. Formation of *E. coli* complexes with and without the non-template DNA strand demonstrated the typical stabilising effect of the non-template strand. In the absence of this strand, the complexes were unable to withstand high ionic strength (compare lane 12 of Figure 8-5a to lane 12 of Figure 8-5b). Together, these results point towards the stability of ORF6 RNAP being determined by unusual interactions with the nucleic acid scaffold.

To confirm that the immobilised ORF6 RNAP complexes were catalytically active, washed complexes were incubated with  $Mg^{2+}$  and NTPs for 30 minutes (Figure 8-5b). As before, RNA13 was extended by 1-2 nucleotides showing the immobilised ORF6 RNAP were active. However, not all RNA13 was extended in 30 minutes. As only specific, assembled elongation complexes remain after washing, this showed that the majority of assembled ORF6 RNAP complexes are either incredibly slow or have entered an inactive state.

In addition to polymerisation, RNAP is able to hydrolyse RNA phosphodiester bonds as a function of proof reading (Orlova et al., 1995, Zenkin et al., 2006b). To see if ORF6 RNAP can catalyse this reaction, washed complexes were incubated with  $Mg^{2+}$  alone. In an  $Mg^{2+}$  dependant manner, RNA13 was shorted by dinucleotides to produce RNA11, RNA9 and RNA7 (lane 3, Figure 8-5a and b). This hydrolysis pattern is characteristic of the multisubunit RNAP, which catalyses hydrolysis of the penultimate phosphodiester bond to release a dinucleotide (Zenkin et al., 2006b). This suggests that despite not being able to polymerise efficiently, ORF6 RNAP is proficient at RNA hydrolysis. On this scaffold, ORF6 RNAP appeared to be more efficient at hydrolysis than *E. coli* RNAP. However, this scaffold does not favour hydrolysis by *E. coli* RNAP. The non-complementarity between the 5' end of RNA13 and the template DNA prevents the backtracking required for hydrolysis (refer to Figure 8-5a for the sequences of the scaffold). RNAP utilises the same active centre to synthesise and hydrolyse RNA (Steitz, 1998, Sosunov et al., 2003, Sosunov et al., 2005). This raises an interesting question regarding how ORF6 RNAP is able to accomplish hydrolysis but not polymerisation.

A possible explanation for the catalytic activities observed in ORF6 RNAP reactions is the presence of contaminating enzymes not removed by Ni<sup>2+</sup>-NTA chromatography. However, since extension and cleavage was observed after washing with 1M KCl, such contaminating enzymes would be required to tightly bind the hybrid and withstand high ionic strength. This makes it unlikely a contaminant is responsible for the observed activities. Furthermore, the same poor extension of RNA13 was accomplished by ORF6 RNAP purified from *E. coli*. This supports the suggestion that ORF6 RNAP is the enzyme responsible for extension of RNA13 by 1-2 nucleotides.



**Figure 8-5: Stability of ORF6 RNAP elongation complexes. a) Stability of elongation complexes assembled without non-template DNA. Top:** Sequences of the GA36 scaffold used for elongation complexes. **Middle:** Schematic representation of immobilised assembled elongation complexes. **Bottom:** Transcription by immobilised ORF6 RNAP elongation complexes. Elongation complexes were assembled and immobilised to streptavidin sepharose beads through the 5' biotin tag of template DNA. Non-template DNA was not included in the assembly of elongation complexes. Immobilised complexes were washed with low salt (40 mM KCl) or high salt (1 M KCl) transcription buffer. Complexes washed in high salt buffer were returned to low salt conditions prior to transcription. Reactions were started by the addition of 10 mM

Mg<sup>2+</sup> alone or 10 mM Mg<sup>2+</sup> and 500 μM NTPs. Reactions were incubated at 30 °C for 30 minutes before the addition of stop buffer. RNA were resolved by 23 % denaturing PAGE and visualised by phosphorimaging. **b) Stability of elongation complexes formed with non-template DNA strand.** Same for **a)** except the non-template DNA strand was included during assembly of elongation complexes.

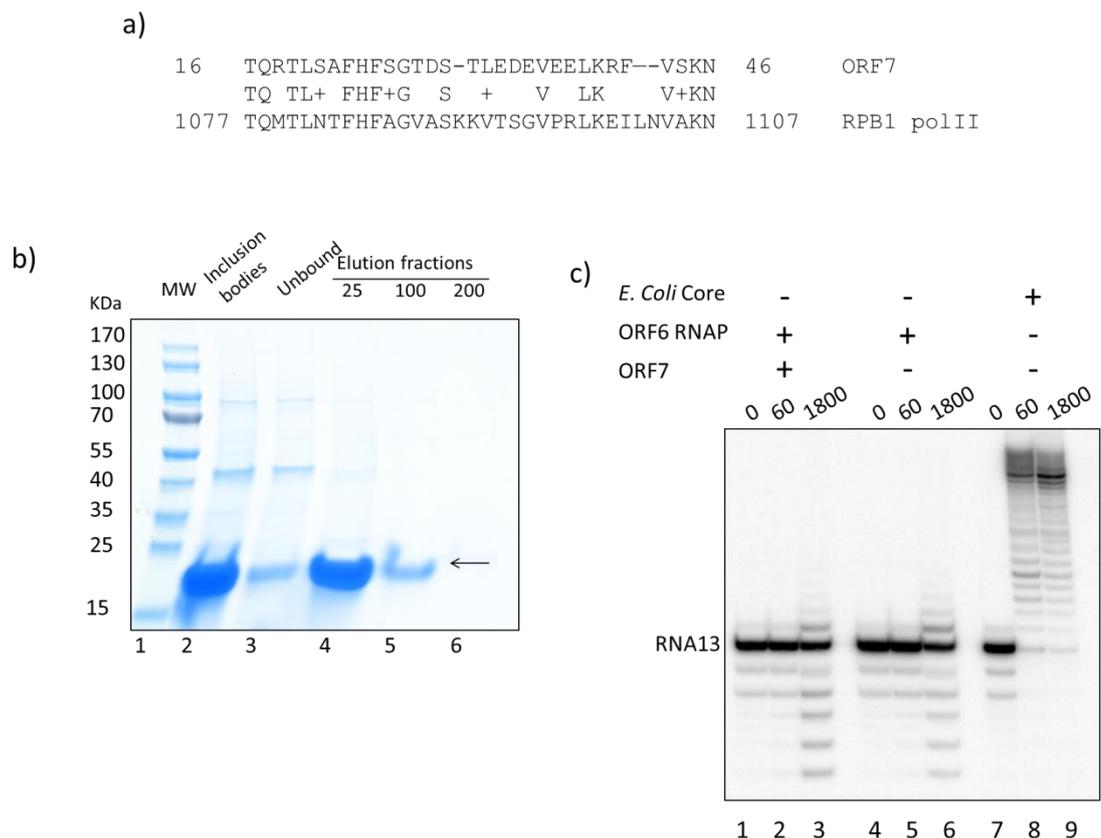
#### 8.5.4. ORF7p and the Trigger Loop

The poor extension of ORF6 RNAP elongation complexes resembles that of an RNAP with a deleted trigger loop (Yuzenkova et al., 2010). Supporting this notion, no amino acid sequence corresponding to the trigger loop has been identified in ORF6 RNAP. Together, these observations led us to hypothesise at least one other subunit, providing a domain functionally analogous to the trigger loop, is required for efficient transcription by ORF6 RNAP. ORF7p has been linked to the transcription system of the *K. lactis* killer system (Schaffrath et al., 1997, Schaffrath et al., 1995a). However, its small size (15 KDa) and limited homology to β' led us to previously assume it was not critical for ORF6 RNAP transcription. The published analysis of ORF7p homology to β' did not delineate the functional motifs conserved. Our updated analysis of the ORF7p amino acid sequence revealed previously unidentified homology to the trigger loop of RNAP II RPB1 (homologue of β') (Figure 8-6a).

To determine if ORF7p is able to contribute to catalysis by ORF6 RNAP, we set out to express and purify ORF7p. The predicted trigger loop is located towards the N-terminus of ORF7p. Given its small size, we sought to avoid altering the structure of a potential trigger loop through introduction of a His tag at the N-terminal end. Therefore, ORF7 was amplified from pGKL2 and cloned into pET-21 to allow for IPTG inducible overexpression of a C-terminal His-tagged ORF7p in *E. coli*. Similarly to ORF6 RNAP, ORF7p formed inclusion bodies when expression was induced by IPTG. A common technique to tackle inclusion bodies is purification under denaturing conditions followed by dialysis to induce refolding (Borukhov and Goldfarb, 1993). To this end, inclusion bodies were isolated, solubilised in 8 M urea and ORF7p was purified through Ni<sup>2+</sup> - NTA chromatography under denaturing conditions (Figure 8-6b). ORF7p weakly bound the column as 25 mM imidazole was sufficient for elution. ORF7p was diluted and dialysed against a standard storage buffer which was sufficient to remove the urea and promote refolding of ORF7p.

To see the effect of ORF7p on ORF6 RNAP transcription, an excess of ORF7p was incubated with ORF6 RNAP elongation complexes. Subsequent to complex assembly, transcription reactions were initiated with  $Mg^{2+}$  and NTPs and incubated for up to 30 minutes. Despite the presence of ORF7p, ORF6 RNAP still exhibited poor polymerisation activity (Figure 8-6c) which suggests ORF7p is not sufficient to increase the catalytic activity of ORF6 RNAP.

The addition of ORF7p to ORF6p *in vitro* assumes the two proteins would specifically bind and form a functional enzyme. The  $\omega$  subunit promotes assembly of RNAP by stabilising the  $\beta'$  subunit (Minakhin et al., 2001). If ORF6 RNAP and ORF7p require an analogous chaperone for proper assembly *in vivo*, addition of ORF7p *in vitro* would not be sufficient for the reconstitution of a fully functional ORF6/ORF7 enzyme. Furthermore, ORF7p was purified under denaturing conditions and so its improper refolding cannot be ruled out.



**Figure 8-6: ORF7p does not increase the catalytic activity of ORF6 RNAP. a) Homology of ORF7p to the trigger loop.** Manual inspection of the ORF7p and RPB1 (RNAP II  $\beta'$  homologue) amino acid sequence alignment revealed potential conservation of the trigger loop in ORF7p. **b) Purification of ORF7p.** ORF7p expression in *E. coli* cells was

induced by 1 mM IPTG. Resultant inclusion bodies were isolated, solubilised in 8M urea and subjected to Ni<sup>2+</sup>-NTA chromatography. ORF7p was eluted from the column by 25 mM imidazole. Denatured ORF7p was diluted in 8M urea and dialysed against ORF7 storage buffer. **c) Effect of ORF7p on the activities of ORF6 RNAP elongation complexes.** A 10 times molar excess of ORF7p was added to ORF6 RNAP during assembly of elongation complexes. Transcription reactions were started by the addition of 10 mM Mg<sup>2+</sup> and 500 μM NTPs and were allowed to proceed at 30 °C for the indicated time. Reaction products were resolved on a 23 % denaturing PAGE and visualised by phosphorimaging.

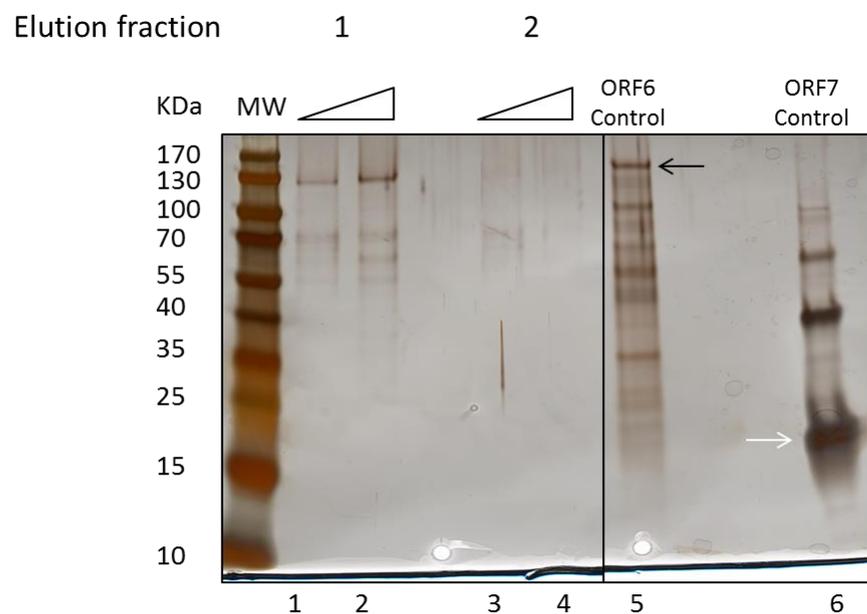
### **8.6. Co-expression of ORF6 RNAP and ORF7p in *S. cerevisiae***

The *in vitro* addition of ORF7 to elongation complexes assumes ORF7 is readily able to bind ORF6 RNAP to form an active enzyme. It was reasoned formation of a functional ORF6/ORF7 RNAP enzyme may require chaperones for correct co-folding and assembly. To address this we aimed to co-express and subsequently co-purify ORF6 RNAP and ORF7 from *S. cerevisiae*. The commercial plasmid pESC-URA was obtained to achieve this. pESC – URA is a circular plasmid that contains two divergent galactose induced promoters p<sub>GAL1</sub> and p<sub>GAL10</sub>. Cloning of ORF6<sup>C+</sup> under the control of p<sub>GAL10</sub> introduced an N-terminal FLAG tag. Cloning of ORF7 under the control of p<sub>GAL1</sub> introduced a C – Terminal C-myc tag. The cloned construct was transformed into the commercial *S. cerevisiae* YPH499 strain. Induction of expression was performed by growing cells in SC – URA medium supplemented with raffinose and galactose as for the expression of ORF6 RNAP from the pYES2/NT vector.

The hypothesis ORF7p assembles with ORF6 RNAP is based solely on small regions of homology between ORF7p and β'. Therefore, there is a possibility ORF7p may not actually bind to ORF6 RNAP. However, if our hypothesis is correct and the two proteins co-fold during expression in *S. cerevisiae*, it was reasoned ORF7p would co-purify with ORF6 RNAP. To purify FLAG tagged ORF6 RNAP, cells were disrupted by French press and the lysate was incubated in batch with anti-FLAG antibody resin. Subsequent to washing, the resin was packed into a gravity flow column and the bound protein was eluted with glycine pH 3. As before, the eluted proteins were resolved by SDS PAGE. However, the amount of eluted protein was too low to be detected by standard Coomassie blue staining. Therefore, protein eluted from the anti-FLAG resin was visualised by silver staining. A protein with the correct molecular weight of ORF6 RNAP was eluted from the anti-FLAG resin (lane 1 and 2, Figure 8-7). Compared to Ni<sup>2+</sup>-NTA chromatography, FLAG affinity chromatography resulted in a greater level of purity.

However, the yield of ORF6 RNAP purified by FLAG affinity chromatography was very low which reflects the poor expression of heterologous proteins in *S. cerevisiae* and the low protein binding capacity of antibody based resins (Lichty et al., 2005).

The sensitivity of the silver stain allowed us to confirm ORF7p did not elute from the anti-FLAG resin. From this purification, it cannot be discerned if failure to co-purify ORF7p was due to a lack of expression or lack of assembly with ORF6 RNAP. The AT rich sequence of ORF7 was not codon optimised for expression in *S. cerevisiae*. As AT rich DNA sequences negatively affect transcription, this suggests RNAP II's inability to transcribe ORF7 is responsible for the absence of co-purification.



**Figure 8-7: FLAG – tag affinity purification of ORF6 RNAP expressed by *S. cerevisiae*.** Co-expression of ORF6 RNAP and ORF7p was attempted in *S. cerevisiae*. ORF6<sup>C+</sup> and ORF7 were cloned under the control of galactose inducible pGAL10 and pGAL1 promoters in the pESC – URA vector. An N-terminal FLAG tag and C – terminal C-Myc tag were introduced on to ORF6 RNAP and ORF7p, respectively. To purify ORF6 RNAP, lysates from induced cells were incubated with anti-flag antibody resin. The resin was collected and bound proteins were eluted by washing with glycine pH 3. The proteins present in elution fractions were resolved by SDS PAGE and visualised using silver staining. ORF6 RNAP inclusion bodies solubilised in laemmli buffer and purified ORF7 were loaded as size controls. ORF6 RNAP and ORF7 in these samples are identified by black and white arrows, respectively.

### 8.7. Co-expression of ORF6 RNAP and ORF7p in *E. coli*

It was previously proposed that the AT rich DNA sequence of ORF7 prevented its transcription during our attempts to co-express ORF6 RNAP with ORF7p in *S. cerevisiae*. As both ORF6 RNAP and ORF7p were individually expressed by the T7 RNAP transcription system, co-expression was attempted in *E. coli*. To this end, ORF6<sup>C+</sup> and ORF7 were cloned into the pACYC Duet vector, which features two independent T7 promoters to allow for the simultaneous expression of two proteins upon IPTG induction. To ensure any co-purification is a result of ORF7p directly interacting with ORF6 RNAP, an N-terminal His-tag was only introduced onto ORF6 RNAP.

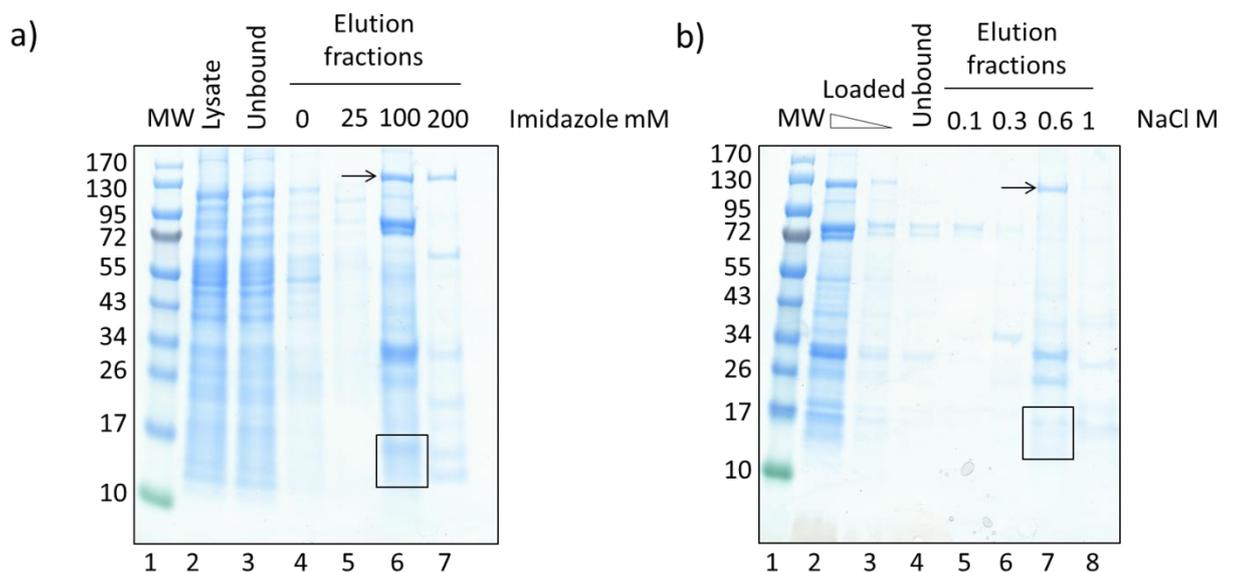
Previously, IPTG induced overexpression of ORF6 RNAP and ORF7p in *E. coli* resulted in the formation of inclusion bodies. To address this, co-expression was attempted overnight at 18 °C with and without 1 mM IPTG induction. The T7 promoter is not tightly regulated, resulting in background expression in the absence of IPTG induction. This leaky expression can be exploited to express difficult proteins such as those which would form inclusion bodies upon overexpression. Additionally, a shift to a lower temperature (18 °C) can also reduce inclusion body formation, as demonstrated by the expression of YonO.

Induction of ORF6 RNAP and ORF7 expression with 1 mM IPTG resulted in inclusion body formation, which was determined through visual inspection of pelleted cell debris after disruption of harvested cells. In contrast, inclusion bodies were not observed in the cell debris of uninduced cells. Therefore, the clarified lysates of uninduced cells were subjected to Ni<sup>2+</sup> - NTA chromatography. Washing the Ni<sup>2+</sup> - NTA column with 100 mM imidazole eluted a 115 KDa protein, which peptide mass fingerprinting confirmed to be ORF6 RNAP (Figure 8-8a).

PMF also revealed that a 15 KDa protein eluted by 100 mM imidazole was ORF7p. ORF7p was not expressed with a His tag, suggesting its co-purification with ORF6 RNAP is a result of the two proteins interacting. To purify the ORF6 RNAP/ORF7p complex further, the 100 mM imidazole elution fraction was applied to a Heparin HiTrap column. Both ORF6 RNAP and ORF7p were eluted from the column by 600 mM NaCl, providing further evidence of a direct interaction between the two proteins (Figure 8-8b).

It was hypothesised that the inability of ORF7p to increase the processivity and rate of ORF6 RNAP transcription was due to a lack of binding between the two proteins *in vitro*. To determine if the ORF6 RNAP/ORF7p complex had increased polymerisation activity, elongation complexes were assembled on the GA36 scaffold. 30 minutes of incubation in the presence of a high concentration of NTPs (500  $\mu$ M) revealed that the ORF6 RNAP/ORF7p elongation complex was no more active than ORF6 RNAP by itself (data not shown).

Analysis of the ORF7p amino acid sequence raised the possibility that this subunit provides ORF6 RNAP with a trigger loop, the mobile domain crucial for multisubunit RNAP catalysis (Temiakov et al., 2005, Yuzenkova et al., 2010, Yuzenkova and Zenkin, 2010). Whilst the co-purification of ORF6 RNAP and ORF7p after two chromatographies does not provide direct evidence of an interaction between the two proteins, it strongly suggests that this is the case. Therefore, the poor activity of the ORF6 RNAP/ORF7p complex indicates additional proteins maybe required for efficient ORF6 RNAP polymerisation.



**Figure 8-8: Co-purification of ORF6 RNAP and ORF7p expressed in *E. coli*.** a) Ni<sup>2+</sup>-NTA chromatography. T7 express cells carrying the pACYC:ORF6<sup>C+</sup>:ORF7 construct were grown overnight at 18 °C. The lysate of these cells was subjected to Ni<sup>2+</sup>-NTA chromatography. Once the lysate had been applied, the Ni<sup>2+</sup>-NTA column was washed by increasing concentrations of imidazole. The protein content of the elution fractions was visualised by SDS PAGE. Peptide mass finger printing was used to confirm that the band indicated by the black arrow corresponded to ORF6 RNAP. PMF also confirmed

ORF7p was present in the portion of the gel enclosed in the black square. **b) Heparin affinity chromatography.** The 100 mM imidazole fraction was diluted 4 times (to reduce the salt concentration) and loaded onto the Heparin HiTrap column. Proteins were eluted from the column by washings of increasing concentrations of NaCl. SDS PAGE showed that ORF6 RNAP and ORF7p were eluted from the Heparin HiTrap column by 600 mM NaCl. As for Figure 8-8a, the arrow and box indicate the bands confirmed by PMF to correspond to ORF6 RNAP and ORF7p.

## 8.8. Discussion

ORF6 RNAP is a putative, non-canonical RNAP of the *K. lactis* killer system. ORF6 RNAP is of interest due to it being a single protein that bears homology to both  $\beta'$  and  $\beta$  subunits of RNAP, whilst appearing to utilise only half of the conserved RNAP functional motifs. *In vitro* characterisation of this protein may result in the uncovering of novel mechanisms of transcription accomplished by a minimal RNAP.

Previously, the high AT content of the ORF6 DNA sequence prevented the expression of ORF6 RNAP in *S. cerevisiae* (MRes Thesis, 2011). The detrimental effect of high AT content on gene expression has not been limited to the expression of ORF6 RNAP. Other investigators of the *K. lactis* killer system have tackled the complications arising from high AT content by resorting to protein expression in less conventional organisms. Expression of ORF3p was achieved by *Bacillus megaterium*, an organism with a naturally increased AT content (Tiggemann et al., 2001) whilst the baculoviral insect expression system was used for the expression of ORF5p (Schaffrath and Meacock, 1995). To overcome high AT content of ORF6, we optimised the ORF6 DNA sequence to match the codon bias of *S. cerevisiae*. This approach was successful, allowing expression of ORF6 RNAP in *S. cerevisiae*.

When assayed with assembled elongation complexes, the *in vitro* activity of ORF6 RNAP purified from *S. cerevisiae* was poor, with the enzyme displaying the same slow, unprocessive activity as ORF6 RNAP purified from *E. coli* (MRes Thesis, 2011). Therefore, the poor activity of ORF6 RNAP purified from *E. coli* was not due to improper expression of a eukaryotic protein as we previously hypothesised. Rather, it is likely ORF6 RNAP's inefficiency stems from a failure to complete the nucleotide addition cycle (described in more detail in chapter 1.5). From the *in vitro* experiments presented, it cannot be determined if ORF6 RNAP's failure to polymerise is due to an inability to accomplish substrate binding, catalysis or translocation.

Despite its activity being slow, unprocessive and not comparable to other RNAPs, we still conclude ORF6 RNAP is a DNA dependant RNAP. This is based on its substantial amino acid homology to  $\beta'$  and  $\beta$  and our evidence showing ORF6 RNAP is able to bind an RNA:DNA hybrid to form a stable elongation complex. Immobilisation of complexes to streptavidin beads revealed that ORF6 RNAP contrasts to multisubunit RNAPs in that the non-template DNA strand reduces the stability of elongation complexes.

Additionally, a scaffold consisting of only the template DNA strand hybridised to RNA was sufficient for salt stable ORF6 RNAP complexes. At present these observations cannot be fully explained. However, they may point towards ORF6 RNAP elongation complexes having unusual properties, including an alternative binding and organisation of nucleic acid strands. If ORF6 RNAP elongation complexes do have unusual characteristics, the present nucleic acid scaffolds may not be suitable for processive polymerisation. Whilst a scaffold with a reduced hybrid length and differing downstream DNA sequence did not increase ORF6 RNAP processivity, scaffolds with alternative conformations of the non-template DNA strand were not tested. The influence of the non-template DNA strand on the stability of T7 RNAP elongation complexes was investigated by using nucleic acid scaffolds with alternative conformations. This included varying the length of the non-template strand along with its complementarity and position on the template DNA strand (Temiakov et al., 2002). A similar approach may allow the elucidation of the factors influencing ORF6 RNAP elongation complex stability and processivity.

Alternatively, an additional subunit may be required for efficient transcription by ORF6 RNAP. Inspection of its amino acid sequence homology with  $\beta'$  suggests ORF7p may provide ORF6 RNAP with a trigger loop. The fusion of the  $\beta'$  and  $\beta$  subunit similar to that seen in ORF6 RNAP has been reported previously (Ruprich-Robert and Thuriaux, 2010). However, to our knowledge, ORF7p providing a detachable trigger loop would be the first example of an RNAP organised in this way. ORF7p did not increase the activity of ORF6 RNAP, possibly due to an inability to correctly assemble into a functional enzyme *in vitro*. It was thought co-expression of the two proteins in *S. cerevisiae* would address this, as any chaperones required for co-folding or assembly would be present. However, our attempts to co-express ORF6 RNAP and ORF7p were unfruitful, as ORF7p did not co-purify during FLAG affinity chromatography of ORF6

RNAP. The AT rich DNA sequence of ORF7 may have prevented its expression during galactose induction, although this was considered unlikely due to the short length of ORF7. Alternatively, a *K. lactis* specific chaperone or factor may be required for efficient assembly. However, the *K. lactis* killer system can be transferred and stably maintained in *S. cerevisiae* strains with deleted mitochondrial DNA (Gunge et al., 1982, Gunge and Sakaguchi, 1981). The maintenance of pGKL1 and pGKL2 in *S. cerevisiae* implies a protein specific to *K. lactis* is not required for correct ORF6 RNAP and ORF7p expression. Therefore this points towards a pGKL2 encoded protein being required for correct ORF6 RNAP/ORF7p assembly.

Co-expression in *E. coli* did result in the co-purification of ORF7p with ORF6 RNAP, suggesting the two proteins interacted to form a complex. Above, it was proposed the absence of co-purification of ORF6 RNAP and ORF7p from *S. cerevisiae* was due to the requirement of a pGKL2 encoded protein. The apparent interaction of ORF6 RNAP and ORF7p during co-expression in bacteria undermines this hypothesis. Rather, it supports the alternative suggestion that the AT rich DNA sequence coding for ORF7p prevented its expression and subsequent co-purification in *S. cerevisiae*. Co-purification of ORF7p did not result in an increase in the catalytic activity of ORF6 RNAP. Whilst additional work would be required to directly show ORF6 RNAP and ORF7p are specifically interacting, the co-purification of ORF7p after two steps of chromatography suggests this to be the case. Therefore, the poor activity of ORF6 RNAP indicates the trigger loop is not sufficient for transcription. Based on the assumption that the nucleic acid scaffolds are suitable for the investigation of ORF6 RNAP, the weak activity of ORF6 RNAP bound to ORF7p could indicate additional proteins are required for efficient transcription. Besides ORF6 RNAP and ORF7p, there are at least 2 additional proteins encoded by pGKL2 that are predicted to function in transcription. Both ORF3p and ORF4p are related to the Vaccinia virus transcription machinery. Whilst ORF3p has been shown to possess mRNA capping activity, ORF4p is homologous to two Vaccinia proteins: the D6R subunit of ETF (early transcription factor) and NPH I (Larsen et al., 1998, Tiggemann et al., 2001, Stark et al., 1990). NPH I is a single stranded DNA dependant ATPase with predicted helicase activity (The Vaccinia virus transcription machinery is reviewed in Broyles, 2003). Despite being required for transcription termination, NPH I has also been shown to reduce pausing of the Vaccinia virus

elongation complex. It has been proposed that both the capping enzyme and NPH I belong to the elongation complex. Additionally, both proteins belong to the Vaccinia virus RNAP holoenzyme that is present in Vaccinia virus virions. Applying this information to the ORF6 RNAP transcription system leads to the suggestion that ORF3p and ORF4p belong to the ORF6 RNAP elongation complex and as such these proteins may be essential for efficient transcription.

Throughout this work, our investigation into the ORF6 RNAP transcription system has been hindered due to technical difficulties. The propensity for the pGKL2 encoded proteins to form inclusion bodies, combined with AT rich DNA sequences and poor expression of heterologous proteins by *S. cerevisiae*, has prevented us from easily obtaining pure and soluble ORF6 RNAP required for *in vitro* characterisation. Future work will include an *in vivo* approach to identify additional proteins required by ORF6 RNAP. However, this too has the potential to be technically challenging. Co-immunoprecipitation could be used to identify additional proteins interacting with ORF6 RNAP. However, this technique would require the production of antibodies raised against ORF6 RNAP, which typically depends on large quantities of protein purified to homogeneity. Antibodies can be generated from protein denatured in 8M urea and resolved by SDS PAGE. This would allow ORF6 RNAP inclusion bodies to be used for the production of antibodies. The need for antibodies against ORF6 RNAP could be circumvented by expressing ORF6 RNAP tagged with an epitope for which antibodies are commercially available. This can be achieved using the plasmid shuffle system that was developed as a tool for the genetic manipulation of pGKL2 (Schaffrath and Meacock, 1996). This system has already been used to transfer the ORF6 DNA sequence onto pGKL1, which complemented the deletion of the native ORF6 on pGKL2 (Schaffrath et al., 2000). It was supposed the shuffling of modified DNA could be exploited to introduce an epitope tag on ORF6 RNAP.

In addition to co-immunoprecipitation, antibodies against ORF6 RNAP would make a range of techniques available for the investigation of ORF6 RNAP. Specifically, the antibodies could be used for immunofluorescence microscopy. This would allow us to observe the localisation of ORF6 RNAP, which may shed light on the relationship between the *K. lactis* killer plasmids and the mitochondria. Alternatively, visualisation

of ORF6 RNAP *in vivo* could be achieved through the introduction of a GFP (green fluorescent protein) tag using the same plasmid shuffle system described above.

## Chapter 9: Concluding Discussion

This body of work was dedicated to the characterisation of two putative novel RNAPs, YonO and ORF6 RNAP. Whilst we have been largely unsuccessful in our attempts to characterise ORF6 RNAP, we have confirmed that YonO, a hypothetical protein of the SP $\beta$  bacteriophage, is expressed and essential for SP $\beta$  development. RNAseq analysis explained this essentiality by indicating that YonO is required for the expression of the late genes during the development of SP $\beta$ . YonO, in stark contrast to ORF6 RNAP, was proficient in transcription despite not possessing any detectable catalytic domains of the multisubunit RNAP aside from the aspartate triad. We showed that YonO functions as a DNA dependant RNAP, rather than a rdRNAP as predicted bioinformatically. We also showed that YonO is capable of initiating transcription on double stranded DNA independent of accessory factors. This is in sharp contrast to multisubunit RNAPs, that without exception require accessory factors to recognise and melt DNA and assist in the initial catalysis. RNAP recognition of DNA independent of  $\sigma^{70}$  has been reported for the synthesis of the M13 bacteriophage DNA replication primer (Zenkin and Severinov, 2004). However, even in this instance,  $\sigma^{70}$  was required for RNAP to pass through abortive initiation and into elongation. Furthermore,  $\sigma^{70}$  facilitates the binding of the first substrate NTP, initial catalysis and the progression into elongation (Kulbachinskiy and Mustaev, 2006). This and the ability to efficiently transcribe without essential domains of the multisubunit RNAP serve to highlight that YonO transcription is very unusual.

Despite ORF6 RNAP possessing significant homology to RNAP including half of the conserved active centre, it was unable to perform rapid and processive polymerisation. Provision of ORF7p, an additional protein encoded by pGKL2 bearing similarities to the trigger loop, did not increase ORF6 RNAP activity. From our results alone, we cannot conclude if ORF6 RNAP is a functional RNAP of the *K. lactis* killer system, as it did not polymerise RNA in manner characteristic of an RNAP. However, its essentiality, homology and the residual DNA dependant RNAP activity we observed strongly suggest it is indeed an RNAP (Schaffrath and Meacock, 1996, Schaffrath et al., 2000, Ruprich-Robert and Thuriaux, 2010). ORF6 RNAP and ORF7p are not the sole proteins of the *K. lactis* cytoplasmic transcription machinery. ORF3p and ORF4p have been

shown to be transcription proteins through experimental evidence and homology, respectively (Larsen et al., 1998, Tiggemann et al., 2001). Four pGKL2 ORFs are unlike any other proteins and so remain without assigned functions. Any number of these or nuclear encoded proteins may stimulate the activities of ORF6 RNAP through provision of catalytic residues like the bacterial Gre factors, acting as a protein chaperone akin to the  $\omega$  subunit or through an unknown mechanism. Considering a potential requirement for additional subunits, it may be that ORF6 RNAP is a 'non-canonical' multisubunit RNAP. Such RNAPs have been uncovered bioinformatically. 'Non-canonical' multisubunit RNAPs are typically viral and have differing subunit organisations and various degrees of homology to the  $\beta'$  and  $\beta$  subunits. This is demonstrated by the multisubunit RNAP of the insect infecting baculovirus (Ruprich-Robert and Thuriaux, 2010). Baculovirus RNAP is comprised of four subunits, two of which are distantly related to Rpb1 (RNAP II homologue of  $\beta'$ ) and Rpb2 (RNAP II homologue of  $\beta$ ) through conservation of the DPBB folds whilst the remaining subunits are unlike any other transcription proteins. In contrast to other non-canonical multisubunit RNAPs, ORF6 RNAP has significant homology to both  $\beta'$  and  $\beta$  subunits putting ORF6 RNAP into a stand-alone subfamily of 'non-canonical' RNAPs.

YonO and ORF6 RNAP have homologues in other bacterial and yeast species, respectively. The closest homologues of YonO reside in other firmicutes, including multiple *Bacillus* and *Clostridia* species. A YonO homologue is carried on the genome of the *Clostridium botulinum* strain D-1873 bacteriophage. This bacteriophage is maintained as a circular plasmid during pseudolysogeny and is responsible for encoding powerful neurotoxins (Hwang et al., 2007). Our work infers that the YonO homologue encoded by this phage is also capable of accomplishing transcription and contributes to the development of the bacteriophage. If this is the case, then investigations into YonO may have medical importance given the induction of D-1873 bacteriophage can spread toxigenicity amongst *C. botulinum* species (Oguma et al., 1976). More distant homologues of YonO have been identified in cyanobacteria. Reflecting their relationship to YonO, these proteins have also been labelled putative rdRNAPs (Iyer and Aravind, 2012). However, as demonstrated by our work, this does not rule out DNA dependant RNA polymerisation. A number of the cyanobacterial YonO homologues have acquired RNase H-like domains. Given RNase H degrades RNA

hybridised to DNA, the presence of an RNase H-like domain conjoined to a putative RNAP is intriguing. Given the use of cyanobacteria for biofuel production, novel RNAPs may be of biotechnological value. Linear cytoplasmic plasmids have been identified across a wide range of yeast genera including *Pichia etchellsii*, *Pichia acacia*, *Pichia inositovora* and *Saccharomyces kluyveri* (Hishinuma and Hirai, 1991, Jeske and Meinhardt, 2006, Klassen et al., 2001). Common to all of these species is the maintenance of a large cytoplasmic plasmid that is closely related to pGKL2. Amino acid sequence alignments revealed each plasmid encodes 10 – 11 proteins which, with the exception of ORF1, correspond to the proteins encoded by pGKL2. As ORF6 RNAP is conserved, it suggests that the utilisation of a cytoplasmic transcription system is common amongst yeast species harbouring linear cytoplasmic plasmids.

Given the presence of homologues across a wide range of organisms, YonO and ORF6 RNAP can be viewed as representatives for two new groups of RNAPs. In particular, our work confirming the RNAP activity of YonO can be used as a starting point for a new field of investigations into one of this new group of RNAPs. In this regard YonO is similar to T7 RNAP. T7 RNAP represents a whole family of nucleic acid polymerases including bacterial RNAPs, mitochondrial RNAP and DNAP I. The discoveries made from its investigation have shed light on the biology of a whole family of enzymes. An additional parallel which can be drawn between YonO and T7 RNAP is at the time of its discovery, T7 RNAP was also a novel RNAP encoded by a bacteriophage. The discovery and subsequent characterisation of T7 RNAP has greatly increased our understanding of transcription, with T7 RNAP still being investigated after 40 years of research. Furthermore, the understanding of the T7 RNAP transcription system led to the development of T7 RNAP based overexpression of recombinant proteins in *E. coli*. By providing a method of obtaining large quantities of RNA or protein, T7 RNAP overexpression has revolutionised molecular biology. According to the parallels running between YonO and T7 RNAP, the characterisation of YonO also has the potential to, whilst advancing our understanding of fundamental mechanisms of transcription, open up a new area of research. Depending on the outcomes of future investigations, such as a direct comparison between the expression levels of YonO and T7 RNAP, YonO could also be exploited for use as an alternative system for the overexpression of recombinant proteins.

## References

- Aalto, A. P., Poranen, M. M., Grimes, J. M., Stuart, D. I. & Bamford, D. H. 2010. In vitro activities of the multifunctional RNA silencing polymerase QDE-1 of *Neurospora crassa*. *J Biol Chem*, 285, 29367-74.
- Anagnostopoulos, C. & Spizizen, J. 1961. Requirements for Transformation in *Bacillus subtilis*. *J Bacteriol*, 81, 741-6.
- Andrecka, J., Treutlein, B., Arcusa, M. A., Muschielok, A., Lewis, R., Cheung, A. C., Cramer, P. & Michaelis, J. 2009. Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex. *Nucleic Acids Res*, 37, 5803-9.
- Aravind, L., Anantharaman, V., Zhang, D., De Souza, R. F. & Iyer, L. M. 2012. Gene flow and biological conflict systems in the origin and evolution of eukaryotes. *Front Cell Infect Microbiol*, 2, 89.
- Artsimovitch, I. & Landick, R. 2000. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A*, 97, 7090-5.
- Au, N., Kuester-Schoeck, E., Mandava, V., Bothwell, L. E., Canny, S. P., Chachu, K., Colavito, S. A., Fuller, S. N., Groban, E. S., Hensley, L. A., O'brien, T. C., Shah, A., Tierney, J. T., Tomm, L. L., O'gara, T. M., Goranov, A. I., Grossman, A. D. & Lovett, C. M. 2005. Genetic composition of the *Bacillus subtilis* SOS system. *J Bacteriol*, 187, 7655-66.
- Auchtung, J. M., Lee, C. A., Monson, R. E., Lehman, A. P. & Grossman, A. D. 2005. Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. *Proc Natl Acad Sci U S A*, 102, 12554-9.
- Awad, W., Svensson Birkedal, G., Thunnissen, M. M., Mani, K. & Logan, D. T. 2013. Improvements in the order, isotropy and electron density of glypican-1 crystals by controlled dehydration. *Acta Crystallogr D Biol Crystallogr*, 69, 2524-33.
- Bandwar, R. P. & Patel, S. S. 2001. Peculiar 2-aminopurine fluorescence monitors the dynamics of open complex formation by bacteriophage T7 RNA polymerase. *J Biol Chem*, 276, 14075-82.
- Banzhaf, M., Van Den Berg Van Saparoea, B., Terrak, M., Fraipont, C., Egan, A., Philippe, J., Zapun, A., Breukink, E., Nguyen-Disteche, M., Den Blaauwen, T. & Vollmer, W. 2012. Cooperativity of peptidoglycan synthases active in bacterial cell elongation. *Mol Microbiol*, 85, 179-94.
- 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, 183-93.
- Bar-Nahum, G. & Nudler, E. 2001. Isolation and characterization of sigma(70)-retaining transcription elongation complexes from *Escherichia coli*. *Cell*, 106, 443-51.

- Batada, N. N., Westover, K. D., Bushnell, D. A., Levitt, M. & Kornberg, R. D. 2004. Diffusion of nucleoside triphosphates and role of the entry site to the RNA polymerase II active center. *Proc Natl Acad Sci U S A*, 101, 17361-4.
- Bergfors, T. 2003. Seeds to crystals. *J Struct Biol*, 142, 66-76.
- Bochkareva, A., Yuzenkova, Y., Tadigotla, V. R. & Zenkin, N. 2012. Factor-independent transcription pausing caused by recognition of the RNA-DNA hybrid sequence. *EMBO J*, 31, 630-9.
- Borukhov, S. & Goldfarb, A. 1993. Recombinant *Escherichia coli* RNA polymerase: purification of individually overexpressed subunits and in vitro assembly. *Protein Expr Purif*, 4, 503-11.
- Borukhov, S. & Nudler, E. 2008. RNA polymerase: the vehicle of transcription. *Trends Microbiol*, 16, 126-34.
- Broyles, S. S. 2003. Vaccinia virus transcription. *J Gen Virol*, 84, 2293-303.
- Butler, A. R., O'donnell, R. W., Martin, V. J., Gooday, G. W. & Stark, M. J. 1991. *Kluyveromyces lactis* toxin has an essential chitinase activity. *Eur J Biochem*, 199, 483-8.
- Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L. & Darst, S. A. 2002. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell*, 9, 527-39.
- Castillo, R. M., Mizuguchi, K., Dhanaraj, V., Albert, A., Blundell, T. L. & Murzin, A. G. 1999. A six-stranded double-psi beta barrel is shared by several protein superfamilies. *Structure*, 7, 227-36.
- Chakraborty, A., Wang, D., Ebright, Y. W., Korlann, Y., Kortkhonjia, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Jansen, R., Nixon, B. T., Knight, J., Weiss, S. & Ebright, R. H. 2012. Opening and closing of the bacterial RNA polymerase clamp. *Science*, 337, 591-5.
- Chamberlin, M., Mcgrath, J. & Waskell, L. 1970. New RNA polymerase from *Escherichia coli* infected with bacteriophage T7. *Nature*, 228, 227-31.
- Chandry, P. S., Moore, S. C., Boyce, J. D., Davidson, B. E. & Hillier, A. J. 1997. Analysis of the DNA sequence, gene expression, origin of replication and modular structure of the *Lactococcus lactis* lytic bacteriophage sk1. *Mol Microbiol*, 26, 49-64.
- Cheetham, G. M. & Steitz, T. A. 1999. Structure of a transcribing T7 RNA polymerase initiation complex. *Science*, 286, 2305-9.
- Cipres-Palacin, G. & Kane, C. M. 1994. Cleavage of the nascent transcript induced by TFIIS is insufficient to promote read-through of intrinsic blocks to elongation by RNA polymerase II. *Proc Natl Acad Sci U S A*, 91, 8087-91.

Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R. & Kornberg, R. D. 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science*, 288, 640-9.

Cramer, P., Bushnell, D. A. & Kornberg, R. D. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*, 292, 1863-76.

D'arcy, A., Villard, F. & Marsh, M. 2007. An automated microseed matrix-screening method for protein crystallization. *Acta Crystallogr D Biol Crystallogr*, 63, 550-4.

Dale, G. E., Oefner, C. & D'arcy, A. 2003. The protein as a variable in protein crystallization. *J Struct Biol*, 142, 88-97.

Darst, S. A., Kubalek, E. W. & Kornberg, R. D. 1989. Three-dimensional structure of *Escherichia coli* RNA polymerase holoenzyme determined by electron crystallography. *Nature*, 340, 730-2.

Dasgupta, S., Iyer, G. H., Bryant, S. H., Lawrence, C. E. & Bell, J. A. 1997. Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins*, 28, 494-514.

Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. 1990. An attempt to unify the structure of polymerases. *Protein Eng*, 3, 461-7.

Derewenda, Z. S. & Vekilov, P. G. 2006. Entropy and surface engineering in protein crystallization. *Acta Crystallogr D Biol Crystallogr*, 62, 116-24.

Dong, A., Xu, X., Edwards, A. M., Midwest Center for Structural, G., Structural Genomics, C., Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., Egorova, O., Evdokimova, E., Filippova, E., Gu, J., Guthrie, J., Ignatchenko, A., Joachimiak, A., Klostermann, N., Kim, Y., Korniyenko, Y., Minor, W., Que, Q., Savchenko, A., Skarina, T., Tan, K., Yakunin, A., Yee, A., Yim, V., Zhang, R., Zheng, H., Akutsu, M., Arrowsmith, C., Avvakumov, G. V., Bochkarev, A., Dahlgren, L. G., Dhe-Paganon, S., Dimov, S., Dombrowski, L., Finerty, P., Jr., Flodin, S., Flores, A., Graslund, S., Hammerstrom, M., Herman, M. D., Hong, B. S., Hui, R., Johansson, I., Liu, Y., Nilsson, M., Nedyalkova, L., Nordlund, P., Nyman, T., Min, J., Ouyang, H., Park, H. W., Qi, C., Rabeh, W., Shen, L., Shen, Y., Sukumard, D., Tempel, W., Tong, Y., Tresagues, L., Vedadi, M., Walker, J. R., Weigelt, J., Welin, M., Wu, H., Xiao, T., Zeng, H. & Zhu, H. 2007. In situ proteolysis for protein crystallization and structure determination. *Nat Methods*, 4, 1019-21.

Douangamath, A., Aller, P., Lukacik, P., Sanchez-Weatherby, J., Moraes, I. & Brandao-Neto, J. 2013. Using high-throughput in situ plate screening to evaluate the effect of dehydration on protein crystals. *Acta Crystallographica Section D*, 69, 920-923.

Durniak, K. J., Bailey, S. & Steitz, T. A. 2008. The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. *Science*, 322, 553-7.

Ederth, J., Artsimovitch, I., Isaksson, L. A. & Landick, R. 2002. The downstream DNA jaw of bacterial RNA polymerase facilitates both transcriptional initiation and pausing. *J Biol Chem*, 277, 37456-63.

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

Epshtein, V., Dutta, D., Wade, J. & Nudler, E. 2010. An allosteric mechanism of Rho-dependent transcription termination. *Nature*, 463, 245-9.

Ferrer-Orta, C., Arias, A., Escarmis, C. & Verdaguer, N. 2006. A comparison of viral RNA-dependent RNA polymerases. *Curr Opin Struct Biol*, 16, 27-34.

Fukuhara, H. 1987. The RF1 gene of the killer DNA of yeast may encode a DNA polymerase. *Nucleic Acids Res*, 15, 10046.

Gietz, D., St Jean, A., Woods, R. A. & Schiestl, R. H. 1992. Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res*, 20, 1425.

Gildehaus, N., Neusser, T., Wurm, R. & Wagner, R. 2007. Studies on the function of the riboregulator 6S RNA from *E. coli*: RNA polymerase binding, inhibition of in vitro transcription and synthesis of RNA-directed de novo transcripts. *Nucleic Acids Res*, 35, 1885-96.

Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. 2001. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science*, 292, 1876-82.

Gong, X. Q., Zhang, C., Feig, M. & Burton, Z. F. 2005. Dynamic error correction and regulation of downstream bubble opening by human RNA polymerase II. *Mol Cell*, 18, 461-70.

Goranov, A. I., Kuester-Schoeck, E., Wang, J. D. & Grossman, A. D. 2006. Characterization of the global transcriptional responses to different types of DNA damage and disruption of replication in *Bacillus subtilis*. *J Bacteriol*, 188, 5595-605.

Griffin, L. & Lawson, A. 2011. Antibody fragments as tools in crystallography. *Clin Exp Immunol*, 165, 285-91.

Gunge, N., Murata, K. & Sakaguchi, K. 1982. Transformation of *Saccharomyces cerevisiae* with linear DNA killer plasmids from *Kluyveromyces lactis*. *J Bacteriol*, 151, 462-4.

Gunge, N. & Sakaguchi, K. 1981. Intergeneric transfer of deoxyribonucleic acid killer plasmids, pGKI1 and pGKI2, from *Kluyveromyces lactis* into *Saccharomyces cerevisiae* by cell fusion. *J Bacteriol*, 147, 155-60.

Gunge, N., Tamaru, A., Ozawa, F. & Sakaguchi, K. 1981. Isolation and characterization of linear deoxyribonucleic acid plasmids from *Kluyveromyces lactis* and the plasmid-associated killer character. *J Bacteriol*, 145, 382-90.

- Gusarov, I. & Nudler, E. 1999. The mechanism of intrinsic transcription termination. *Mol Cell*, 3, 495-504.
- Haag, J. R. & Pikaard, C. S. 2011. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol*, 12, 483-92.
- Haag, J. R., Ream, T. S., Marasco, M., Nicora, C. D., Norbeck, A. D., Pasa-Tolic, L. & Pikaard, C. S. 2012. In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. *Mol Cell*, 48, 811-8.
- Harris, J. R. 1997. Negative Staining and Cryoelectron Microscopy: The Thin Film Techniques. *RMS Microscopy Handbook*, 35.
- Harris, J. R. 2008. Cholesterol binding to amyloid-beta fibrils: a TEM study. *Micron*, 39, 1192-6.
- Harwood, C. R. & Cutting, S. M. 1991. Chapter 10: Bacteriophages. *Molecular Biological Methods for Bacillus*. Wiley.
- Heras, B. & Martin, J. L. 2005. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr D Biol Crystallogr*, 61, 1173-80.
- Herbert, K. M., Zhou, J., Mooney, R. A., Porta, A. L., Landick, R. & Block, S. M. 2010. *E. coli* NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. *J Mol Biol*, 399, 17-30.
- Hishinuma, F. & Hirai, K. 1991. Genome organization of the linear plasmid, pSKL, isolated from *Saccharomyces kluyveri*. *Mol Gen Genet*, 226, 97-106.
- Hishinuma, F., Nakamura, K., Hirai, K., Nishizawa, R., Gunge, N. & Maeda, T. 1984. Cloning and nucleotide sequences of the linear DNA killer plasmids from yeast. *Nucleic Acids Res*, 12, 7581-97.
- Huang, J., Brieba, L. G. & Sousa, R. 2000. Misincorporation by wild-type and mutant T7 RNA polymerases: identification of interactions that reduce misincorporation rates by stabilizing the catalytically incompetent open conformation. *Biochemistry*, 39, 11571-80.
- Hwang, H. J., Lee, J. C., Yamamoto, Y., Sarker, M. R., Tsuchiya, T. & Oguma, K. 2007. Identification of structural genes for *Clostridium botulinum* type C neurotoxin-converting phage particles. *FEMS Microbiol Lett*, 270, 82-9.
- Ireton, G. C. & Stoddard, B. L. 2004. Microseed matrix screening to improve crystals of yeast cytosine deaminase. *Acta Crystallogr D Biol Crystallogr*, 60, 601-5.
- Iyer, L. M. & Aravind, L. 2012. Insights from the architecture of the bacterial transcription apparatus. *J Struct Biol*, 179, 299-319.

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 Struct Biol*, 3, 1.

Izban, M. G. & Luse, D. S. 1992. The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII. *Genes Dev*, 6, 1342-56.

Jeske, S. & Meinhardt, F. 2006. Autonomous cytoplasmic linear plasmid pPac1-1 of *Pichia acaciae*: molecular structure and expression studies. *Yeast*, 23, 479-86.

Juang, Y. L. & Helmann, J. D. 1994. A promoter melting region in the primary sigma factor of *Bacillus subtilis*. Identification of functionally important aromatic amino acids. *J Mol Biol*, 235, 1470-88.

Juers, D. H. & Matthews, B. W. 2001. Reversible lattice repacking illustrates the temperature dependence of macromolecular interactions. *J Mol Biol*, 311, 851-62.

Jung, G. H., Leavitt, M. C. & Ito, J. 1987. Yeast killer plasmid pGKL1 encodes a DNA polymerase belonging to the family B DNA polymerases. *Nucleic Acids Res*, 15, 9088.

Kashkina, E., Anikin, M., Tahirov, T. H., Kochetkov, S. N., Vassilyev, D. G. & Temiakov, D. 2006. Elongation complexes of *Thermus thermophilus* RNA polymerase that possess distinct translocation conformations. *Nucleic Acids Res*, 34, 4036-45.

Kettenberger, H., Armache, K. J. & Cramer, P. 2003. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell*, 114, 347-57.

Kettenberger, H., Armache, K. J. & Cramer, P. 2004. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell*, 16, 955-65.

Kireeva, M. L. & Kashlev, M. 2009. Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proc Natl Acad Sci U S A*, 106, 8900-5.

Klassen, R., Tontsidou, L., Larsen, M. & Meinhardt, F. 2001. Genome organization of the linear cytoplasmic element pPE1B from *Pichia etchellsii*. *Yeast*, 18, 953-61.

Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A. & Darst, S. A. 2000. A structural model of transcription elongation. *Science*, 289, 619-25.

Korzheva, N., Mustaev, A., Nudler, E., Nikiforov, V. & Goldfarb, A. 1998. Mechanistic model of the elongation complex of *Escherichia coli* RNA polymerase. *Cold Spring Harb Symp Quant Biol*, 63, 337-45.

Kriminski, S., Caylor, C. L., Nonato, M. C., Finkelstein, K. D. & Thorne, R. E. 2002. Flash-cooling and annealing of protein crystals. *Acta Crystallogr D Biol Crystallogr*, 58, 459-71.

Krogh, S., Jorgensen, S. T. & Devine, K. M. 1998. Lysis genes of the *Bacillus subtilis* defective prophage PBSX. *J Bacteriol*, 180, 2110-7.

Krogh, S., O'reilly, M., Nolan, N. & Devine, K. M. 1996. The phage-like element PBSX and part of the skin element, which are resident at different locations on the *Bacillus subtilis* chromosome, are highly homologous. *Microbiology*, 142 ( Pt 8), 2031-40.

Kuehner, J. N., Pearson, E. L. & Moore, C. 2011. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol*, 12, 283-94.

Kulbachinskiy, A. & Mustaev, A. 2006. Region 3.2 of the sigma subunit contributes to the binding of the 3'-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *J Biol Chem*, 281, 18273-6.

Kunkel, B., Losick, R. & Stragier, P. 1990. The *Bacillus subtilis* gene for the development transcription factor sigma K is generated by excision of a dispensable DNA element containing a sporulation recombinase gene. *Genes Dev*, 4, 525-35.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S. K., Cordani, J. J., Connerton, I. F., Cummings, N. J., Daniel, R. A., Denziot, F., Devine, K. M., Dusterhoft, A., Ehrlich, S. D., Emmerson, P. T., Entian, K. D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S. Y., Glaser, P., Goffeau, A., Golightly, E. J., Grandi, G., Guiseppi, G., Guy, B. J., Haga, K., Haiech, J., Harwood, C. R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M. F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S. M., Levine, A., Liu, H., Masuda, S., Mael, C., Medigue, C., Medina, N., Mellado, R. P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S. H., Parro, V., Pohl, T. M., Portelle, D., Porwollik, S., Prescott, A. M., Presecan, E., Pujic, P., Purnelle, B., et al. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature*, 390, 249-56.

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. *EMBO J*, 21, 1369-78.

Landick, R. 2006. The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc Trans*, 34, 1062-6.

Lang, W. H., Morrow, B. E., Ju, Q., Warner, J. R. & Reeder, R. H. 1994. A model for transcription termination by RNA polymerase I. *Cell*, 79, 527-34.

Laptenko, O., Lee, J., Lomakin, I. & Borukhov, S. 2003. Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J*, 22, 6322-34.

- Larsen, M., Gunge, N. & Meinhardt, F. 1998. *Kluyveromyces lactis* killer plasmid pGKL2: evidence for a viral-like capping enzyme encoded by ORF3. *Plasmid*, 40, 243-6.
- Larsen, M. & Meinhardt, F. 2000. *Kluyveromyces lactis* killer system: identification of a new gene encoded by pGKL2. *Curr Genet*, 38, 271-5.
- Laurila, M. R., Salgado, P. S., Makeyev, E. V., Nettelship, J., Stuart, D. I., Grimes, J. M. & Bamford, D. H. 2005. Gene silencing pathway RNA-dependent RNA polymerase of *Neurospora crassa*: yeast expression and crystallization of selenomethionated QDE-1 protein. *J Struct Biol*, 149, 111-5.
- Lawson, D. M., Artymiuk, P. J., Yewdall, S. J., Smith, J. M., Livingstone, J. C., Treffry, A., Luzzago, A., Levi, S., Arosio, P., Cesareni, G. & Et Al. 1991. Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature*, 349, 541-4.
- Lazarevic, V., Dusterhoft, A., Soldo, B., Hilbert, H., Mael, C. & Karamata, D. 1999. Nucleotide sequence of the *Bacillus subtilis* temperate bacteriophage SPbetac2. *Microbiology*, 145 ( Pt 5), 1055-67.
- Lee, C. A., Auchtung, J. M., Monson, R. E. & Grossman, A. D. 2007. Identification and characterization of int (integrase), xis (excisionase) and chromosomal attachment sites of the integrative and conjugative element ICEBs1 of *Bacillus subtilis*. *Mol Microbiol*, 66, 1356-69.
- Lee, H. C., Chang, S. S., Choudhary, S., Aalto, A. P., Maiti, M., Bamford, D. H. & Liu, Y. 2009. qiRNA is a new type of small interfering RNA induced by DNA damage. *Nature*, 459, 274-7.
- Lehmann, E., Brueckner, F. & Cramer, P. 2007. Molecular basis of RNA-dependent RNA polymerase II activity. *Nature*, 450, 445-9.
- Lenhart, J. S., Schroeder, J. W., Walsh, B. W. & Simmons, L. A. 2012. DNA repair and genome maintenance in *Bacillus subtilis*. *Microbiol Mol Biol Rev*, 76, 530-64.
- Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J. & Tan, S. 2005. Comparison of affinity tags for protein purification. *Protein Expr Purif*, 41, 98-105.
- Longchamp, P. F., Mael, C. & Karamata, D. 1994. Lytic enzymes associated with defective prophages of *Bacillus subtilis*: sequencing and characterization of the region comprising the N-acetylmuramoyl-L-alanine amidase gene of prophage PBSX. *Microbiology*, 140 ( Pt 8), 1855-67.
- Loris, R., Marianovsky, I., Lah, J., Laeremans, T., Engelberg-Kulka, H., Glaser, G., Muyldermans, S. & Wyns, L. 2003. Crystal structure of the intrinsically flexible addiction antidote MazE. *J Biol Chem*, 278, 28252-7.
- Magliani, W., Conti, S., Gerloni, M., Bertolotti, D. & Polonelli, L. 1997. Yeast killer systems. *Clin Microbiol Rev*, 10, 369-400.

- Maida, Y., Yasukawa, M., Furuuchi, M., Lassmann, T., Possemato, R., Okamoto, N., Kasim, V., Hayashizaki, Y., Hahn, W. C. & Masutomi, K. 2009. An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. *Nature*, 461, 230-5.
- Martinez-Rucobo, F. W. & Cramer, P. 2013. Structural basis of transcription elongation. *Biochim Biophys Acta*, 1829, 9-19.
- Matthews, B. W. 1968. Solvent content of protein crystals. *J Mol Biol*, 33, 491-7.
- Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. 2007. Phaser crystallographic software. *J Appl Crystallogr*, 40, 658-674.
- Mcneel, D. G. & Tamanoi, F. 1991. Terminal region recognition factor 1, a DNA-binding protein recognizing the inverted terminal repeats of the pGKI linear DNA plasmids. *Proc Natl Acad Sci U S A*, 88, 11398-402.
- Minakhin, L., Bhagat, S., Brunning, A., Campbell, E. A., Darst, S. A., Ebright, R. H. & Severinov, K. 2001. Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci U S A*, 98, 892-7.
- Moss, B. 1990. Regulation of vaccinia virus transcription. *Annu Rev Biochem*, 59, 661-88.
- MRes Thesis: Unusual Cytoplasmic RNA Polymerase of the Yeast Killer System. 2011.  
David Forrest, Nikolay Zenkin
- Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. & Darst, S. A. 2002a. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science*, 296, 1285-90.
- Murakami, K. S., Masuda, S. & Darst, S. A. 2002b. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science*, 296, 1280-4.
- Mustaev, A., Kozlov, M., Markovtsov, V., Zaychikov, E., Denissova, L. & Goldfarb, A. 1997. Modular organization of the catalytic center of RNA polymerase. *Proc Natl Acad Sci U S A*, 94, 6641-5.
- Nechaev, S., Chlenov, M. & Severinov, K. 2000. Dissection of two hallmarks of the open promoter complex by mutation in an RNA polymerase core subunit. *J Biol Chem*, 275, 25516-22.
- Nemeth, A., Perez-Fernandez, J., Merkl, P., Hamperl, S., Gerber, J., Griesenbeck, J. & Tschochner, H. 2013. RNA polymerase I termination: Where is the end? *Biochim Biophys Acta*, 1829, 306-17.

Nguyen, H.M. & Kang, C. 2014. Lysis delay and burst shrinkage of coliphage T7 by deletion of terminator Tphi reversed by deletion of early genes. *J Virol*, 88, 2107-15.

Nicolas, P., Mader, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeke, M., Aymerich, S., Becher, D., Bisicchia, P., Botella, E., Delumeau, O., Doherty, G., Denham, E. L., Fogg, M. J., Fromion, V., Goelzer, A., Hansen, A., Hartig, E., Harwood, C. R., Homuth, G., Jarmer, H., Jules, M., Klipp, E., Le Chat, L., Lecointe, F., Lewis, P., Liebermeister, W., March, A., Mars, R. A., Nannapaneni, P., Noone, D., Pohl, S., Rinn, B., Rugheimer, F., Sappa, P. K., Samson, F., Schaffer, M., Schwikowski, B., Steil, L., Stulke, J., Wiegert, T., Devine, K. M., Wilkinson, A. J., Van Dijl, J. M., Hecker, M., Volker, U., Bessieres, P. & Noirot, P. 2012. Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science*, 335, 1103-6.

Nielsen, S., Yuzenkova, Y. & Zenkin, N. 2013. Mechanism of eukaryotic RNA polymerase III transcription termination. *Science*, 340, 1577-80.

Nielsen, S. & Zenkin, N. 2013. Transcript assisted phosphodiester bond hydrolysis by eukaryotic RNA polymerase II. *Transcription*, 4, 209-12.

Nudler, E., Mustaev, A., Lukhtanov, E. & Goldfarb, A. 1997. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell*, 89, 33-41.

Oguma, K., Iida, H., Shiozaki, M. & Inoue, K. 1976. Antigenicity of converting phages obtained from *Clostridium botulinum* types C and D. *Infect Immun*, 13, 855-60.

Orlova, M., Newlands, J., Das, A., Goldfarb, A. & Borukhov, S. 1995. Intrinsic transcript cleavage activity of RNA polymerase. *Proc Natl Acad Sci U S A*, 92, 4596-600.

Paik, S. H., Chakicherla, A. & Hansen, J. N. 1998. Identification and characterization of the structural and transporter genes for, and the chemical and biological properties of, sublancin 168, a novel lantibiotic produced by *Bacillus subtilis* 168. *J Biol Chem*, 273, 23134-42.

Panaghie, G., Aiyar, S. E., Bobb, K. L., Hayward, R. S. & De Haseth, P. L. 2000. Aromatic amino acids in region 2.3 of *Escherichia coli* sigma 70 participate collectively in the formation of an RNA polymerase-promoter open complex. *J Mol Biol*, 299, 1217-30.

Peters, J. M., Mooney, R. A., Kuan, P. F., Rowland, J. L., Keles, S. & Landick, R. 2009. Rho directs widespread termination of intragenic and stable RNA transcription. *Proc Natl Acad Sci U S A*, 106, 15406-11.

Pupov, D., Kuzin, I., Bass, I. & Kulbachinskiy, A. 2014. Distinct functions of the RNA polymerase sigma subunit region 3.2 in RNA priming and promoter escape. *Nucleic Acids Res*, 42, 4494-504.

- Ream, T. S., Haag, J. R., Wierzbicki, A. T., Nicora, C. D., Norbeck, A. D., Zhu, J. K., Hagen, G., Guilfoyle, T. J., Pasa-Tolic, L. & Pikaard, C. S. 2009. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell*, 33, 192-203.
- Regamey, A. & Karamata, D. 1998. The N-acetylmuramoyl-L-alanine amidase encoded by the *Bacillus subtilis* 168 prophage SP beta. *Microbiology*, 144 ( Pt 4), 885-93.
- Reines, D. 1992. Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II. *J Biol Chem*, 267, 3795-800.
- Roberts, M. S., Nakamura, L. K. & Cohan, F. M. 1996. *Bacillus vallismortis* sp. nov., a close relative of *Bacillus subtilis*, isolated from soil in Death Valley, California. *Int J Syst Bacteriol*, 46, 470-5.
- Roghanian, M., Yuzenkova, Y. & Zenkin, N. 2011. Controlled interplay between trigger loop and Gre factor in the RNA polymerase active centre. *Nucleic Acids Res*, 39, 4352-9.
- Romanos, M. A. & Boyd, A. 1988. A transcriptional barrier to expression of cloned toxin genes of the linear plasmid k1 of *Kluyveromyces lactis*: evidence that native k1 has novel promoters. *Nucleic Acids Res*, 16, 7333-50.
- Rosenthal, R., Toye, P. A., Korman, R. Z. & Zahler, S. A. 1979. The prophage of SP beta c2dcitK1, A defective specialized transducing phage of *Bacillus subtilis*. *Genetics*, 92, 721-39.
- Ruprich-Robert, G. & Thuriaux, P. 2010. Non-canonical DNA transcription enzymes and the conservation of two-barrel RNA polymerases. *Nucleic Acids Res*, 38, 4559-69.
- Sainsbury, S., Niesser, J. & Cramer, P. 2013. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature*, 493, 437-40.
- Salgado, P. S., Koivunen, M. R., Makeyev, E. V., Bamford, D. H., Stuart, D. I. & Grimes, J. M. 2006. The structure of an RNAi polymerase links RNA silencing and transcription. *PLoS Biol*, 4, e434.
- Sanchez-Weatherby, J., Bowler, M. W., Huet, J., Gobbo, A., Felisaz, F., Lavault, B., Moya, R., Kadlec, J., Ravelli, R. B. & Cipriani, F. 2009. Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta Crystallogr D Biol Crystallogr*, 65, 1237-46.
- Santangelo, T. J. & Reeve, J. N. 2006. Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. *J Mol Biol*, 355, 196-210.
- Sato, T., Samori, Y. & Kobayashi, Y. 1990. The *cisA* cistron of *Bacillus subtilis* sporulation gene *spoIVC* encodes a protein homologous to a site-specific recombinase. *J Bacteriol*, 172, 1092-8.

- Schaffrath, R. & Breunig, K. D. 2000. Genetics and molecular physiology of the yeast *Kluyveromyces lactis*. *Fungal Genet Biol*, 30, 173-90.
- Schaffrath, R. & Meacock, P. A. 1995. *Kluyveromyces lactis* killer plasmid pGKL2: molecular analysis of an essential gene, ORF5. *Yeast*, 11, 615-28.
- Schaffrath, R. & Meacock, P. A. 1996. A cytoplasmic gene-shuffle system in *Kluyveromyces lactis*: use of epitope tagging to detect a killer plasmid-encoded gene product. *Mol Microbiol*, 19, 545-54.
- Schaffrath, R. & Meacock, P. A. 2001. An SSB encoded by and operating on linear killer plasmids from *Kluyveromyces lactis*. *Yeast*, 18, 1239-47.
- Schaffrath, R., Meinhardt, F. & Meacock, P. A. 1997. ORF7 of yeast plasmid pGKL2: analysis of gene expression in vivo. *Curr Genet*, 31, 190-2.
- Schaffrath, R., Sasnauskas, K. & Meacock, P. A. 2000. Use of gene shuffles to study the cytoplasmic transcription system operating on *Kluyveromyces lactis* linear DNA plasmids. *Enzyme Microb Technol*, 26, 664-670.
- Schaffrath, R., Soond, S. M. & Meacock, P. A. 1995a. Cytoplasmic gene expression in yeast: a plasmid-encoded transcription system in *Kluyveromyces lactis*. *Biochem Soc Trans*, 23, 128S.
- Schaffrath, R., Soond, S. M. & Meacock, P. A. 1995b. The DNA and RNA polymerase genes of yeast plasmid pGKL2 are essential loci for plasmid integrity and maintenance. *Microbiology*, 141 ( Pt 10), 2591-9.
- Schaffrath, R., Stark, M. J., Gunge, N. & Meinhardt, F. 1992. *Kluyveromyces lactis* killer system: ORF1 of pGKL2 has no function in immunity expression and is dispensable for killer plasmid replication and maintenance. *Curr Genet*, 21, 357-63.
- Schick, B. & Journé, F. 1994. Extension of the diffraction resolution of crystals. *Acta Crystallographica Section D*, 50, 563-568.
- Schickel, J., Helmig, C. & Meinhardt, F. 1996. *Kluyveromyces lactis* killer system: analysis of cytoplasmic promoters of the linear plasmids. *Nucleic Acids Res*, 24, 1879-86.
- Seaman, E., Tarmy, E. & Marmur, J. 1964. Inducible Phages of *Bacillus subtilis*. *Biochemistry*, 3, 607-13.
- Sharma, C. M., Hoffman, S., Darfeuille, F., Reignier, J., Findeiß, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., Stadler, P., Vogel, J. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*, 464, 250-256.
- Sidorenkov, I., Komissarova, N. & Kashlev, M. 1998. Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Mol Cell*, 2, 55-64.

- Smyth, D. R., Mrozkiewicz, M. K., Mcgrath, W. J., Listwan, P. & Kobe, B. 2003. Crystal structures of fusion proteins with large-affinity tags. *Protein Sci*, 12, 1313-22.
- Sor, F. & Fukuhara, H. 1985. Structure of a linear plasmid of the yeast *Kluyveromyces lactis*; Compact organization of the killer genome. *Current Genetics*, 9, 147-155.
- Sosunov, V., Sosunova, E., Mustaev, A., Bass, I., Nikiforov, V. & Goldfarb, A. 2003. Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J*, 22, 2234-44.
- 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 Res*, 33, 4202-11.
- 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. *Proc Natl Acad Sci U S A*, 100, 15469-74.
- Sousa, R., Chung, Y. J., Rose, J. P. & Wang, B. C. 1993. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature*, 364, 593-9.
- Stam, J. C., Kwakman, J., Meijer, M. & Stuitje, A. R. 1986. Efficient isolation of the linear DNA killer plasmid of *Kluyveromyces lactis*: evidence for location and expression in the cytoplasm and characterization of their terminally bound proteins. *Nucleic Acids Res*, 14, 6871-84.
- Stano, N. M., Levin, M. K. & Patel, S. S. 2002. The +2 NTP binding drives open complex formation in T7 RNA polymerase. *J Biol Chem*, 277, 37292-300.
- Stark, M. J. & Boyd, A. 1986. The killer toxin of *Kluyveromyces lactis*: characterization of the toxin subunits and identification of the genes which encode them. *EMBO J*, 5, 1995-2002.
- Stark, M. J., Boyd, A., Mileham, A. J. & Romanos, M. A. 1990. The plasmid-encoded killer system of *Kluyveromyces lactis*: a review. *Yeast*, 6, 1-29.
- Stark, M. J., Mileham, A. J., Romanos, M. A. & Boyd, A. 1984. Nucleotide sequence and transcription analysis of a linear DNA plasmid associated with the killer character of the yeast *Kluyveromyces lactis*. *Nucleic Acids Res*, 12, 6011-30.
- Steitz, T. A. 1998. A mechanism for all polymerases. *Nature*, 391, 231-2.
- Steitz, T. A. & Steitz, J. A. 1993. A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A*, 90, 6498-502.
- Stragier, P., Kunkel, B., Kroos, L. & Losick, R. 1989. Chromosomal rearrangement generating a composite gene for a developmental transcription factor. *Science*, 243, 507-12.

- Sudiarta, I. P., Fukushima, T. & Sekiguchi, J. 2010. *Bacillus subtilis* CwIP of the SP- $\beta$  prophage has two novel peptidoglycan hydrolase domains, muramidase and cross-linkage digesting DD-endopeptidase. *J Biol Chem*, 285, 41232-43.
- Tahirov, T. H., Temiakov, D., Anikin, M., Patlan, V., Mcallister, W. T., Vassilyev, D. G. & Yokoyama, S. 2002. Structure of a T7 RNAP polymerase elongation complex at 2.9A resolution. *Nature*, 420, 43-50.
- Takeda, M., Hiraishi, H., Takesako, T., Tanase, S. & Gunge, N. 1996. The terminal protein of the linear DNA plasmid pGKL2 shares an N-terminal domain of the plasmid-encoded DNA polymerase. *Yeast*, 12, 241-6.
- Takemaru, K., Mizuno, M., Sato, T., Takeuchi, M. & Kobayashi, Y. 1995. Complete nucleotide sequence of a skin element excised by DNA rearrangement during sporulation in *Bacillus subtilis*. *Microbiology*, 141 ( Pt 2), 323-7.
- Tanguy-Rougeau, C., Chen, X. J., Wesolowski-Louvel, M. & Fukuhara, H. 1990. Expression of a foreign KmR gene in linear killer DNA plasmids in yeast. *Gene*, 91, 43-50.
- Temiakov, D., Anikin, M. & Mcallister, W. T. 2002. Characterization of T7 RNA polymerase transcription complexes assembled on nucleic acid scaffolds. *J Biol Chem*, 277, 47035-43.
- Temiakov, D., Zenkin, N., Vassilyeva, M. N., Perederina, A., Tahirov, T. H., Kashkina, E., Savkina, M., Zorov, S., Nikiforov, V., Igarashi, N., Matsugaki, N., Wakatsuki, S., Severinov, K. & Vassilyev, D. G. 2005. Structural Basis of Transcription Inhibition by Antibiotic Streptolydigin. *Molecular Cell*, 19, 655-666.
- Tiggemann, M., Jeske, S., Larsen, M. & Meinhardt, F. 2001. *Kluyveromyces lactis* cytoplasmic plasmid pGKL2: heterologous expression of Orf3p and proof of guanylyltransferase and mRNA-triphosphatase activities. *Yeast*, 18, 815-25.
- Tokunaga, M., Wada, N. & Hishinuma, F. 1987. Expression and identification of immunity determinants on linear DNA killer plasmids pGKL1 and pGKL2 in *Kluyveromyces lactis*. *Nucleic Acids Res*, 15, 1031-46.
- Tommasino, M., Ricci, S. & Galeotti, C. L. 1988. Genome organization of the killer plasmid pGK12 from *Kluyveromyces lactis*. *Nucleic Acids Res*, 16, 5863-78.
- Tovar-Rojo, F. & Setlow, P. 1991. Effects of mutant small, acid-soluble spore proteins from *Bacillus subtilis* on DNA in vivo and in vitro. *J Bacteriol*, 173, 4827-35.
- Vannini, A. & Cramer, P. 2012. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell*, 45, 439-46.
- Vassilyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassilyeva, M. N., Borukhov, S. & Yokoyama, S. 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature*, 417, 712-9.

- 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, 157-62.
- 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, 163-8.
- Villemain, J., Guajardo, R. & Sousa, R. 1997. Role of open complex instability in kinetic promoter selection by bacteriophage T7 RNA polymerase. *J Mol Biol*, 273, 958-77.
- Walter, T. S., Meier, C., Assenberg, R., Au, K.-F., Ren, J., Verma, A., Nettleship, Joanne e., Owens, R. J., Stuart, David i. & Grimes, J. M. 2006. Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. *Structure*, 14, 1617-1622.
- 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, 941-54.
- Wang, I. N., Smith, D. L. & Young, R. 2000. Holins: the protein clocks of bacteriophage infections. *Annu Rev Microbiol*, 54, 799-825.
- Warner, F., Kitos, G., Romano, M. & Hemphill, E. 1977. Characterization of SP $\beta$ : a temperate bacteriophage from *Bacillus subtilis* 168M. *Canadian Journal of Microbiology*, 21 - 25, 45 - 51.
- Wassarman, K. M. & Saecker, R. M. 2006. Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science*, 314, 1601-3.
- Werner, F. 2008. Structural evolution of multisubunit RNA polymerases. *Trends Microbiol*, 16, 247-50.
- Werner, F. & Weinzierl, R. O. 2005. Direct modulation of RNA polymerase core functions by basal transcription factors. *Mol Cell Biol*, 25, 8344-55.
- Wernimont, A. & Edwards, A. 2009. In situ proteolysis to generate crystals for structure determination: an update. *PLoS One*, 4, e5094.
- Wesolowski, M., Algeri, A., Goffrini, P. & Fukuhara, H. 1982. Killer DNA Plasmids of the Yeast *Kluyveromyces lactis* : I. Mutations Affecting the Killer Phenotype. *Curr Genet*, 5, 191-7.
- Westers, H., Dorenbos, R., Van Dijl, J. M., Kabel, J., Flanagan, T., Devine, K. M., Jude, F., Seror, S. J., Beekman, A. C., Darmon, E., Eschevins, C., De Jong, A., Bron, S., Kuipers, O. P., Albertini, A. M., Antelmann, H., Hecker, M., Zamboni, N., Sauer, U., Bruand, C., Ehrlich, D. S., Alonso, J. C., Salas, M. & Quax, W. J. 2003. Genome engineering reveals large dispensable regions in *Bacillus subtilis*. *Mol Biol Evol*, 20, 2076-90.

- 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, 481-9.
- Wheeler, M. J., Russi, S., Bowler, M. G. & Bowler, M. W. 2012. Measurement of the equilibrium relative humidity for common precipitant concentrations: facilitating controlled dehydration experiments. *Acta Crystallographica Section F*, 68, 111-114.
- Wilson, D. W. & Meacock, P. A. 1988. Extranuclear gene expression in yeast: evidence for a plasmid-encoded RNA polymerase of unique structure. *Nucleic Acids Res*, 16, 8097-112.
- Wood, H. E., Dawson, M. T., Devine, K. M. & McConnell, D. J. 1990. Characterization of PBSX, a defective prophage of *Bacillus subtilis*. *J Bacteriol*, 172, 2667-74.
- Yin, Y. W. & Steitz, T. A. 2002. Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science*, 298, 1387-95.
- Yuzenkova, Y., Bochkareva, A., Tadigotla, V. R., Roghanian, M., Zorov, S., Severinov, K. & Zenkin, N. 2010. Stepwise mechanism for transcription fidelity. *BMC Biol*, 8, 54.
- Yuzenkova, Y., Tadigotla, V. R., Severinov, K. & Zenkin, N. 2011. A new basal promoter element recognized by RNA polymerase core enzyme. *EMBO J*, 30, 3766-75.
- Yuzenkova, Y. & Zenkin, N. 2010. Central role of the RNA polymerase trigger loop in intrinsic RNA hydrolysis. *Proc Natl Acad Sci U S A*, 107, 10878-83.
- Zahler, S. A., Korman, R. Z., Rosenthal, R. & Hemphill, H. E. 1977. *Bacillus subtilis* bacteriophage SPbeta: localization of the prophage attachment site, and specialized transduction. *J Bacteriol*, 129, 556-8.
- 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, 107-9.
- Zenkin, N. 2014a. Ancient RNA stems that terminate transcription. *RNA Biol*, 11, 295-7.
- Zenkin, N. 2014b. RNA secondary structure-dependent termination of transcription. *Cell Cycle*, 13, 3-4.
- Zenkin, N., Naryshkina, T., Kuznedelov, K. & Severinov, K. 2006a. The mechanism of DNA replication primer synthesis by RNA polymerase. *Nature*, 439, 617-20.
- Zenkin, N. & Severinov, K. 2004. The role of RNA polymerase sigma subunit in promoter-independent initiation of transcription. *Proc Natl Acad Sci U S A*, 101, 4396-400.
- Zenkin, N., Yuzenkova, Y. & Severinov, K. 2006b. Transcript-assisted transcriptional proofreading. *Science*, 313, 518-20.

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, 811-24.

Zhang, J., Palangat, M. & Landick, R. 2010. Role of the RNA polymerase trigger loop in catalysis and pausing. *Nat Struct Mol Biol*, 17, 99-104.