STUDIES ON TRANSCRIPTION IN *ESCHERICHIA COLI*

BY

BANDAR SENDY

A THESIS

SUBMITTED TO THE UNIVERSITY OF BIRMINGHAM FOR

THE DEGREE OF

DOCTOR OF PHILOSOPHY

INSTITUTE OF MICROBIOLOGYAND INFECTION
SCHOOL OF BIOSCIENCES
UNIVERSITY OF BIRMINGHAM
MAY 2017

Abstract

In bacteria, the expression of genes is tightly controlled, predominantly at the point of transcription. This gives the bacterium the ability to adapt to changing environments. RNA polymerase (RNAP) is one of the key players in transcription and must first bind to a deoxyribonucleic acid (DNA) promoter upstream of a gene to transcribe it. However, the ability of RNAP binding is dictated by the core promoter DNA sequence, the presence of transcription activator or repressor proteins and numerous other factors that can influence the level of gene expression. The strength of promoters has been indirectly measured by estimating the fold change in the expression of transcripts or gene products. Only a few studies have attempted to directly address the RNAP flux through transcription units, and further studies are still required.

In the current study, the aim was to directly correlate RNAP gene transcription with the strength of core promoter elements. To measure the RNAP binding to any DNA segment in *Escherichia coli in vivo,* I employed the direct method of chromatin immunoprecipitation (ChIP), followed by quantification of immunoprecipitated DNA. For promoter regions, this method directly measures the occupancy by RNAP; for regions within transcription units, the flux of the RNAP was deduced. This method was validated using a hierarchy of semi-synthetic promoters inserted in a pRW50 plasmid and a chromosomal transcription unit (i.e., the *lac* operon).

A range of semisynthetic promoters, with different combinations of core promoter elements to obtain different levels of expression, was used in this study. This direct method enabled the calculation of “promoter competitivity”, “promoter occupancy index” (POI), RNAP “escape index” (EI), “fragment occupancy percentage” (FOP) and the

time interval between transcribing RNAPs (T_{int}). On the basis of time intervals between RNAPs during transcription, I have been able to calculate the number of RNAPs crossing any DNA sequence of interest per second (polymerase per second; PoPS). Surprisingly, the results of the present study revealed that the RNAPs are well separated during transcription of the *lac* operon.

This study has increased our knowledge about RNAP recruitment, escape and flux through a transcription unit. The results of this study have implications for synthetic biology systems, where precise amounts of gene expression can be tailored to fit a particular application.

I dedicate this document to my Mom, Dad and my beloved Wife

Acknowledgements

This thesis is the result of four years of work, and I have done all the work unless otherwise mentioned. During all these years, I have been working with many people and want to thank them for sharing their knowledge, showing technical skill and equipment operation.

First, I want to thank my wife, Shimaa Nagro for her support, patience and understanding during my thesis writing and my sons, Khaled and Omar for their patience in waiting for finishing my thesis writing.

I want to thank my supervisor, Prof. Steve Busby for accepting me as a PhD student in his lab and for his relentless support and expert guidance to enable me to understand the project. He has been very kind, friendly and encouraging to me. I would also like to thank my second supervisor Dr. David Lee, who showed me most of the bench techniques I used in my project and his guidance during different phases of my PhD.

I am grateful to Dr. Jack Bryant for helping me in optimizing the ChIP experiment and being very good colleague and friend. I want to say thanks to all the members of Busby's group present and past, Dr. Douglas Browning, Dr. Laura Sellers, Yasir, Gaddah, and Karen for making the lab very comfortable and fun place to work. Special thanks goes to Rita Godfrey, for her support, help and keeping the supply of all useable items in the lab. I want to say thanks to Ian Cadby, Basema and Anmol for their support.

Many thanks go to the very kind team at the Functional Genomics and Proteomics Facility at the University of Birmingham for all their assistance. I am very grateful to the Ministry of Higher Education and King Abdulaziz City for Science and Technology (KACST) for sponsoring me for this scholarship.

Finally, I thank my family, my parents, my mother in law and sisters for supporting me during all this time, when I was staying away from them. My PhD would not be the same, if I did not have supportive and loving wife and understanding sons.

TABLE OF CONTENT

ABSTRACT II

ACKNOWLEDGEMENTS V

TABLE OF CONTENT VI

LIST OF FIGURES XI

LIST OF TABLES XIII

LIST OF ABBREVIATIONS XVI

CHAPTER 1 1

1.1. Deoxyribonucleic acid 2

1.2. Bacteria 4

1.3. Gene expression 6

1.3.1. DNA-dependent RNA polymerase 6

1.3.2. Promoter elements 8

1.4. The transcription cycle 9

1.4.1. Transcription initiation 13

1.4.2. Transcription elongation 15

1.4.3. Transcription termination 21

1.5. Gene regulation in bacteria 24

1.5.1. Initiation regulation 26

1.5.1.1. Sigma factors 26

1.5.1.2. Transcription factors (TFs) 27

1.5.1.3. Transcription initiation regulation by altering DNA topology 32

1.5.1.4. Transcription initiation regulation by promoter modification 33

1.5.2. Elongation regulation 35

1.5.3. Termination regulation 38

1.5.4. Gene regulation by location 40

1.6. Quantitative aspects of gene expression 41

1.6.1. Effect of rifampicin on transcription 41

1.6.2. Promoter strength 42

1.6.3. Promoter strength measurements 44

CHAPTER 2 48

2.1. Suppliers 49

2.2. Bacterial strains, media, antibiotics, growth conditions and bacterial storage 49

2.2.1. Bacterial strains 49

2.2.2. Media 49

2.2.3. Antibiotics used in this study 51

2.2.4. Growth conditions 51

2.2.5. Bacterial storage (glycerol stock preparation) 52

2.3. Plasmid DNA 52

2.4. Buffers and Solutions 53

2.4.1. Solutions for general use 53

2.4.2. ChIP buffers 53

2.4.3. Reporter gene assay solutions 57

2.4.4. Competent cells preparation buffers 57

2.4.5. Gel electrophoresis 58

2.4.6. Stock acrylamide solution (250 ml) 59

2.5. Promoter manipulation 60

2.5.1. Oligonucleotides design 60

2.5.2. Conventional polymerase chain reaction 60

2.5.3. Digestion of the newly constructed DNA with restriction enzymes 62

2.5.4. Large acrylamide gel for digestion products 62

2.5.5. Electroelution 66

2.5.6. Small acrylamide gel 68

2.5.7. pRW50 EcoRI-HindIII vector preparation 68

2.5.8. Vector-Insert ligation 69

2.5.9. Transformation 71

2.6. Plasmid DNA extraction 71

2.7. Sequencing 72

2.8. Rubidium chloride-treated competent cells 72

2.9. Agarose gel electrophoresis 73

2.10. Promoters used in this study 73

2.11. Primers used 73

2.12. Reporter gene assay 75

2.13. Chromatin immunoprecipitation 79

2.13.1. Cell preparation 79

2.13.2. Crosslinking 80

2.13.3. Protein A Sepharose (beads) preparation 81

2.13.4. Immunoprecipitation 81

2.13.5. De-crosslinking 82

2.14. Quantitative Polymerase Chain Reaction 82

2.15. RNAP occupancy unit calculations 83

2.16. Calculation of promoter occupancy index 85

2.17. Calculation of promoter escape index 85

2.18. Calculation of fragment occupancy percentage 85

2.19. Calculation of polymerases per second 86

CHAPTER 3 87

3.1. Introduction 88

3.2. Method 90

3.3. Optimisation of the method 92

3.3.1. Rifampicin treatment time optimisation: 92

3.3.2. Sonication time optimisation: 93

3.4. Measurement of RNAP flux through the *lac* operon before and after induction 93

3.5. Measurement of RNAP flux through the *lac* operon before and after rifampicin treatment 94

3.6. Discussion 96

CHAPTER 4 102

4.1. Introduction 103

4.2. Reporter gene assay 106

4.3. Effect of improving extended -10 element on promoter activity 108

4.4. Effect of downgrading the extended -10 element on promoter activity 115

4.5. Effect of improving -35 element on promoter activity 119

4.6. Effect of downgrading -10 element on promoter activity 123

4.7. Effect of improving the -35 element of a promoter having a corrupted extended -10 element 131

4.8. Effect of improving -10 element on promoter activity 135

4.9. Effect of having fully consensus promoter sequences on promoter activity 138

4.10. Effect of having a consensus -10 hexamer only on promoter activity 146

4.11. Discussion 153

CHAPTER 5 162

BIBLIOGRAPHY 169

APPENDIX I 185

Sendy, B., Lee, D.J., Busby, S.J. and Bryant, J.A., 2016. RNA polymerase supply and flux through the lac operon in Escherichia coli. *Phil. Trans. R. Soc. B*, *371*(1707), p.20160080.

List of Figures

Figure 1. Promoter elements. 10

Figure 2. Promoter recognition by RNAP. 11

Figure 3. Transcription cycle steps. 12

Figure 4. Bacterial phage lambda N protein cooperation with Nus factors. 20

Figure 5. Bacterial transcription termination mechanisms. 22

Figure 6. Transcription elongation complex (TEC) pausing and termination. 25

Figure 7. The *lac* operon. 31

Figure 8. The map of pRW50 plasmid used in this study 54

Figure 9. Approximate locations of the lac0-5 probes. 91

Figure 10. RNAP flux on the *lac* operon on the chromosome before and after IPTG treatment. 95

Figure 11. RNAP flux on the *lac* operon on the chromosome before and after rifampicin treatment. 97

Figure 12. β-galactosidase measured activities for all 9 promoters used in this study. 107

Figure 13. RNAP occupancies of promoters BKS101 and BKS102. 110

Figure 14. RNAP occupancies of promoters BKS101 and BKS106. 117

Figure 15. RNAP occupancies of promoters BKS101 and BKS104. 122

Figure 16. RNAP occupancies of promoters BKS101 and BKS105. 128

Figure 17. RNAP occupancies of promoters BKS106 and BKS107. 133

Figure 18. RNAP occupancies of promoters BKS101 and BKS103. 139

Figure 19. RNAP occupancies of promoters BKS103 and BKS108. 144

Figure 20. RNAP occupancies of promoters BKS103 and BKS109. 150

List of Tables

Table1. Oligonucleotides used to design the new promoters. 61

Table2. Conventional PCR reaction components. 63

Table3. Conventional PCR thermal conditions. 64

Table4. Components of each digestion reaction. 65

Table5. Components of the ligation reactions. 70

Table6. Sequences of the semisynthetic promoters used in this study. 74

Table7. Probes used in qPCR experiment to measure RNAP flux on the *lac* operon. 76

Table8. Probes used in qPCR experiment to measure RNAP flux on the *lac* operon on the chromosome. 77

Table9. Cycling program used for QPCR experiments 84

Table10. Occupancy values and percentages, T_{int} and PoPS values for the *lac* operon on the chromosome before and after rifampicin treatment. 98

Table11. Starting promoter BKS101 sequence from the *Eco*R1 site to the -1 base. 105

Table12. Sequences of promoters BKS101 and BKS102. 109

Table13. β-galactosidase levels, POI and EI values of promoters BKS101 and BKS102. 111

Table14. PoPS for promoter BKS101. 113

Table15. PoPS for promoter BKS102. 114

Table16. Sequences of promoters BKS101 and BKS106. 116

Table17. β-galactosidase levels, POI and EI values of promoters BKS101 and BKS106. 118

Table18. PoPS for promoter BKS106. 120

Table19. Sequences of promoters BKS101 and BKS104. 121

Table20. β-galactosidase levels, POI and EI values of promoters BKS101 and BKS104. 124

Table21. PoPS for promoter BKS104. 125

Table22. Sequences of promoters BKS101 and BKS105. 126

Table23. β-galactosidase levels, POI and EI values of promoters BKS101 and BKS105. 129

Table24. PoPS for promoter BKS105. 130

Table25. Sequences of promoters BKS106 and BKS107. 132

Table26. β-galactosidase levels, POI and EI values of promoters BKS106 and BKS107. 134

Table27. PoPS for promoter BKS107. 136

Table28. Sequences of promoters BKS101 and BKS103. 137

Table29. β-galactosidase levels, POI and EI values of promoters BKS101 and BKS103. 140

Table30. PoPS for promoter BKS103. 141

Table31. Sequences of promoters BKS103 and BKS108. 143

Table32. β-galactosidase levels, POI and EI values of promoters BKS103 and BKS108. 145

Table33. PoPS for promoter BKS108. 147

Table34. Sequences of promoters BKS103 and BKS109. 148

Table35. β-galactosidase levels, POI and EI values of promoters BKS103 and BKS109. 151

Table36. PoPS for promoter BKS109. 152

Table37. Summary of calculated parameters for all promoters used in this study. 161

List of Abbreviations

A	adenine
AMP	Adenosine monophosphate
APS	ammonium persulphate
ATP	adenosine triphosphate
ATP – BD	ATP binding domain
bp	base – pair
C	cytosine
$CaCl_2$	calcium chloride
CAP	catabolite activator protein
ChIP	chromatin Immunoprecipitation
CIP	commonly known as CAP – calf alkaline phosphatase
CRE	core recognition element
CRP	cyclic-AMP receptor protein
Ct	cycle threshold
CTD	carboxy terminal domain
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
EB	elution buffer
EC	elongation complex
EDTA	ethylenediaminetetraacetic acid
EI	promoter escape index
EMSA	electrophoretic mobility shift assay

FOP	fragment occupancy percentage
g	gram
G	guanine
HCl	Hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF *Eco*RI	high fidelity *Eco*RI restriction enzyme
HF *Hin*dIII	high fidelity *Hin*dIII restriction enzyme
HNS	histone for nucleoid structuring protein
IHF	integration host factor
IP	immunoprecipitation
IPTG	isopropyl-β-D-l-thiogalactoside
KCl	potassium chloride
kDA	kilo Dalton
KoAc	potassium acetate
L	liter
LiCl	lithium chloride
M	molar
mg	milligram
$MgSo_4$	magnesium sulphate
ml	milliliter
mM	millimolar
$MnCl_2$	manganese (II) chloride
MOPS	3-(N-morpholino) propanesulfonic acid
mRNA	messenger RNA

Na_2CO_3	sodium carbonate
Na_2HPO_4	disodium phosphate
NaCl	sodium chloride
NaH_2PO_4	monosodium phosphate
NTD	N-terminal domain
^oC	degree Celsius
ONPG	O-nitrophenyl-β-D-galactopyranoside
PCR	polymerase chain reaction
Phenol:$CHCl_3$	phenol chloroform
POI	promoter occupancy index
PoPS	polymerase per second
qPCR	quantitative Polymerase Chain Reaction
RbCl	rubidium chloride
RNA	ribonucleic acid
RNA – BD	RNA binding domain
RNAP	ribonucleic acid polymerase
RPc	RNAP – promoter closed complex
RPo	RNAP – promoter open complex
rRNA	ribosomal RNA
rut	rho utilization site
SDS	sodium dodecyl sulfate
T	thymine
TBE	Tris/Borate/EDTA
TBS	tris-buffered saline

TEC	transcription elongation complex
TEMED	Tetramethylethylenediamine
T_{int}	time interval between RNAPs coming to the start of a particular DNA region
U	uracil
UV	ultraviolet
μg	microgram
μl	microliter
μM	mircomolar

Chapter 1

Introduction

1. Introduction

1.1. Deoxyribonucleic acid

Deoxyribonucleic acid (DNA) is the molecular component present in all living cells that carries all the genetic information required for the metabolism, growth and reproduction of organisms. Friedrich Miescher, a Swiss physician who was studying leucocyte proteins in attempt to discover the so-called "building blocks of life", first noticed and isolated this very important molecule inside cells' nuclei in 1869. He investigated the molecule and found that it had specific properties differentiating it from cellular proteins. After his discovery, he called this fundamental material "nuclein" (Dahm, 2005). Several decades later, it was shown that genes and chromosomes carrying genetic information were made up of DNA (Avery *et al.*, 1944). Years later, in 1953, Watson and Crick first proposed the double-helical structure of DNA, claiming that it consisted of two anti-parallel helical chains coiled around the same axis. Watson and Crick also published detailed information about the key structural features of DNA (Watson and Crick, 1953). This information was proven correct by subsequent research on DNA structure. DNA, as it is currently understood, is a double-stranded molecule in which two strands are intertwined to form a double-helix. Each DNA strand is a polynucleotide, meaning that it composed of a chain of nucleotides linked one to another by phosphodiester bonds. Each nucleotide contains 3 constituents: a phosphate group, a five-carbon sugar and one of four possible nitrogenous bases: adenine (A), thymine (T), guanine (G) or cytosine (C). The phosphate group is attached to the 5' carbon of the sugar molecule and to the 3' carbon of the next sugar molecule. These connections are used to describe the direction of the DNA strand (i.e., 5'➔3' or 3'➔5'). Therefore,

phosphate groups and sugar molecules comprise the backbone of the DNA strand. On the inner side of the DNA strand, one of the four possible nitrogenous bases is attached to the 1' carbon of the sugar. The rationale for the "deoxyribose" name of the sugar molecule is based on a missing hydroxyl group at its 2' carbon. Adenine and guanine nitrogenous bases are known as purines and are characterized by their double-ring structure. Cytosine and thymine are categorized as pyrimidines and are characterized by their single-ring structure. These bases link the two DNA strands by binding to each other through non-covalent hydrogen bonds. Adenine in one DNA strand typically binds to thymine on the other strand via two hydrogen bonds, and guanine binds to cytosine on the opposite strand via three hydrogen bonds. As the two DNA strands are coiled around the same axis, this coiling forms two alternating spaces repeated along the DNA called major and minor grooves.

The significance of DNA is derived from its ability to store the genetic information that directs virtually all cells' activities (Avery *et al.*, 1944). These sequences are transcribed into ribonucleic acid (RNA) with the help of a specific cellular protein called RNA polymerase (RNAP). Although it is often assumed that all produced RNAs encode for proteins, most of these RNAs are non-coding and hence do not code for proteins (Mattick and Makunin, 2006). For example, in average growing *E. coli*, less than 20% of produced RNAs code for proteins (Browning and Busby, 2016). There are several types of non-coding RNAs (ncRNAs), such as, ribosomal RNA (rRNA), transfer RNA (tRNA) and small non-coding RNA (SncRNA) or small RNA (sRNA), which measure approximately 50 to 300 nucleotides in length (Storz *et al.*, 2011). However, it has been found that these non-coding RNAs are of a great importance to bacterial cells. For instance, the 6S RNA interacts with RNAP during stationary phase (Browning and Busby, 2016). These

sRNAs can also regulate messenger RNAs (mRNAs) and their translation by different mechanisms, such as base-pairing with mRNAs and protein activity modulation (Waters and Storz, 2009).

To get the full benefit of DNA, bacterial cells have evolved perfectly to express it. DNA expression is the process in which DNA is transcribed to produce different RNAs, including the coding mRNAs that are translated by cellular ribosomes to produce proteins.

1.2. Bacteria

Bacteria are unicellular microorganisms that are found everywhere on earth. A myriad of different bacterial species have been identified. These bacterial species can be classified depending on their different characteristics, such as morphology, life-style and their ecological niche (Cohan, 2001). Interestingly, many bacterial species live inside and on the surface of other organisms, including mammals such as humans. In humans, bacteria can be found on the skin surface and in the mucosa and digestive tract (Dethlefsen *et al.*, 2006). Specifically in the gastrointestinal tract, the number of bacteria is normally greater than that of human cells (Savage, 1977; Hooper and Gordon, 2001). In fact, approximately 500 to 1000 bacterial species colonize the human intestines, with bacterial genes estimated to outnumber human genes by an order of 50 to 100 (Hooper and Gordon, 2001). The presence of this bacterial flora on the inner intestinal surface is essential for human health. The bacteria are critical for the development of the immune system in the gut during the very early stages of human life (Chung *et al.*, 2012) and play a role in protecting the host against pathogenic bacteria; this protection is improved

when two or three species colonize the same area rather than one single species (Dillon *et al.*, 2005). Moreover, the composition of these bacterial species inside human gut has been found to differ greatly between obese and non-obese individuals, illustrating the relationship between host health and normal bacterial flora (DiBaise *et al.*, 2008). Some of these bacteria can essentially produce "antibiotics" against other bacteria. These antibiotics can be used medically to protect against bacterial infections (Sykes *et al.*, 1981). In addition to the advantages of normal intestinal flora in terms of protection against pathogens, the bacteria can also assist in digestion and human intestinal epithelial cell proliferation (Dethlefsen *et al.*, 2006). All of these benefits elucidate the intricate bacteria-host relationship and illustrate why it is medically relevant to study bacteria and bacterial systems in particular.

Among bacterial species, *Escherichia coli* has been extensively studied. The species was first isolated in 1885 from a child's feces by an Austrian pediatrician named Theodor Escherich who initially named it *Bacterium coli communis*; the bacterial species was eventually named after him (Welch, 2006; Daegelen *et al.*, 2009). *E. coli* has been used as a model organism since the 1950s. During that era, Francois Jacob, Jacques Monod and several other researchers used the organism to expound on the fundamentals of molecular biology by studying different aspects of bacterial molecular activities and the regulation of gene activity (Jacob and Monod, 1961). Many different *E. coli* strains have then been discovered; however, *E. coli* K-12 remains one of the most extensively studied (Riley *et al.*, 2006). This strain was first isolated in 1922 from the stool of a convalescent diphtheria patient (Daegelen *et al.*, 2009).

Although there are many *E. coli* strains that are harmless and normal commensal inhabitants of mammalian, including humans, intestines, there are a number of

pathogenic *E. coli* strains. Pathogenic strains include enteropathogenic *E. coli* (EPEC), enteroinvasive *E. coli* (EIEC), enterotoxigenic *E. coli* (ETEC), diffusely adherent *E. coli* (DAEC), enteroaggregative *E. coli* (EAEC) and enterohaemorrhagic *E. coli* (EHEC). These strains are found to cause three different diseases: urinary tract infection, sepsis/meningitis and diarrhoea (Kaper *et al.*, 2004).

The rationale for studying *E. coli* lies in the fact that the species is native to the human gastrointestinal tract and also capable of causing serious infections. Furthermore, *E. coli* are maintainable in clinical laboratories and some strains can multiply in a broad range of temperatures ranging from approximately 8°C to 48°C. In addition, the means by which basic cellular functions are achieved in *E. coli* is very similar to that of human cells with respect to the storage of genetic information and protein synthesis, however there are clearly differences, such as, complexity of RNAP and ribosomes (Berg, 2008).

1.3. Gene expression

1.3.1. DNA-dependent RNA polymerase

DNA-dependent RNA polymerase (RNAP) is the key enzyme of bacterial transcription, which is the process of copying DNA sequences into RNA sequences. RNAP is a multisubunit enzyme that consists of different subunits assembled in a specific way to allow RNAP to achieve its role in DNA transcription. In *E. coli*, RNAP can be found in two different forms: the core enzyme and the holoenzyme. The core enzyme is composed of five subunits: α2, β, β′ and ω. These subunits are highly conserved in sequence, assembly and function among prokaryotes and eukaryotes, including humans (Allison *et.al*, 1985; Archambault and Friesen, 1993). β and β′ are the largest RNAP subunits and

are structured in a way that resembles a crab claw, with the subunits assuming the form of the respective "pincers". The location of these two subunits is fundamental for their function as they surround RNAP active site cleft (Geszvain and Landick, 2005; Mukhopadhyay *et al.*, 2008). One pincer of this claw, which is represented by the β′ subunit, has the ability of moving to open and close the claw. β subunit domains 2 and 3 represent the other pincer of the claw (Cramer, 2002; Geszvain and Landick, 2005; Chakraborty *et al.*, 2012).

Each RNAP α subunit consists of two domains that play distinct role in transcription: the C-terminal domain (αCTD) and the N-terminal domain (αNTD). These two domains are connected by a linker whose flexibility permits variations in αCTD binding sites and permits diverse promoter architectures (Blatter *et al.*, 1994). The main role of the αNTD is to assemble the β and β′ subunits together, while the main function of the αCTD is to bind to the upstream promoter (UP) element, assisting in promoter recognition and binding. The ω subunit plays a role in promoting RNAP assembly and assists in β′ subunit folding (Minakhin *et al.*, 2001). In addition to these two roles, the ω subunit affects transcription regulation. RNAPs lacking an ω subunit have been found to be unresponsive to guanosine tetraphosphate (ppGpp) and guanosine pentaphosphate (pppGpp) *in vitro*. ppGpp is a global regulator during stringent response, which is a cellular response to amino acid starvation. However, RNAPs lacking an ω subunit have also been found to demonstrate stringent response *in vivo* as a result of the DksA (a transcription factor) protein effect (Vrentas *et al.*, 2005).

The RNAP core enzyme alone lacks the ability to initiate specific transcription from specific promoters. To initiate specific transcription, RNAP needs to bind to a specific transcription initiation factor, called the sigma (σ) factor, prior to transcription

initiation to form the holoenzyme. (Darst, 2001; Feklistov *et al.*, 2014). Without this binding, specific transcription is not possible, as the σ factor plays a fundamental role in promoter recognition, positioning of RNAP on DNA and unwinding of the double helix DNA at the transcription start site.

In general, RNAP has a dynamic structure. This means that all of the subunits can move slightly within the complex throughout the different transcription steps. This dynamic movement clearly explains how the β′ subunit clamps down after the detachment of the σ factor during the elongation step (Chakraborty *et al.*, 2012).

1.3.2. Promoter elements

Promoters are DNA sequences located just before the transcription start site of any transcription unit. To clearly understand how the transcription process starts, it is important to be familiar with the promoter structure and the role of each promoter element in the transcription process. Each promoter has four different regions that are mostly conserved among all bacterial species. These regions are known as the -10, extended -10, -35 and UP elements, which play a fundamental role in transcription as they are responsible for recognizing and recruiting RNAP to transcribe following genes. The -10 and -35 elements are the most important promoter elements in transcription. The -10 hexamer is located 10 base pairs (bp) upstream of the transcription start site and normally starts from base -12 to base -7, with the consensus sequence of 5' TATAAT 3' with three highly conserved bases (T_{-12} A_{-11} and T_{-7}) (Shultzaberger *et al.*, 2007). The -35 hexamer is located 35 bp in the same direction of the -10 hexamer and starts from base -30 to base -35 with a consensus sequence of 5' TTGACA 3'. The

extended -10 element starts from base -14 to base -17 and the optimal sequence is TRTG (where R is A or G). The UP element is located upstream of the -35 element in an AT-rich region participating in recognizing a specific unit of RNAP. These elements are found to control promoter affinity to RNAP, and when their sequences are in consensus, promoter recognition is optimal (Phan *et al.*, 2012) (Figure 1).

In addition to these promoter elements, there are two other elements both located downstream of the -10 element that play a significant role in the transcription initiation process. These elements are the discriminator region and the core recognition element (CRE). The discriminator region is located directly downstream of the -10 element from base -6 to base -4, and the CRE element is located downstream of the discriminator region from base -4 to base +2 (Feklistov, 2013).

With respect to recognition by RNAP, the -10, extended -10 and -35 elements are recognized by three domains of the RNAP σ factor, whereas the UP element is recognized by the αCTD domain of RNAP (Figure 2) (Browning and Busby, 2004).

1.4. The transcription cycle

DNA transcription is the first step of gene expression, where RNAP transcribes a target DNA fragment producing a complementary chain of RNA (Suzuki, 2015). One type of these newly synthesized RNAs is mRNA, which encodes for protein synthesis by bacterial ribosomes (Cukras and Green, 2005). There are three main steps in the bacterial transcription cycle: transcription initiation, elongation and termination (Figure 3). Each step can be divided into several additional steps.

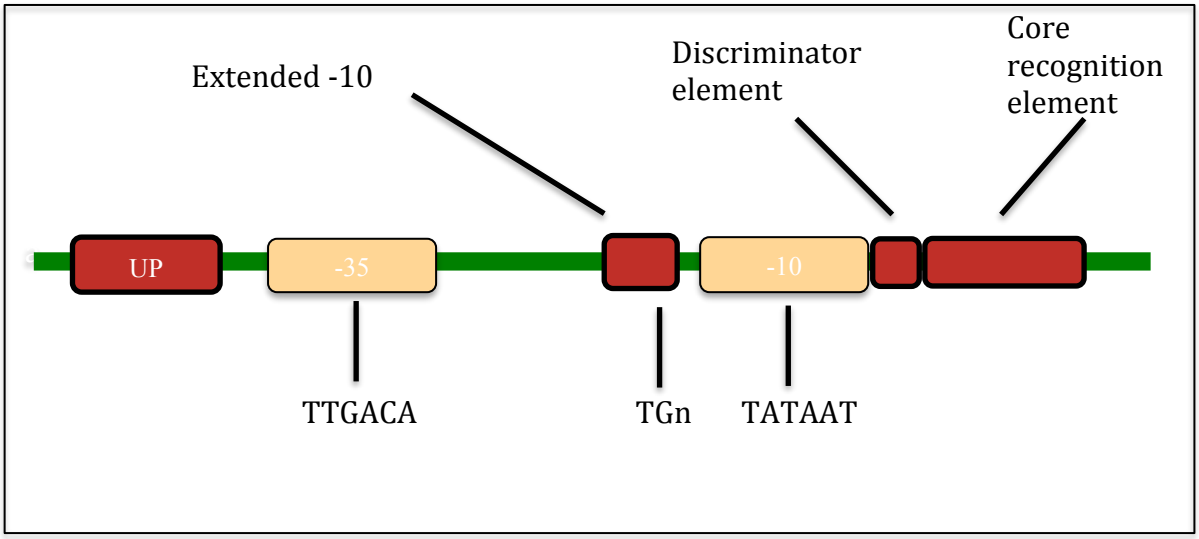

Figure 1. Promoter elements.

This figure illustrates the main promoter elements and their consensus sequences. These elements play an important role in RNA polymerase binding.

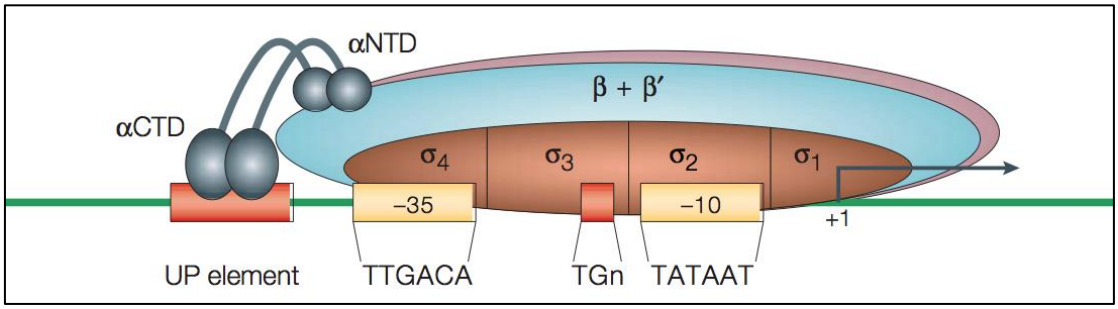

Figure 2. Promoter recognition by RNAP.

This figure shows the different promoter elements and their recognition by RNA polymerase subunits. Adapted from (Browning and Busby, 2004).

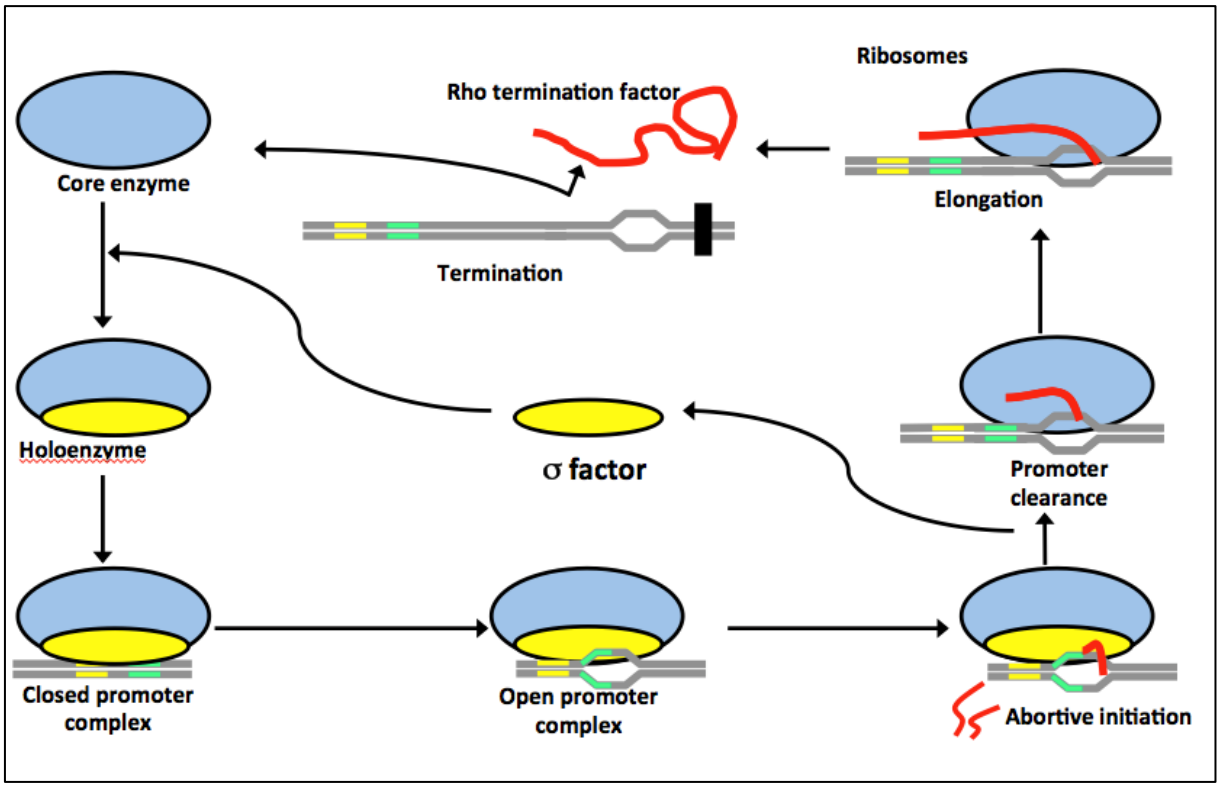

Figure 3. Transcription cycle steps.

This figure shows the three main steps of the bacterial transcription cycle: transcription initiation, elongation and termination.

1.4.1. Transcription initiation

The transcription initiation step is composed of three different steps: promoter recognition and open complex (RPo) formation, transcript initiation and scrunching; and promoter clearance. Prior to the promoter recognition step, core RNAP must bind to the σ factor to form the holoenzyme.

In *E. coli*, most σ factors belong to the housekeeping σ^{70} family (Paget and Helmann, 2003; Feklistov *et al.*, 2014). σ^{70} factor is known as a housekeeping factor because it is responsible for the transcription of most of the bacterial genes during exponential growth of the bacteria (Feklistov *et al.*, 2014). Approximately 1643 binding sites for σ^{70} factor alone have been recognized in the entire *E. coli* genome (Cho *et al.*, 2014). However, σ^{70} is not the only σ factor in *E. coli*. The *E. coli* genome encodes 6 other alternative σ factors to activate transcription of specific genes when the bacterial cell is under stress or growing in strict environmental conditions (Paget and Helmann, 2003; Gruber and Gross, 2003; Wade *et al.*, 2006). One of these alternative σ factors is σ^{54}, which is responsible for directing RNAPs to genes involved in nitrogen starvation and other functions (Brown *et al.*, 2014). There are many other σ^{54}-dependent genes. Many of the final products of these genes frequently participate in cellular nitrogen adjustment (Reitzer and Schneider, 2001). Unlike σ^{70}, σ^{54} recognizes promoters having -12 and -24 instead of -10 and -35 conserved regions. In addition, σ^{54} does not have domain 2, which induces the open complex formation and then positions the template DNA strand in RNAP active site (Cook and Ussery, 2013). Thus, σ^{54} needs a transcription factor called an enhancer binding protein for transcription to go forward. This alternative σ factor was reported to have around 180 binding sites within the whole *E. coli* genome (Cho *et al.*, 2014).

The σ factors, related to the σ^{70} family, have four different conserved domains (1, 2, 3 and 4). Each conserved domain has conserved sub-domains. Domain 1 has two sub-domains (1.1 and 1.2), domain 2 has four sub-domains (2.1, 2.2, 2.3 and 2.4), domain 3 has three sub-domains (3.0, 3.1 and 3.2), and domain 4 has two sub-domains (4.1 and 4.2). Among these four domains, domains 1 and 2 has the greatest interaction with the coiled coil motif of the β' subunit of core RNAP. The other two domains have only a few polar interactions with core RNAP (Vassylyev *et al.*, 2002; Feklistov and Darst, 2011).

As σ^{70} is the recognition subunit of the holoenzyme, it needs to contact specific promoter elements on the DNA to lead RNAP to the target promoter. The 4.2 conserved region of domain 4 of σ^{70} factor recognizes the conserved -35 promoter element. The σ^{70} sub-domains 2.1, 2.3 – 2.4 and 2.4 – 3.0 are inserted into the DNA major groove and assist in the recognition of the -10 element. The promoter-extended -10 element is recognized by domain 3 of the σ^{70} subunit (Fenton *et al.*, 2000; Vassylyev *et al.*, 2002; Busby and Savery, 2007; Taliaferro *et al.*, 2012). As mentioned previously, before making contact with promoters, σ^{70} must first dock to core RNAP. It has been determined that σ^{70} factor initially contacts three specific regions of the core RNAP β and β' subunits. These contact regions are the β flap region, the β' coiled coil region and a β region located between amino acids 1060 and 1240. These three regions contact the σ^{70} sub-domains 1.1, 2.1 – 2.2 and 3.1, respectively (Burgess and Anthony, 2001).

It has been shown that UP, -35 and extended -10 elements are initially recognized when the DNA is still double stranded (dsDNA). This recognition leads to the formation of the closed complex (RPc). Afterwards, Upstream DNA is wrapped on RNAP and downstream bases -10 to +20 are inserted into the RNAP active site (Ruff *et al.*, 2015). Then, RNAP in

association with σ factor unwinds around 12 to 13 bp of the DNA to form what is known as the "transcription bubble" or "open complex", where the +1 base is positioned in the RNAP active site (Gries *et al.*, 2010; Zhang *et al.*, 2012). The -10 element and discriminator region are recognized at this point when the DNA is single-stranded (ssDNA) (Feklistov, 2013). This recognition starts when the two conserved bases of the -10 element (-11 and -7) are flipped out of the ssDNA and inserted into specific pockets in the σ subunit of RNAP (Feklistov and Darst, 2011). Then, RNAP starts the transcription process by synthesizing short non-productive transcripts until the σ factor detaches from the holoenzyme (Vassylyev *et al.*, 2002). By this time, RNAP is tightly bound to the DNA and capable of escaping the promoter and proceeding to the elongation process.

1.4.2. Transcription elongation

Transcription elongation is the next step in the transcription process. This step derives its importance from being the productive step of transcription as it is responsible for the production of RNA, the target of any successful transcription. The transcription elongation step can be primarily divided into three main stages: promoter clearance (escape), promoter-proximal pausing and constructive elongation. RNAP undergoes several conformational changes during transition from the initiation complex to the elongation complex (EC), which takes place after the synthesis of 13- to 15-nucleotide long RNA fragments, with 7 to 9 nucleotides out of these 13 to 15 nucleotides participating in an RNA/DNA hybrid. However, these conformational changes do not extremely alter the structure of the core enzyme, as all of the structures needed for RNA

synthesis already exist (Saunders *et al.*, 2006; Vassylyev *et al.*, 2002). Nevertheless, some of these structures, such as the RNA exit channel, the downstream DNA binding site and the RNA/DNA hybrid binding site are in contact with the σ factor of the holoenzyme. Thus, RNAP needs to release its σ factor and close its clamp on the DNA strand so it can engage itself in a more stable complex to carry out transcription elongation (Nudler, 1999).

After σ factor detachment from RNAP and the escape of RNAP from the promoter, RNAP, template DNA and the newly synthesized mRNA form the EC. This EC shows no particular affinity to any part of the template DNA and travels along the DNA strand in a slide-like movement (Nudler, 1999). Two previous studies analyzed transcription elongation rate *in vivo* and found that for genes producing mRNA, such as the *lacZ* gene, the average transcription elongation rate is 30 to 50 nucleotides per second (Murakawa *et al.*, 1991; Vogel and Jensen, 1997). This finding was confirmed *in vitro* using single-molecule studies on RNAP (Larson *et al.*, 2011). The implication of this speed is that the ribosomes can catch up with the elongation complex reducing the amount of naked RNA (McGary and Nudler, 2013). However, this rate is increased to approximately 80 nucleotides per second when transcribing ribosomal operons, such as the *rrn* operon, or producing ribosomal RNA (rRNA) messages that will not be translated (Vogel and Jensen, 1994; Condon *et al.*, 1995; Voulgaris *et al.*, 1999).

During transcription elongation, the EC does not move smoothly or at constant speed along the DNA. Instead, the EC may encounter several phenomena that can disturb elongation regularity. These events include transcriptional pauses, transcriptional arrests and transcript termination. In transcriptional pauses, RNAP rests for a short period of time without synthesizing RNA and then continues its process normally. In

transcriptional arrests, the same phenomenon occurs, but the enzyme cannot continue transcribing without the help of other factors. These events may occur as a result of some proteins bound to the DNA that obstruct the path of RNAP (Uptain *et al.*, 1997). Overall, in transcriptional pauses and arrests, RNAP molecules stop synthesizing RNA transcripts but do not release them and are still able to continue producing RNA. Conversely, in transcript termination, RNAP stops synthesizing RNA and releases the synthesized transcript and therefore cannot resume synthesizing RNA (Kerppola and Kane, 1991).

In addition, the EC occasionally stops transcribing and moves backwards on the DNA, changing the last added base or bases in the nascent RNA chain to ensure that there are no mistakes in the nascent RNA sequence. This mechanism is known as backtracking and is critical for transcription, as the correctness of the newly synthesized RNA sequence is vitally important for all bacterial biological processes (Voliotis *et al.*, 2009). Backtracking is characterized by two main steps: diffusion of the EC backwards on the DNA and the extrusion of the 3' end bases of the nascent RNA strand (Sahoo and Klumpp, 2013). Backtracking usually takes place after long RNAP halts. One study has suggested a role for these long pauses in backtracking by looking at similarities between the rate of ribonucleotide misincorporation and the frequency of these long pauses (Shaevitz *et al.*, 2003). During backtracking, the 3' end of the nascent RNA strand detaches from the stalled RNAP active site and the EC moves backwards. At this time, the 3' end of the RNA strand can be internally cleaved by some elongation factors, such as GreA or GreB factors, to relieve the pause. Then, the cleaved RNA fragment detaches from the EC and the EC carries on transcribing downstream DNA by adding the right bases to the 3' end of the RNA transcript (Komissarova and Kashlev, 1997; Poteete,

2011). Promoter–proximal pauses constitute a large proportion of backtracking occurrences. These pauses take place at the beginning of the elongation step when the EC is close to the promoter (Adelman and Lis, 2012). This area of the DNA is susceptible to this kind of pausing as a result of three main reasons. First, the nascent RNA strand is still short and unable to form a structure that prevents the EC from moving backwards. Second, the robust contact of the promoter elements with the initiation factors and the EC may hold the EC back. Finally, there are still no ribosomes and other ECs following the paused EC in this region; these tailing ribosomes and ECs usually prevent the backward movement by physically pushing the stalled EC (Nudler, 2012).

Transcription elongation factors, such as NusA, NusB, NusE and NusG, play a significant role in the transcription process. The NusA factor is a protein that consists of six domains with a total atomic mass of approximately 55 kDa in *E. coli* (Burmann and Rosch, 2011). These domains are composed of the following: the amino-terminal domain (NTD), S1, KH1, KH2, acidic repeat1 (AR1) and acidic repeat2 (AR2). The NTD exerts its effect by interacting with RNAP at the nascent RNA exit channel, leading to a transcriptional pause that may eventually result in termination (Ha *et al.*, 2010). The NusA domains S1, KH1 and KH2 represent the newly synthesized RNA chain binding domains (Prasch *et al.*, 2009). While the AR1 domain function is not very clear *in vitro* or *in vivo*, AR2 has been found to play two different roles in transcription. The first role is preventing the three RNA binding domains from binding to the nascent RNA strand by covering them. The second role is to interfere with the RNAP αCTD subunit and prevent it from binding to the promoter's UP element (Mah *et al.*, 2000; Schweimer *et al.*, 2011). This transcription elongation factor has been shown to negatively affect transcription elongation by increasing the energy needed for EC forward translocation (Mooney *et al.*,

2009). Nus factors have a special significance in transcription units that are reliant on a dedicated anti-terminator, such as the case for phage lambda (λ), where the λ N protein exploits the host Nus factors to stabilize the anti-termination complex and ensure full read-through as illustrated in Figure 4.

NusB is a small (15.7 kDa) and often supplementary protein that frequently acts by joining other factors to achieve its function (Burmann and Rosch, 2011). NusB consists of seven helices, α1 to α7, arranged in two subdomains: the N-terminal (α1–α3) and C-terminal (α4–α7) subdomains (Altieri *et al.*, 2000). NusB on its own has the ability to bind to an RNA motif containing anti-termination signals called *BoxA*. It can also bind to another Nus factor called NusE to form a heterodimer that have the ability to act as an anti-terminator in ribosomal RNA operons (*rrn*) containing the *BoxA* motifs (Burmann *et al.*, 2009). Furthermore, NusB can interact with the S10 protein of the 30s ribosomal subunit, forming a complex that can bind to *BoxA* on RNA strands and act as an anti-terminator body (Luo *et al.*, 2008).

NusE is slightly smaller than the NusB protein with an atomic mass of approximately 11.7 kDa. NusE is the active member of the NusE–NusB complex, and NusB guides NusE to the EC (Burmann and Rosch, 2011). The high similarity of NusE to the S10 protein, a 30S ribosomal protein, allows it to bind to NusB in a manner identical to the NusB–S10 complex and bind to the *BoxA* motif (Luo *et al.*, 2008; Burmann *et al.*, 2009).

NusG is another transcription factor that consists of two domains, the NTD and the CTD, linked to one another by a flexible linker (Burmann *et al.*, 2011). The size of these two domains and the linker is approximately 20.5 kDa (Burmann and Rosch, 2011). The NTD domain of this transcription factor believed to make a direct surface interaction with RNAP (Belogurov *et al.*, 2009).

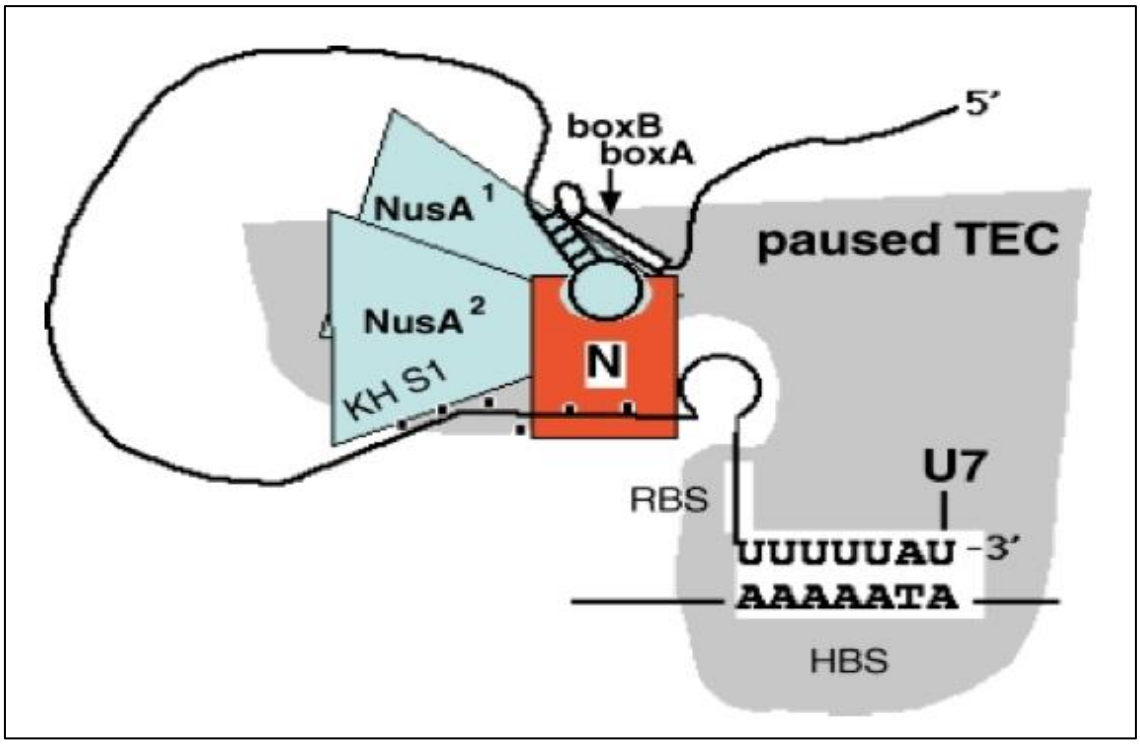

Figure 4. Bacterial phage lambda N protein cooperation with Nus factors.

This diagram illustrates how lambda N protein cooperates with the host Nus factors to ensure full-length transcript. Adapted from (Nudler and Gottesman, 2002).

NusG is one of the key protein factors during the transcriptional process due to its positive effect on RNAP. This factor reduces transcriptional pause events and increases the RNAP elongation rate by promoting its forward translocation (Herbert *et al.*, 2010; Burmann and Rosch, 2011). It was shown that this factor, assisted by the NusE factor, can couple transcription to translation. While the NTD domain of NusG interacts with RNAP, its CTD interacts with the NusE factor, which is identical to the S10 protein of the 30S ribosomal subunit. According to this similarity, NusG–CTD can interact with the S10 protein and link transcription to translation (Burmann *et al.*, 2010).

1.4.3. Transcription termination

Transcription termination is the final step of the transcription process that occurs at the end of the transcribed gene. In this step, the EC disappears due to three main modifications to the transcription bubble. These three changes can be illustrated as follows: the newly constructed mRNA dissociates from the template DNA, the transcription bubble collapses and RNAP itself detaches from the DNA. Bacterial transcription termination is carried out through one of two mechanisms: intrinsic termination and Rho-dependent termination (Figure 5).

The preferable mechanism of transcription termination, intrinsic termination, occurs when the emerging RNA forms a loop stimulated by signals encoded within the nascent RNA itself. These signals are usually represented as a guanine- and cytosine-rich region followed by a chain of sequential uracil residues. When RNAP encounters the U-rich region, it pauses transcription, giving the nascent RNA the chance to fold and forming the loop generated from the binding of the Gs and Cs in the G- and C-rich region

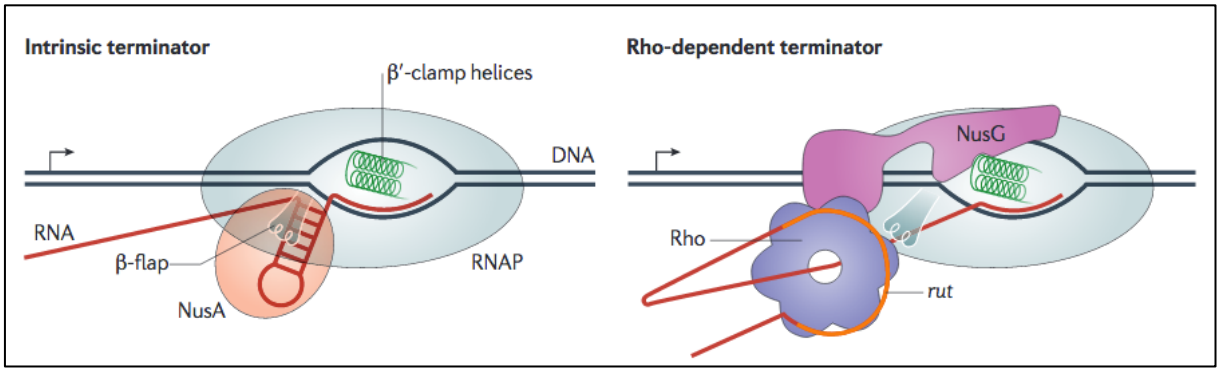

Figure 5. Bacterial transcription termination mechanisms.

This figure was adapted from (Santangelo and Artsimovitch, 2011) to illustrate the two mechanisms of bacterial transcription termination.

(Artsimovitch and Henkin, 2009; Ray *et al.*, 2016). The formation of the DNA/RNA hybrid in the U-rich region has a much lower stability than the hybrid with other normal sequences, as the bonds between the U and A residues are considerably weaker. Taking both the formation of the loop and the weak association between the DNA and RNA in the U-rich region into account, dissociation of the nascent RNA from the elongation complex is unavoidable (Martin and Tinoco Jr, 1980; Carafa *et al.*, 1990).

The second mechanism of transcription termination is known as Rho-dependent termination. The Rho protein belongs to the helicase family whose function is to unwind annealed nucleic acid strands, such as double-stranded DNA or DNA/RNA hybrids. This helicase enzyme, Rho, controls approximately 20% to nearly 50% of transcription termination processes in *E. coli* and acts by unwinding the DNA/RNA hybrid at the 5' end of the nascent RNA strand (Brennan *et al.*, 1987; Koslover *et al.*, 2012; Hollands *et al.*, 2014). Rho has six subunits that need to be organized in a hexameric structure to be capable of separating nascent RNA from template DNA (Geiselmann *et al.*, 1992). The complete functional structure of the Rho termination factor must include two important domains: the amino-terminal domain, also known as RNA-binding domain (RNA-BD), as it binds to the nascent RNA, and a domain embracing the ATP-binding site, known as the ATP-binding domain (ATP-BD) (Dolan *et al.*, 1990; Koslover *et al.*, 2012). The Rho factor is recruited by a segment of the nascent RNA that is rich in cytidine residues, which is also known as the Rho utilization (*rut*) site. Rho moves along the nascent RNA following RNAP (5'➔3'). This movement requires energy that is provided by ATP hydrolysis of the ATP-BD of the Rho factor, which is activated by the interaction between nascent RNA and the Rho factor. Once RNAP reaches the terminator, the Rho factor catches up with it and starts unwinding the RNA/DNA hybrid. After that, RNAP, the nascent RNA and the

Rho factor are all released from the template DNA (Richardson, 2002). However, competitor proteins, such as the transcription elongation factor NusA, can work as an antiterminator by inhibiting Rho-dependent termination. This factor binds to a binding site on the nascent mRNA called the N utilization (*nut*) site. The inhibition takes place when the NusA and Rho factors compete to bind at the overlapping *rut* and *nut* sites (Qayyum *et al.*, 2016). Interestingly, NusA was reported to help to achieve transcription-intrinsic termination by interacting with the hairpin structure formed in this termination mechanism (Gusarov and Nudler, 1999). In addition, this factor was found to work in favor of transcription elongation by enhancing the overall EC stability (Gusarov and Nudler, 2001).

Termination via these two mechanisms depends on RNAP pausing and disruption of the EC Figure 6. Controlling the sensitivity of RNAP to the DNA sequence and its structural elements that direct RNAP pausing can control termination at this level. Controlling pausing or escape from the pausing state using protein factor-RNAP interaction can modulate the sensitivity (Henkin, 2000).

1.5. Gene regulation in bacteria

Bacterial DNA encodes for the production of many cellular proteins through transcription and translation processes; however, these proteins do not need to be produced all the time. As some of these proteins need to be produced in higher or lower quantities than usual in specific bacterial life stages or to adapt to certain environmental conditions, bacteria need to be able to control the expression of genes responsible for the production of these proteins. During gene expression, several targets are available for regulating bacterial protein production. As a step of the gene expression process,

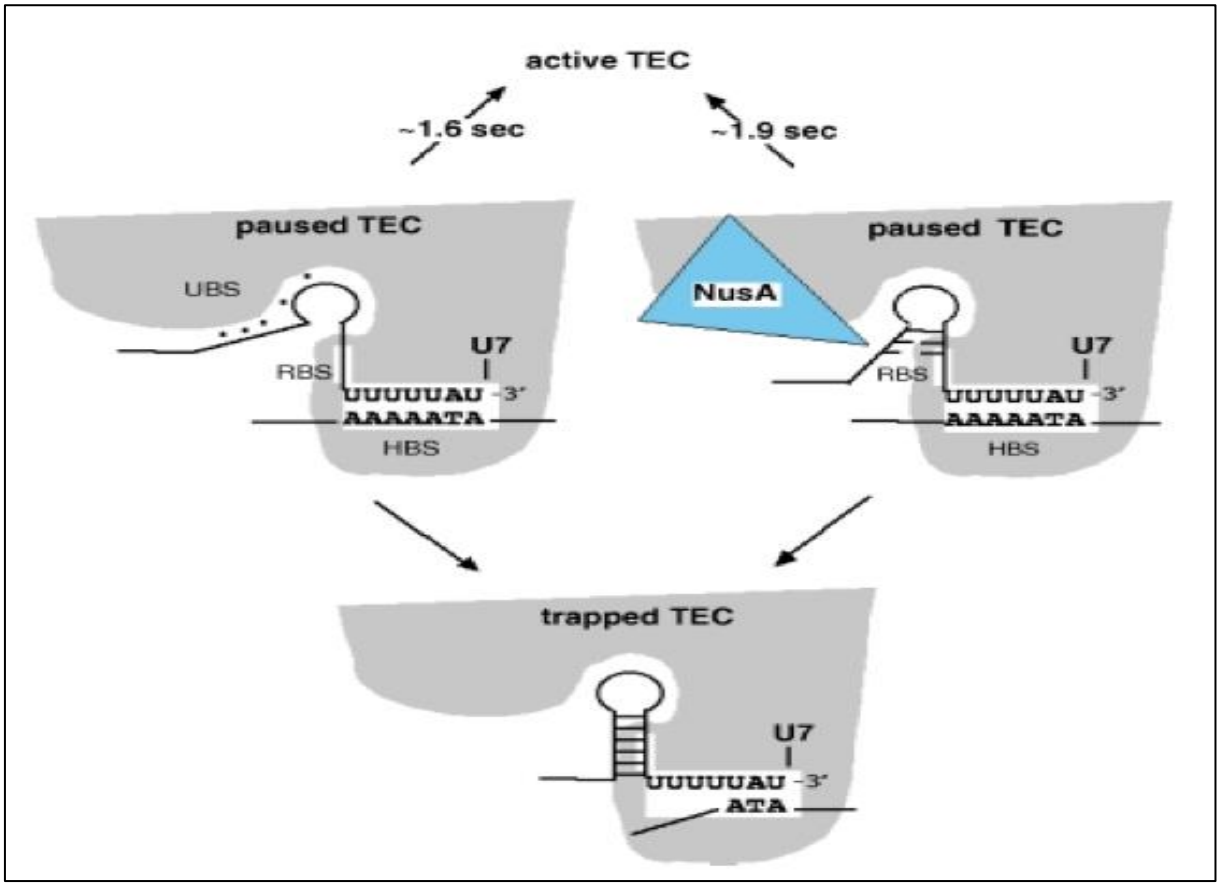

Figure 6. Transcription elongation complex (TEC) pausing and termination.

This diagram illustrates how paused TECs can lead to transcription termination. Adapted from (Nudler and Gottesman, 2002).

DNA transcription represents a target to be controlled in order to regulate gene expression. Three main transcriptional steps can be targeted to regulate bacterial DNA transcription: initiation, elongation and termination.

1.5.1. Initiation regulation

Bacterial transcription initiation is a multistep process starting with promoter recognition by DNA-dependent RNAP. This step is followed by a sequence of conformational changes in the RNAP complex and the DNA promoter itself in order for RNA synthesis to be initiated (Saecker, 2011). RNAP transcribes the DNA region located between +1 and the termination sites, producing a nascent RNA strand. At the recognition phase, the DNA is still in its original double-stranded form. After that, the DNA is unwound at the +1 site, creating a bubble that is known as the open complex.

1.5.1.1. Sigma factors

Transcription initiation can be controlled by RNAP affinity to promoters, attained by recruiting a σ factor, and promoter strength. As already mentioned, σ factors recognize different promoters and have a high affinity to specific promoter elements. The first regulatory line for transcription initiation is to control the amounts and affinity of these σ factors to the promoters to which they bind. σ factors with low affinity to promoters are assisted by different transcription factors and regulatory proteins.

All bacteria contain one housekeeping σ and several other alternative σ factors. *E. coli* has six alternative factors that are highly regulated. For example, σ^{38} is regulated by its synthesis, sRNAs and proteolysis. Other alternative σ factors are regulated by anti-sigma

factors. For example, σ^{28} (FliA) is regulated by an anti-sigma factor called flgM (Tlapak *et al.*, 2017); and σ^{24} (σE) is controlled by an anti-sigma factor called RseA (El-Mowafi *et al.*, 2015).

1.5.1.2. Transcription factors (TFs)

While σ factors play an important role in the regulation of transcription initiation, the principal regulators for most responses are transcription factors. Thousands of genes are encoded in any bacterial genome, and these produce all of the proteins necessary for bacterial survival. These proteins are needed in different quantities, depending on the surrounding environment and bacterial conditions. Based on this differential need for various proteins, not all genes are expressed all of the time; such an unregulated approach to gene expression would jeopardize bacterial survival by wasting required energy and by producing more proteins than needed, which could be toxic for the cell. To overcome this issue, bacterial cells adopt several activation mechanisms by recruiting hundreds of activator proteins that assist in activating genes producing proteins limited to bacterial requirements. These activator factors vary in their functions and mechanisms of action.

Transcription factors can positively or negatively affect transcription initiation by repressing or stimulating it. The regulatory activity of some of these proteins has been found to target only one or two promoters, while others have been found to control several promoters (Browning and Busby, 2004). Martinez-Antonio and Collado-Vides (2003) also claimed that seven transcription factors are considered to be global or master transcription factors: IHF, Fis, CRP, ArcA, NarL, Lrp and FNR. These global factors

control more than half of all genes. Conversely, approximately 60 transcription factors are specific to only one promoter (Browning and Busby, 2004).

1.5.1.2.1. Factors that activate transcription initiation

This group of transcription factors includes factors that intervene with the interaction between RNAP and the DNA to stimulate transcription initiation (i.e., positive regulation). A very well reviewed example of this kind of factor is the cyclic-AMP (cAMP) receptor protein (CRP), also known as catabolite activator protein (CAP). At high levels of cAMP in the cell, CRP binds to cAMP and becomes activated as a cAMP-CRP complex. This activated complex stimulates transcription at different promoters, including the *lac* promoter, which controls expression of the genes that are responsible for the utilization of lactose (Ebright, 1993; Saha *et al.*, 2015). This activator has an affinity to the CTD of the RNAP α-subunit. It initially binds to the DNA at a target upstream of the target promoter DNA. Then, by its affinity to the CTD of the RNAP α-subunit, the activator recruits RNAP. When the interaction occurs, other RNAP holoenzyme subunits interact with the downstream promoter sequence.

Transcription activator proteins can be divided into two main groups: activators that directly interact with RNAP and activators that do not directly interact with RNAP. There are two classes of activation among the first group of activators, known as class I and II. In class I, activators bind somewhere upstream of the target promoter, so they can bind to the CTD of the RNAP α-subunit and offer an opportunity for other RNAP subunits to contact the downstream promoter elements. CRP, which enhances the expression of the lactose operon by binding upstream of the *lac* promoter, provides an

extensively studied example of class I activation. (Ebright, 1993; Busby and Savery, 2007).

In class II, activators bind to a target overlapping the -35 promoter element and helps in converting RNAP-promoter closed complex (RPc) into RNAP-promoter open complex (RPo) (Feng *et al.*, 2016). Therefore, the CTD of the RNAP α-subunit binds upstream of the activator. In this scenario, the activator itself can interact with some parts of the RNAP holoenzyme, primarily region 4 of the σ subunit. Two very well understood examples for class II activation are the CRP protein at the promoter of the galactose operon and the bacteriophage λ cI protein at the λ P_{RM} promoter (Niu *et al.*, 1996; Busby and Savery, 2007; Lee *et al.*, 2012).

Members of the other group of activators help in regulating the transcription process without directly binding to RNAP. They bind to the DNA in order to change its conformation and make it easier for RNAP to bind and initiate transcription, or they change the DNA topology to make it easier for RNAP to reach the target promoter.

At some promoters, transcription is suppressed as a result of the highly compact structure of bacterial DNA. This highly compact structure is created by the interference of some nucleoid-associated proteins (NAPs) like the histone for nucleoid structuring protein (HNS) and the integration host factor (IHF). For these promoters to be transcribed, activation factors need to disrupt the compact structure of the DNA. In some cases, the thermally assisted "breathing" of the DNA helps locate the target of initial DNA unwinding. The IHF protein functions by sharply bending the DNA, which can be either helpful or suppressive for transcription processes depending on where it binds. In some cases, bending the upstream DNA sequence can prevent it from hindering bound RNAPs, which eventually facilitates the transcription initiation process. Also,

when transcription factors bind well upstream of the promoter region, bending the DNA with IHF can bring those transcription factors and promoters together, which also favors transcription. In other cases, IHF binds to the promoter itself, preventing RNAP from binding, which suppresses the transcriptional activity of this promoter (Pagel *et al.*, 1992; Ueguchi and Mizuno, 1993; Busby and Savery, 2007; Browning *et al.*, 2010).

Fnr, which is activated by oxygen starvation, NarL and NarP are other transcription factors that regulate *E. coli nir* operon transcription without directly contacting RNAP. The Fnr factor binds upstream of the *nir* transcription start site between bases -41 and -42. The NarL and NarP factors bind between bases -69 and -70 upstream of the *nir* transcription start site (Wu *et al.*, 1998).

1.5.1.2.2. Factors that repress transcription initiation

This group of factors includes proteins that negatively affect transcription initiation. Some of these factors inhibit transcription initiation by binding to a specific target in the DNA overlapping the promoter region. This binding hinders RNAP from binding to the promoter, which eventually deters transcription initiation. An example of this kind of regulation is the effect of the Lac repressor at the *lac* promoter. The *lac* operon (Figure 7), when transcribed, assists the cell in the utilization of lactose, in the absence of glucose, as a source of energy. A protein known as Lac repressor or LacI mediates this switch. This protein binds to three different sites on the *lac* operon to repress the transcription of the three *lac* genes (*lacZ*, *lacY* and *lacA*). The *lacZ* gene is responsible for the production of the β-galactosidase enzyme, which catalyzes the hydrolysis of β-galactosides into monosaccharides. The second gene of the *lac* operon is *lacY*, which

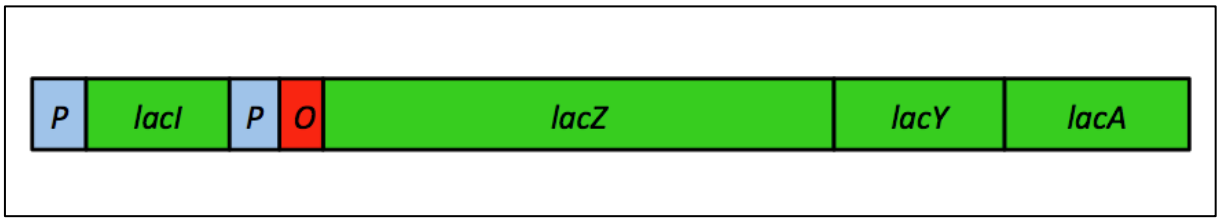

Figure 7. The *lac* operon.

This figure shows the main components of the *lac* operon.

codes for a membrane protein that assists in transporting lactose into the cell known as the lactose permease (LacY) enzyme. The third gene is *lacA*, which is responsible for the production of an enzyme involved in sugar metabolism known as acetyltransferase (Ozbudak *et al.*, 2004). This enzyme is also known as thiogalactoside transacetylase and assists in bacterial cell detoxification by preventing accumulation of metabolic compounds (Andrews and Lin, 1976).

In the absence of lactose, LacI binds to its binding site on the DNA to inhibit the transcription of the *lac* operon. When lactose is present, a lactose isomer called allolactose binds to LacI and releases it from the DNA. This induces transcription of the *lac* operon to utilize cellular lactose, provided that CRP is active due to glucose being absent. LacI release can also be triggered by isopropyl-β-D-1-thiogalactoside (IPTG). IPTG works in favor of transcription initiation of the *lac* operon by binding to the repressor, which negatively affects its affinity to the DNA operator region (Reznikoff *et al.*, 1985; Kercher *et al.*, 1997; Bell and Lewis, 2001; Gatti-Lafranconi *et al.*, 2013).

Transcription initiation regulation can be attained by two other mechanisms: altering DNA topology and improving or demoting the promoter sequence (Reznikoff *et al.*, 1985).

1.5.1.3. Transcription initiation regulation by altering DNA topology

Regulation of transcription initiation can also be achieved by controlling DNA topology at the promoters. Bacterial DNA is highly supercoiled so it can be compacted and fit into a very small cellular space. As a result of this supercoiling, the DNA becomes highly tensed, which can negatively affect the rate of initiating transcription at particular

promoters. Proteins called topoisomerases can control the topology of bacterial DNA to regulate transcription initiation. There are two types of topoisomerases (I and II), and each one relaxes the double-helix DNA structure in a different way. Topoisomerase I cuts one DNA strand and passes the other strand around the temporarily introduced cut to relax the DNA. Topoisomerase II or bacterial gyrase relaxes the double-stranded DNA by cutting both strands and passing another double-stranded DNA through that cut (Wang, 1985; Wang, 1996). An example of a supercoiling-dependent promoter is the DNA gyrase promoter. DNA gyrase is an enzyme responsible for the DNA supercoiling in bacteria. However, DNA supercoiling modulates the synthesis of the gyrase enzyme itself. Up to 10-fold increase in gyrase protein synthesis was noticed after relaxing the DNA by blocking gyrase activity (Menzel and Gellert, 1983).

1.5.1.4. Transcription initiation regulation by promoter modification

Some bacteria have evolved different mechanisms in which they modify promoter bases or promoter structure in order to regulate transcription initiation. One of these mechanisms involves introducing a modification to the promoter bases. An example for this kind of modification is the methylation of these bases. This mechanism represents an evolutionary adaptation by which bacteria protect themselves from foreign viral DNA or transposons; the bacterial cell produces restriction enzymes that cut unmodified foreign DNA but not its own methylated DNA (Low *et al.*, 2001). Methylation can also be used to prevent the binding of repressors at the *pap* and *agn43* promoters in *E. coli*. In these two scenarios, the repressor binding is blocked by the action of the Dam methylase enzyme, leading to activation of the related promoters (Browning and Busby, 2016).

Another mechanism for controlling gene expression includes promoter inversion. Some bacteria have been found to respond to a signal by turning the expression of specific genes on or off. This activation or deactivation of gene expression is achieved by flipping the whole promoter so it sends recruited RNAPs toward or away from its corresponding gene, respectively. This mechanism can be clearly seen in the *fim* operon, where the expression of type 1 fimbrial genes is regulated by inverting their promoter's orientation. FimE and FimB recombinases regulate the direction of this inversion (Cerdeño-Tárraga *et al*., 2005; Browning and Busby, 2016).

To adapt to changing environments, bacteria have evolved another strategy to regulate gene expression levels. This method involves mutations to an area of the promoter or the reading frame known as simple sequence repeats (SSRs), which carry a tract of one or more base repeats. These SSR regions are of different lengths between individual bacteria of the same population as well as from one generation to another. The difference in lengths leads to different promoter activities (Moxon *et al.*, 2006; Browning and Busby, 2016).

Some other promoters have non-optimal spacing between the -35 and -10 elements. Members of a family of activator proteins, known as the MerR family, activate transcription at promoters where the distance between the -35 and -10 elements is greater than the optimal spacing, which is normally 17 ±1 bp. As a result of this abnormal spacing, RNAP binds only to the UP and -35 elements but not the -10 element, rendering the formation of an open complex impossible. This problem can be resolved by the interference of MerR factors, which can correct the distance between the -35 and -10 elements by moving the -10 element to the correct position where it can bind to the RNAP σ subunit (Brown *et al.*, 2003; Busby and Savery, 2007; Lee *et al.*, 2012).

Some of the MerR family members, can act as both repressor and activator. An example of this is CueR, that activates *copA* promoter. CueR binds to and causes some conformational changes to promoter DNA. These changes might positively or negatively affect promoter recognition by RNAP. When CueR binds the DNA in the absence of metal ion, it bends the DNA to a degree that negatively affects RNAP-promoter recognition by moving the -10 element away from domain 2 of σ factor. However, when this binding takes place in the presence of metal ion, CueR acts as an activator by unwinding and kinking the DNA more, which favor RNAP-promoter recognition by bringing the -10 element closer to the σ domain 2 of RNAP. In both cases, some bases of the protein-binding site, located between the proteins' two binding domains, are twisted more than others to cause this kink (Heldwein and Brennan, 2001; Philips *et al.,* 2015).

1.5.2. Elongation regulation

Transcription elongation takes place after the recognition and binding of RNAP to a promoter. In this step, RNAP binds RNA nucleotides to each other in specific order that is complementary to the template DNA base order. Elongation cannot begin unless RNAP is completely clear of the promoter region. This escape is characterized by three different changes occurring to RNAP: dissociation of the recognition subunit (σ factor) from RNAP, stabilization and robust binding of RNAP to the DNA template and primary movement of RNAP along the DNA template (Uptain *et al.*, 1997).

Transcription elongation can be regulated by three different factors: signals, accessory factors and polymerase modifications. Signals that regulate transcription elongation

involve growth conditions and stress responses. For example, the elongation of the *rrn* operon in *E. coli* is directly affected by its stringent response through the ppGpp signaling molecule. This molecule is responsible for the stringent response in *E. coli* that can lower the elongation rate of RNAP in *E. coli* and increase the frequency of *rrnB* operon premature termination (Kingston and Chamberlin, 1981). Similar to growth conditions, stress responses can affect transcription elongation as well. Vidovic and colleagues (2011) proteomically analyzed the wild-type *E. coli* 0157, which is adapted to cold, and the *rpoS* mutant strain and recognized a difference in the expression of 21 proteins between those two strains. They also found that the RpoS σ factor regulates the expression of several proteins when *E. coli* is exposed to a cold shock. Some of these proteins are important for adaptation to cold stress and others are vital for the normal central metabolic pathways of *E. coli*.

Accessory factors that regulate transcription elongation can be identified as proteins or small molecules that inhibit or promote transcription elongation and affect transcription elongation in different ways. Accessory factors can regulate transcription by affecting RNAP and/or its accessory proteins or the DNA itself. Some of these proteins, such as GreA and GreB, increase elongation rates by decreasing pausing or helping arrested transcription complexes resume elongation. Some other proteins affect the elongation process by binding to the nascent RNA transcript (such as the Rho protein described previously) or binding to the DNA template. Other protein factors affect the EC by altering its processivity through direct protein:protein interaction, leading to different changes in the activity of the EC. For instance, in phage λ, the N protein encoded by the phage initially binds to the *nut* site on the RNA strand. After that, a group of the host proteins (Nus factors and ribosomal S10 proteins) join the N protein on the *nut* site. This

protein complex then interacts with the elongating RNAP and alters its protein components, leading to a more stable complex that is resistant to downstream pauses and termination signals (Friedman and Court, 1995; Uptain *et al.*, 1997; Henkin, 2000).

Bacterial RNAP consists of several subunits participating to achieve transcription. It has been found that these subunits have many sequence conservations among bacterial species (Allison *et al.*, 1985; Archambault and Friesen, 1993). This leads to the suggestion that these proteins represent a possible target for transcription elongation regulation. Some of the RNAP subunits experience changes such as phosphorylation and glycosylation that can affect RNAP efficiency. However, in eukaryotes, these types of changes do not affect transcription unless they occur to large RNAP subunits. In eukaryotes, the C-terminal of the largest subunit of RNAP II contains a repeated sequence of heptapeptide amino acids, which are highly phosphorylated. RNAP II with a highly phosphorylated C-terminus has been found to be more efficient in elongation than RNAP II with a non-phosphorylated C-terminus (Uptain *et al.*, 1997). In *E. coli*, phosphorylation of the β and β' subunits was found to be associated with the shutoff of RNA synthesis. This connection emphasizes the association between phosphorylation and elongation regulation (Cozzone, 1988).

The transcription EC has the ability to introduce conformational changes to the transcribed DNA segment. As bacterial DNA is double-stranded and highly supercoiled, RNAP and its nascent RNA need to rotate around the DNA during elongation, which is not possible at all times as the DNA is stored in a highly compact structure and the RNAP itself is very large in comparison to the chromatin filaments. As a result, RNAP, with topoisomerase I ahead of it, forces the DNA to rotate around its axis, resulting in an increasing torsional stress in the downstream DNA. The introduced torsional stress may

slow the movement of the proceeding elongation complex and can play an important role in transcription regulation (Liu and Wang, 1987; Kouzine and Levens, 2007; Kouzine *et al.*, 2013). These facts led to the twin domain hypothesis by Liu and Wang in 1987. This hypothesis illustrates that ECs change the topology of DNA by leaving negative supercoiling behind them and increasing positive supercoiling ahead of them when their rotation is obstructed during active transcription (Liu and Wang, 1987). Moreover, the bacterial chromosome is structured in a way that is compatible with the bacterial cell cycle, which might play a role in elongation regulation. Bacterial chromosome consists of two main structures, which are the topologically constrained loops, or topological domains, and the higher structures known as macrodomains. The topologically constrained loops are formed by connections resulted from the negative supercoiling of the DNA. These loops are stabilized by the attachment of some DNA binding proteins along bacterial DNA to protect the cell against DNA over relaxation during replication and other processes by which the DNA is unwound (Postow *et al.*, 2004). During the repair of DNA double-strand breaks, DNA topological domains help by keeping DNA ends close to each other (Wang *et al.*, 2013). Macrodomains are suggested to be formed by some sequence specific binding proteins. The formation of these macrodomains is central for bacterial DNA compaction. There are four macrodomains in *E. coli* chromosome (Dame *et al.*, 2011; Wang *et al.*, 2013).

1.5.3. Termination regulation

Transcription termination occurs at the end of each transcriptional process or when cells need to abort the transcription process. This process is characterized by two

events: the release of the newly synthesized nascent RNA from the transcription complex and the detachment of the transcribing RNAP from the template DNA. The termination step can be regulated by different mechanisms. One of these mechanisms is directly related to the Rho-dependent termination mentioned previously. When the Rho protein binds to the transcript and moves towards the transcript's 3' end, it catches up with stalled RNAP complexes at pausing sequences. Once the Rho protein contacts RNAP, RNAP terminates transcription and releases both the RNA transcript and template DNA. The regulation of this process can be attained by controlling the sensitivity of RNAP to the pausing sequences and by controlling the Rho protein supply to the RNA transcript or RNAP (Henkin and Yanofsky, 2002).

Another mechanism of regulating transcription termination is by employing the transcription termination–antitermination process. This process is commonly known as transcription attenuation. In this process, bacterial cells respond to a metabolic signal and instruct RNAP to terminate transcription or to continue transcribing the subsequent genes of the same operon (Merino and Yanofsky, 2005). Attenuation occurs when the RNA transcript forms one of two structures (i.e., hairpin loops): the antiterminator loop or the terminator loop. The formation of the antiterminator loop prevents the formation of the terminator loop and allows the transcription process to proceed. Alternatively, the terminator hairpin loop leads to RNAP being stalled and transcription being terminated (Lathe *et al.*, 2002).

1.5.4. Gene regulation by location

Regulation of gene expression can also be attained by means other than the aforementioned mechanisms and factors. Specifically, it has been discovered that gene positioning within the chromosome can affect gene expression in both prokaryotes and eukaryotes (Dobzhansky, 1936; Bryant *et al.*, 2014). Some genes were found to be less active or even silent when their locations changed on the chromosome. Other genes demonstrated increased expression when their location changed. These observations demonstrate that gene expression can be regulated by the location of the gene on the chromosome (Bryant *et al.*, 2014).

During their studies on *Pseudomonas putida* using *lacZ* fusions to different promoters, de Lorenzo and colleagues (1993) found that the activities of several catabolic systems differ significantly depending on the promoter's location on the chromosome. In another study, Sousa and colleagues (1997) suggested that chromosomal positioning can modulate gene expression and change gene activity in response to changing environments.

Gene expression is likely altered according to the gene's position on the chromosome according to two means. First, expression decreases as the distance between the gene and the origin of replication (*oriC*) increases. In other words, genes located near the *oriC* have more copies, particularly at high rates of growth when multiple rounds of chromosome replication are initiated. Second, DNA compaction and supercoiling have been found to affect gene expression. Therefore, gene expression can be altered significantly by changing the location of the gene along the chromosome (Block *et al.*, 2012).

Schmid and Roth (1987) conducted a study on the effect of location on the expression of the *his* gene of *Salmonella typhimurium.* The researchers constructed large numbers of strains carrying translocations of the *his* gene at defined chromosomal positions to test if these translocations had any effect on gene expression. Schmid and Roth eventually observed that gene expression was different in each location and suggested that these differences resulted from the differences in gene dosage of chromosomal sites; this finding again supports the notion that genes located near *oriC* have more copies than distant genes.

1.6. Quantitative aspects of gene expression

1.6.1. Effect of rifampicin on transcription

Rifampicin, also known as rifamycin, is one of the most important antibiotics in clinical use. This antibiotic, which works as an inhibitor of RNAP, was discovered decades ago and has been used as a treatment for serious bacterial infections. It has been found that mutations introduced to the RNAP β subunit lead to the development of rifampicin resistance in *E. coli*, suggesting that this antibiotic interferes with the transcription initiation step (Rabussay and Zillig, 1969). Furthermore, a study published by Hinkle and his coworkers in 1972 proved that rifampicin has no effect on RNAP binding to bacterial DNA (Hinkle *et al.*, 1972). Another study showed that rifampicin binds to a pocket in the RNAP β subunit fairly close to the RNAP active site; once RNAP escapes the promoter region and produces a long RNA chain, it becomes completely resistant to rifampicin. The same study illustrated that, after rifampicin treatment, RNAP produces a very short RNA chain (2–3 nucleotides long) and then stops elongating (Campbell *et al.*,

2001). The sum of findings from these three studies support the fact that rifampicin acts by physically blocking the nascent RNA channel in RNAP. However, more recent studies have reopened the issue of what rifampicin does to inhibit transcription. One of these studies has suggested that steric blocking, of the RNA exit channel by rifampicin, alone is not enough to inhibit transcription. This study has proposed another component of the inhibition mechanism. This component is represented by an allosteric signal resulted from the binding of rifampicin to the RNAP molecule. This signal negatively affects the phosphodiester bond formation by preventing the binding of Mg^{2+} ion, which eventually leads to the release of short RNA transcripts (Artsimovitch *et al.*, 2005). Generally, rifampicin-treated RNAPs will be trapped at the promoter only. This property has allowed researchers to study RNAP distribution along the bacterial chromosome and RNAP binding sites around the whole genome (Grainger *et al.*, 2005; Cho *et al.*, 2009). After rifampicin treatment, each promoter will be occupied by one RNAP. Assuming 100% promoter occupancy after rifampicin treatment will assist in further calculations regarding promoter activity.

1.6.2. Promoter strength

Promoter strength can be defined by the frequency at which RNA polymerase initiates transcription. This strength depends primarily on promoter sequence, which plays a vital role in the distribution of RNAP among the massive numbers of promoters. It has been revealed that each promoter has two conserved regions that are recognized by the σ factor of RNAP. These conserved regions are known as consensus sequences (Wosten, 1998). The two consensus regions are the -10 and -35 promoter elements or hexamers.

The activity of each promoter depends mainly on the sequences of these two hexamers. Having the -10 and -35 hexamers with sequences identical or near to the consensus sequences increases the promoter strength and vice versa. Another two promoter elements were found to play significant role in transcription initiation: the extended -10 element and the UP element. The extended -10 element consists of 3 to 4 bp and is situated directly upstream of the -10 hexamer (Browning and Busby, 2004). In other studies, the extended -10 element was found 1 bp upstream of the -10 hexamer (Burr *et al.*, 2000; Mitchell *et al.*, 2003). Both of these observations are accurate, as the extended -10 element can be found adjacent to the -10 hexamer or separated from it by 1 bp. A functional extended -10 element is not present in some primitive promoters. This element is recognized by domain 3 of σ factor. Some σ factors have domain 2 only to bind to -10 elements and some other have domains 2 and 4 to bind to -10 and -35 elements, respectively. Domain 3 has evolved to recognize the sequences located between these two elements, which involve the extended -10 element. So, the extended -10 element has, presumably, evolved after the -10 and -35 elements.

The UP element consists of approximately 20 bp situated upstream of the -35 hexamer. The length of the spacer sequence between the -10 and -35 hexamers is also important for the transcription process. A previous study proved that the optimal length for this region is 17 bp. This study compared promoters with 16-, 17- and 18-bp spacer regions and found that the rate of open complex formation when the spacer region is 17-bp long is higher than when it is 16- or 18-bp long (Stefano and Gralla, 1982).

Like all other promoter elements, the UP element plays a role in activating transcription. Most *E. coli* promoters have an UP element upstream of the -35 element. This element is

rich in A and T bases which is recognized by the CTD of the RNAP α subunit. It has been found that the presence of the UP element upstream of some *E. coli* promoters increases the promoters' activity. This activity immensely decreased after introducing a mutation to the RNAP α subunit, which supports the notion that the UP element has an important role in RNAP recruitment (by contacting the αCTD domain of RNAP α subunit) and promoter activity and this has been quantified by Ross and her co-workers (Ross *et al.*, 1993).

1.6.3. Promoter strength measurements

There are several techniques that can be used to measure promoter strength, such as reporter gene assays and electrophoretic mobility shift assays (EMSA). Promoter strength can be estimated by measuring any reporter gene product activity, such as the *lacZ* gene product (β-galactosidase) enzyme activity, assuming that this activity reflects promoter strength. The principle of the β-galactosidase assay is to measure the hydrolysis of O-nitrophenyl-β-D-galactopyranoside (ONPG). ONPG is hydrolyzed by water giving galactose (colourless) and O-nitrophenol (yellow). The amount of O-nitrophenol formed is normally proportional to the amount of β-galactosidase and the reaction time. The amount of O-nitrophenol produced can be measured by determining the absorbance at 420 nm. At the end of the assay, the reaction is stopped by adding Na2CO3, leading to the rapid increase in the reaction mixture pH to around 11. At this pH, a certain amount of the O-nitrophenol is converted to the yellow-coloured anionic form and β-galactosidase is inactivated.

EMSA is a technique mainly used to study protein-RNA or protein-DNA interactions. EMSA can assess binding complexes and determine if a protein is capable of binding to a given DNA or RNA sequence. This assay depends on the electrophoretic separation of protein-RNA or protein-DNA mixtures on an agarose or polyacrylamide gel depending on their sizes. Therefore, if the protein is capable of binding to DNA or RNA, the resulting complex will move slowly through the gel and demonstrate a band with a larger size than a band of protein alone. This technique can be applied to assess promoter strength by adding different concentrations of RNAP (protein) with the promoter. Strong promoters will produce binding complexes (promoter-RNAP complexes) with low RNAP concentrations. However, weak promoters will need higher RNAP concentrations to produce complexes.

Several previous studies have tried to indirectly measure promoter strength and RNAP flux through given genes in living cells. Kelly and colleagues (2009) tried to indirectly measure promoter strength by fusing them to genes coding for green fluorescent protein (GFP) and calculate promoter strength depending on the amount of GFP produced by these genes under the control of different promoters. Using their calculations, which involved estimates of the rate of GFP synthesis, folding and maturation, and RNA half-life, they were able to measure polymerase per second (PoPS) for their promoters. Another recent study used computational methods to design genetic circuits, which then inserted in different specific locations on the DNA and transformed into bacterial strain. They were able to indirectly calculate the amount of RNAP per second per the standard promoter used in their study (Nielsen *et al.*, 2016).

To date, to my knowledge, no reliable technique has been developed to directly measure promoter activity *in vivo*. The two aforementioned techniques and many other

techniques rely solely on measuring genes' final products. The reporter gene assay measures the amount of final product produced by a gene, which can be compared to another gene's products to evaluate which gene is stronger. The EMSA technique depends on measuring the sizes of protein-DNA or protein-RNA complexes *in vitro* and changing the protein concentrations to determine which protein concentration produces binding complexes. Hence, both techniques are indirect and *in vitro*.

Considering a lack of reliable methods for measuring promoter activity directly in real-time, as inside the cell, an alternative technique for measuring promoter activity is desirable. In this study, I developed a new, real-time method to directly assess promoter strength depending on RNAP activity. This method allowed us to yield more measurements regarding promoter strength, such as the promoter competitivity, promoter occupancy index (POI), promoter escape index (EI), fragment occupancy percentage (FOP) and PoPS.

All samples were tested using a chromatin immunoprecipitation (ChIP) experiment to immunoprecipitate DNA fragments that were bound to RNAP in each sample. The resultant DNA fragments were quantified by quantitative polymerase chain reaction (qPCR) using specific primers. qPCR employs the same procedural steps as conventional PCR. However, in qPCR, as the DNA is labelled with a fluorescence dye, the amount of the amplified DNA can be accurately counted by monitoring the fluorescence emitted in each thermal cycle. At the end of each qPCR run, readings are given as cycle threshold (Ct), which describes how many cycles emitted a fluorescence exceeding the background signal. The Ct value is relative to the initial amount of the DNA, in that fluorescence will be detected earlier in samples with high initial DNA concentrations (i.e., samples with high DNA concentrations will yield low Ct values). After qPCR, these Ct values were used

to calculate the amount of DNA fragments of interest that were occupied by RNAP in each sample. These experiments were performed in triplicate for each sample as a biological repeat. For each biological repeat, qPCR was performed in triplicate as a technical repeat.

This new direct method developed in this study was applied on two different systems: the *lac* operon on the chromosome and the *lac* operon in pRW50 with different promoters. For the *lac* operon on the chromosome system, the *lac* operon activity was tested on the chromosome twice: one time before and after rifampicin treatment and another time before and after IPTG treatment. In qPCR, seven newly designed primer pairs were used. A primer pair for the promoter region (lac0), five sets of primer pairs distributed along the *lac* operon (lac1, lac2, lac3, lac4, and lac5) and a primer pair for a control region. For the *lac* operon in pRW50, nine synthetic promoters with different activities inserted into the pRW50 plasmid, just before the *lac* operon, were tested with and without rifampicin treatment. Then, qPCR was applied using four sets of primer pairs to quantify how many RNAPs were attached to the promoter, two downstream regions on the *lacZ* gene and a control region.

The Cts obtained from all qPCR tests were used to calculate the occupancy value of each region. These occupancy values were then used to calculate how competitive is the promoter in recruiting RNAPs, how efficiently the RNAP recruited by the promoter (i.e., POI), how efficiently the promoter allows RNAP to escape (i.e., EI), the time interval between RNAPs transcribing downstream regions (T_{int}), downstream fragments occupancy percentage (FOP) and the rate at which RNA polymerase moves past a given position on the DNA (i.e., PoPS).

Chapter 2

Materials and methods

2.1. Suppliers

All chemicals, media, reagents and consumables were purchased from the following suppliers, unless otherwise specified: GE Healthcare, SIGMA-ALDRICH, Agilent Technologies and Neoclone Biotechnology.

2.2. Bacterial strains, media, antibiotics, growth conditions and bacterial storage

2.2.1. Bacterial strains

Two *E. coli* strains were used throughout this study: *E. coli* M182 (genotype: Δ(*lacIPOZY*)*X*74, *galK*, *galU*, *strA*), obtained from J. Beckwith (1969); and *E. coli* MG1655. The M182 strain was transformed with pRW50 plasmids containing promoters used in this study to test the promoters' activities and RNAP flux in the *lacZ* gene. The MG1655 strain was used to test the activity of *lac* promoter on the chromosome and RNAP flux in the *lacZ gene*.

2.2.2. Media

Both solid and liquid media in this study were prepared by dissolving the required amounts of reagents in distilled water.

2x Luria-Bertani (LB) broth (1 L):

20g tryptone

10g NaCl

10g yeast extract.

Distilled H_2O to 1 L

LB Broth with agar (Lennox) (1 L):

15g agar

10g tryptone

5g NaCl

5g yeast extract.

Distilled H_2O to 1 L

MacConkey agar (1 L):

40 g MacConkey agar base

10 g Lactose

Distilled H_2O to 1 L

After preparation, all liquid media were divided into small aliquots to avoid contaminating large amounts of media. All liquid and solid media were sterilized by autoclaving at 121°C for 20 minutes. LB broth with agar was allowed to cool down to 50°C to 55°C before the addition of tetracycline. Then, it was poured into petri dishes

under sterile conditions and allowed to set on the bench. All plates were stored at 4°C and allowed to dry at 37°C in an incubator before use.

2.2.3. Antibiotics used in this study

Tetracycline

200 mg tetracycline

Methanol to 20 ml

Rifampicin (50 mM)

1 g rifampicin

Methanol to 24 ml

Tetracycline and rifampicin were used in this study. Tetracycline was used at a final concentration of 35μg/ml for the selection of plasmid-encoded resistances, as all of the plasmids used in this study held a tetracycline resistance gene. Rifampicin (Sigma) was used at a final concentration of 50 μM to treat the cells for 15 minutes (when required) just before OD_{650} reached 0.4 (Grainger *et al.*, 2005).

2.2.4. Growth conditions

Agar plates

Bacterial cells were streaked onto appropriate agar plates (containing tetracycline where appropriate) for single colonies and incubated at 37°C in an incubator overnight.

Liquid media

For a 5 ml overnight liquid culture, a single colony was picked from the appropriate overnight culture plate and inoculated in 5 ml of the required media (containing tetracycline where appropriate) in 25 ml conical flasks. Flasks were then incubated at 37^{o}C in a shaker overnight.

For media used in the ChIP experiment, a calculated amount from the overnight liquid growth was inoculated in 40 ml of 2x LB media (tetracycline supplemented where appropriate). For testing RNAP flux through the *lac* operon, the 40 ml of media was supplemented with 100 μM IPTG to overcome the *lac* repressor effect (Reznikoff, 1992; Wilson *et al.*, 2007).

2.2.5. Bacterial storage (glycerol stock preparation)

Bacterial cells were plated on agar plates by streaking (for single colonies) and incubated at 37°C overnight. A single colony was picked and inoculated in 5 ml of 2x LB in a 25 ml conical flask and incubated with shaking overnight in a 37^{o}C incubator. Next, 300 μl of the overnight culture was mixed with 700 μl of 50% glycerol (Fisher Chemicals) and kept frozen at -80°C until use.

2.3. Plasmid DNA

The pRW50 plasmid was used as a vector for different promoters used in this study.

This vector contains different regions, such as the *lac* operon located directly downstream of the promoter's region and a control region with no expression at all (Figure 8). This plasmid DNA also contains a tetracycline resistance gene for the selection of desired cells only.

2.4. Buffers and Solutions

2.4.1. Solutions for general use

70% ethanol (100 ml)

70 ml of 100% ethanol

Distilled H_2O to 100 ml

50% glycerol (50 ml)

50 ml of 100% glycerol

50 l distilled H_2O

2.4.2. ChIP buffers

Tris-buffered saline (TBS) (500 ml)

50 mM Tris, pH7.6

150 mM NaCl

Distilled H_2O to 500 ml

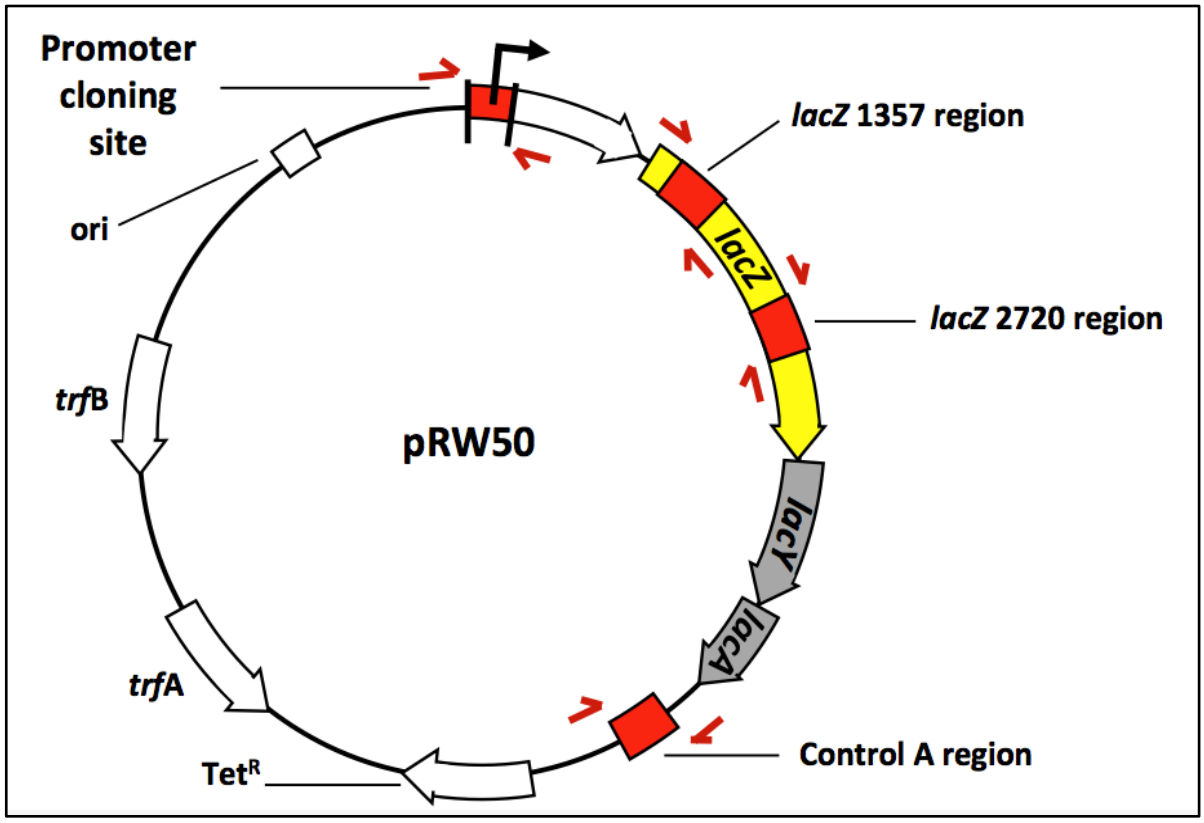

Figure 8. The map of pRW50 plasmid used in this study

This figure shows a diagram of the pRW50 plasmid and the approximate locations of lac1357, lac2720, the control region, promoter's cloning site, the origin of replication (ori) and the tetracycline resistance gene (Tet^R) (Lodge *et al*, 1992).

FA Lysis buffer (100 ml)

10 mM Tris, pH8.0

20% sucrose

50 mM NaCl

10 mM EDTA

Distilled H_2O to 100 ml

20 mg/ml lysozyme (to be added immediately before use)

IP (Immunoprecipitation) buffer (300 ml)

50 mM HEPES-KOH, pH7.5

150 mM NaCl

1m M EDTA

1% Triton X-100

0.1% Na deoxycholate

0.1% SDS

Distilled H_2O to 300 ml

IP buffer + alt (100 ml)

50 mM HEPES-KOS, pH7.5

500 mM NaCl

1 mM EDTA

1% Triton X-100

0.1% Na deoxycholate

0.1% SDS

Distilled H_2O to 100 ml

IP wash buffer (100 ml)

10 mM Tris, pH8.0

250 mM LiCl

1 mM EDTA

0.5% Nonidet P-40

0.5% Na deoxycholate

Distilled H_2O to 100 ml

TE (Tris/EDTA) buffer (100 ml)

10 mM Tris, pH7.5

1 mM EDTA

Distilled H_2O to 100 ml

1x Elution buffer (100 ml)

50 mM Tris, pH7.5

10 mM EDTA

1% SDS

Distilled H_2O to 100 ml

2.4.3. Reporter gene assay solutions

Z buffer (1 L)

0.75 g of KCl

0.246 g of $MgSO_4 \bullet 7H_2O$

8.53 g of Na_2HPO_4

4.866 g of $NaH_2PO_4 \bullet 2H_2O$

2.7 ml of 2-Mercaptoethanol

Distilled H_2O to 1 L

1 M Na_2CO_3 (1 L)

106 g Na_2CO_3

Distilled H_2O to 1 L

ONPG (O-nitrophenyl-β-D-galactopyranoside) (100 ml)

400 mg ONPG

Z buffer to 100 ml

The 100 ml ONPG was divided in small aliquots, stored at 4^oC and used at room temperature.

2.4.4. Competent cells preparation buffers

TFB I buffer (125 ml)

0.368 g KoAc

1.24 g $MnCl_2 \bullet 4H_2O$

1.51 g RbCl

0.18 g $CaCl_2$

18.75 ml 100% glycerol

pH5.8 adjusted with 1 M Acetic acid

Distilled H_2O to 125 ml

Filter sterilized and incubated at room temperature

TFB II buffer

0.21 g Na MOPS

1.10 g $CaCl_2 \bullet 2H_2O$

0.121 g RbCl

15 ml 100% glycerol, pH 7

Distilled H_2O to 125 ml

Filter sterilized and incubated at room temperature

2.4.5. Gel electrophoresis

5x TBE (Tris/Borate/EDTA) buffer (1 L)

54 g Tris

27.5 g boric acid

20 ml of 0.5 M EDTA

pH8.3 adjusted by HCL

Distilled H_2O to 1 L

The working solution for agarose gels was 0.5x and for acrylamide gels was 1x.

2.4.6. Stock acrylamide solution (250 ml)

62.5 ml of 30% acrylamide (PROTOGEL) – 7.5% final concentration

12.5 ml of 5% glycerol – 5% final concentration

50 ml of 5x TBE – 1x final concentration

125 ml distilled H_2O

This was prepared in a 250 ml measuring cylinder and covered with Parafilm to be mixed by inversion. After mixing, the volume was transferred to a Duran bottle and covered by foil. The bottle then was labeled properly and stored at 4°C.

10% Ammonium Persulphate (APS) (100 ml)

1 g APS

100 ml distilled H_2O

The 100 ml was divided into small aliquots (600 µl each) and frozen at -20°C.

For large acrylamide gels

50 ml of 7.5% acrylamide

500 µl of 10% APS

75 µl TEMED

The TEMED and APS should be added just before pouring the gel.

For small acrylamide gels

10 ml of 7.5% acrylamide

100 µl of 10% APS

15 µl TEMED

The TEMED and APS should be added just before pouring the gel.

2.5. Promoter manipulation

2.5.1. Oligonucleotides design

My starting promoter was KAB-TTTG (Burr *et al.*, 2000), which carries a -35 element of TAGACA (consensus is TTGACA), an extended -10 element of TTTG (consensus is TGTG) and a -10 hexamer of TATGGT (consensus is TATAAT). This promoter is designed to be located directly downstream of the *Eco*RI site and 6 bp before the transcription start site of the *lac* operon. The sequence of this promoter was manipulated in different ways (one, two or three mutations each time) to get eight more related promoters with different activities. Five of these new promoters were designed by designing an oligonucleotide (Alta Biosciences, University of Birmingham) carrying the required sequence of each promoter. Four promoters, including the starting promoter, were obtained from two previous studies (Burr *et al.,* 2000; Chan and Busby, 1989). The sequences (5'➔3') of the oligonucleotides used to design the new five promoters are shown in Table 1.

2.5.2. Conventional polymerase chain reaction

Polymerase chain reaction (PCR) was performed on the pRW50 plasmid DNA containing our starting promoter (KAB-TTTG) as a template to change its sequence. A single 100 µl reaction was prepared for each promoter to be designed. Each reaction contained one of the newly designed oligonucleotides along with another reverse oligonucleotide

Promoter name	Oligonucleotide sequence	Oligo code
KAB-TTTG	5' GAATTCTAGACAGCTGCATGCATCTTTGTTATGGTTATTTC 3'	-
BKS103	5' TAGATCTGAATTCTAGACAGCTGCATGCATCTTTGTTATAATTATTTCATACC 3'	D83457DL
BKS104	5' TAGATCTGAATTCTTGACAGCTGCATGCATC 3'	D83149BS
BKS106	5' TAGATCTGAATTCTAGACAGCTGCATGCATCTTTTTTATGGTTATTTCATACC 3'	D83161DL
BKS108	5' TAGATCTGAATTCTTGACAGCTGCATGCATCTGTGTTATAATTATTTCATACC 3'	D83151BS
BKS109	5' TAGATCTGAATTCTACTGTGCTGCATGCATCTTTTTTATAATTATTTCATACC 3'	D83152BS

Table1. **Oligonucleotides used to design the new promoters.**

This table shows the sequences (5'➔3') of the oligonucleotides used to design the new promoters with the short names of the promoters to be designed. It also shows the sequence of the starting promoter (KAB-TTTG). The -35 elements are shown in blue, the extended -10 elements are shown in green (bases -16 and -17) and yellow (bases -14 and -15) and the -10 elements are shown in red.

(D10527) binding downstream of the promoter area in a PCR reaction. The contents and thermal conditions of each PCR reaction are shown in Table 2 and Table 3, respectively. PCR products were then purified using the QIAquick PCR Purification Kit (QIAgen) according to manufacturer's instructions and using buffers supplied with the kit. This step was performed to dispose of the extra PCR reagents. The final product was eluted in 84 μl of distilled H_2O.

2.5.3. Digestion of the newly constructed DNA with restriction enzymes

All PCR products, obtained using the newly designed oligonucleotides, were intended to contain recently constructed promoters to replace the starting promoter (KAB-TTTG). As these promoters are located between the *Eco*RI and *Hin*dIII restriction sites, high fidelity *Eco*RI (HF *Eco*RI) and HF *Hin*dIII restriction endonucleases were used to digest all PCR products in preparation to insert them into the pRW50 plasmid DNA. Table 4 shows the volume of each component used in each digestion reaction. Digestion reactions were then incubated at 37°C for three hours, cleaned up using the QIAquick PCR Purification kit (QIAGEN) according to manufacturer's instructions and eluted in 30 μl of elution buffer (EB).

2.5.4. Large acrylamide gel for digestion products

A large acrylamide gel was prepared to run the digestion products of the DNA. This was

Component	Volume
Template DNA (pRW50)	5 μl
10 μM new designed forward primer	10 μl
10 μM reverse primer (D10527)	10 μl
10x buffer	10 μl
5 mM dNTPs	2 μl
Accuzyme (last solution to add)	2 μl
H_2O	61 μl
Total	100 μl

Table2. **Conventional PCR reaction components.**

This table illustrates the contents of each (100 μl) conventional PCR reaction.

PCR step	Temperature	Time	Repeat
Initialization step	95°C	5 minutes	1
Denaturation step	95°C	30 seconds	35
Annealing step	55°C	30 seconds	
Extension step	68°C	30 seconds	
Final extension step	68°C	10 minutes	1
Final hold	4°C	∞	--

Table3. **Conventional PCR thermal conditions.**

This table illustrates the thermal conditions of each conventional PCR run. It shows the temperature , time and repeat of each step of the PCR run.

Component	Volume
DNA (PCR product)	82 µl
Cut smart buffer	10 µl
HF *Eco*RI	4 µl
HF *Hin*dIII	4 µl

Table4. **Components of each digestion reaction.**

This table shows the components of each (100 µl) digestion reaction using HF *Eco*RI and HF *Hin*dIII.

done to find the *Eco*RI-*Hin*dIII band, cut it and extract the DNA from the gel slice by electroelution. Acrylamide gel glass plates were first cleaned with ethanol. Then, the spacers were greased on both sides and at the junctions between the side and bottom spacers. The bottom and both sides of the glass plates were clamped, making sure there was no gap between the side and bottom spacers. After assembly, the plates were placed diagonally on one side with the top lifted up to allow the pouring of the gel. The acrylamide gel was prepared, mixed gently by swirling and poured into the plate assembly continuously, avoiding air bubbles. Finally, the comb was inserted and clamped to allow the formation of several wells for the samples to be loaded into. The gel was then allowed to set for ~30 minutes and the comb was removed. To load the samples, the bottom spacer was removed and the plates were installed vertically in an electrophoresis apparatus. Next, 6 μl gel loading dye, blue (6x) (NEW ENGLAND BioLabs), and 3 μl of 50% glycerol were added to each digested DNA sample and loaded on the gel with 25 μl of HyperLadder 100 bp (BIOLINE). The gel was run at 40 mA for 3 to 4 hours (depending on the size of the fragments). At the end of the run, the gel was stained in ethidium bromide for 15 minutes. The gel was then removed from the ethidium bromide stain and laid flat on a UV light box after cleaning with ethanol. After identifying the *Eco*RI-*Hin*dIII band, the band was cut out with a clean razor blade and placed into a clean, labelled tube.

2.5.5. Electroelution

Electroelution was performed to extract the *Eco*RI-*Hin*dIII DNA fragments from the acrylamide gel slices obtained previously. A piece of dialysis tubing with a clip attached

to one end was washed thoroughly with distilled H_2O, and 200 μl of EB buffer was added into it. The gel slice was then transferred with a small clean spatula to the dialysis tube. After that, another clip was attached to the other end of the tube to close it, taking care to eliminate air bubbles. The dialysis tube was then placed into an electroelution tank filled with 0.1x TBE buffer (Tris base, boric acid and ethylenediaminetetraacetic acid [EDTA]) with the gel slice pushed to one side of the tube leaving, space for DNA to migrate into. The dialysis tube was then weighed down into the TBE buffer using a glass slide and electroelution was started at 30 mA for 30 minutes. The dialysis tube was then removed, one cap was taken off and the liquid was transferred from the dialysis tube into a clean, labelled eppendorf tube. The dialysis tube was washed with 200 μl of EB and the washing was added to the sample. To obtain the DNA, 400 μl of phenol:$CHCl_3$ was added to the sample and the sample was vortexed for 10 seconds. Sample was then centrifuged at 13000 rpm for 5 minutes. After centrifugation, three layers appeared in the tube: the upper aqueous phase containing DNA, the white interface and the phenol:$CHCl_3$. The upper aqueous layer was removed into a fresh tube taking care to not pipette any phenol. Next, 1/10 of the sample volume of Na acetate (pH 5.2) and two volumes of the final mixture of ice-cold ethanol were added to the sample. To help precipitate the DNA, 1 μl of glycogen was added and mixed well. The mixture was then stored at -20°C overnight or at -80°C for ~1 hour. Next, the mixture was centrifuged at 4°C and 13000 rpm for 15 minutes. The supernatant was poured off, leaving a white pellet behind, and 1 ml of ice-cold 70% ethanol was added without mixing followed by centrifugation at 4°C and 13000 rpm for 10 minutes. The supernatant was poured off and the pellet was dried using a Spedivac for approximately 12 minutes. The pellet was

then resuspended in 25 μl of sterile distilled H_2O. The final product included the *Eco*RI-HindIII DNA fragments containing the new designed promoters (insert DNA).

2.5.6. Small acrylamide gel

In a subsequent step, 1 μl of the final product was run on a mini acrylamide gel to determine the quantity of the prepared insert DNA. The previously described polyacrylamide gel procedure was performed again, but small acrylamide gel glass plates were used in this step. In addition, the gel was checked on the UV light box, without cutting the bands, to determine the quantity of the insert DNA.

2.5.7. pRW50 EcoRI-HindIII vector preparation

The pRW50 vector was prepared to be ligated with the previously prepared inserts.

First, pRW50 was extracted from the *E. coli* RLG221 strain using QIAfilter Plasmid Maxi Kit (QIAGEN). The extraction was performed according to the manufacturer's instructions using the buffers supplied with the kit. Next, 50 μl of purified pRW50 was mixed with 10 μl of CutSmart buffer (NEW ENGLAND BioLabs) and 34 μl of H_2O. Then, 2 μl of the mixture was removed and deposited into a clean PCR tube and labelled as (A). The mixture was then divided into two clean PCR tubes, adding 47 μl to each. Next, 3 μl of *Eco*RI restriction enzyme was added to the first tube and 3 μl of *Hin*dIII restriction enzyme was added to the other tube. Then, 2 μl of each new mixture was transferred into two clean PCR tubes and labelled as (B) and (C), respectively. The two mixtures were then mixed together and another 2 μl was removed and deposited into a fresh PCR

tube and labelled as (D). The final mixture and all of the 2 μl-samples (A-D) were incubated at 37°C for 3 hours to allow the enzymes to cut the pRW50 plasmids at the *Eco*RI and *Hin*dIII restriction sites. After the incubation, aliquots A through D were ran on a 0.8 % agarose gel to verify that the digestion had worked. After digestion, 4 μl of CIP (commonly known as CAP; calf alkaline phosphatase) was added to the sample and incubated at 37°C for one hour to dephosphorylate the DNA ends. Then, the sample was put through the QIAquick PCR Purification Kit (QIAGEN) to remove the enzymes and buffer. The DNA was then eluted off the column with 30 μl of sterile distilled H_2O and 1 μl was run through a gel to check quality and quantity. After verifying that the quality and quantity of the DNA was adequate, the vector was ready to be ligated with the insert DNA.

2.5.8. Vector-Insert ligation

The DNA inserts and vector plasmids were cut using *Eco*RI and *Hin*dIII restriction enzymes. Therefore, when ligated, the insert would be located between those two restriction sites. Three 20 μl ligation reactions were prepared in 0.5 ml tubes for each promoter sample. One ligation reaction containing the vector plasmid only as a negative control and two reactions containing two different concentrations of the insert DNA were prepared for each promoter sample. These ligation reactions were set up as shown in Table 5. Reaction mixes were then incubated at 16°C overnight.

	Reaction		
Component	1 (Control)	2	3
Vector	0.5 μl	0.5 μl	0.5 μl
Insert	-	0.5 μl	1 μl (1/50)
10x buffer	2 μl	2 μl	2 μl
H_2O	17 μl	16.5 μl	16 μl
T4 DNA Ligase	0.5 μl	0.5 μl	0.5 μl
Total	20 μl	20 μl	20 μl

Table5. **Components of the ligation reactions.**

This table illustrates the components of each ligation reaction prepared to ligate DNA inserts into pRW50 vectors. The components shown are for the three different reactions prepared for each promoter: one negative control reaction and two reactions containing different concentrations of DNA inserts.

2.5.9. Transformation

All ligation reactions (plasmids containing new promoters) were transformed into competent cells of a standard lab cloning *E. coli* strain (M182 in this study). For transformation, 20 μl of ligated plasmids were mixed with 200 μl of competent cells (RLG221) and chilled on ice for 60 minutes. A heat shock was introduced by putting the samples in a 42°C heat block for 2 minutes and returning them back to ice for five minutes. Then, 1 ml of 2x LB media was added and samples were incubated in a 37°C shaker for one hour. After that, the samples were centrifuged for 1 minute at full speed using a benchtop centrifuge and the supernatant was removed, leaving 100 to 200 μl. An empty vector (PRW50 with no inserts) was transformed into cells to be used as a negative control. Next, the samples were plated onto tetracycline-supplemented MacConkey agar plates and incubated at 37°C overnight; growth was checked the next day.

2.6. Plasmid DNA extraction

All of the plasmids were extracted from the *E. coli* M182 cells to be sent for sequencing as means of confirming successful transformation. Plasmids were extracted using the ISOLATE Plasmid Mini Kit (BIOLINE). This was performed according to the manufacturer's instructions and using buffers provided with the kit.

2.7. Sequencing

Sequencing was performed to confirm that all of the plasmids contained the correct promoter sequences. The Functional Genomics and Proteomics Laboratory (University of Birmingham, UK) carried out the sequencing. In preparation for sequencing, 7 μl of the plasmid miniprep was mixed with 3 μl of 1 μM sequencing primer. Two tubes were prepared for each sample. Each tube contained one of the primer pairs.

2.8. Rubidium chloride-treated competent cells

E. coli M182 competent cells were prepared for transformation with the plasmids constructed in this study containing the newly designed promoters. To prepare competent cells, a 5 ml overnight culture of the desired strain was grown. Then, 1 ml of the overnight culture was added to 100 ml of drug-free 2x LB and grown with shaking at 37°C till the OD_{650} reached approximately 0.5. After that, the 100 ml culture was divided into two 50 ml falcon tubes and chilled on ice for 20 minutes. Cells were harvested by centrifugation in clinical centrifuge at 4000 rpm for 15 minutes and the supernatant was poured off. Each cell pellet was resuspended in 15 ml of TFB I and shaken on ice for another 20 minutes. Cells were harvested again and the supernatant was discarded. Cell pellets were gently resuspended in 2 ml of TFB II and kept on ice for 30 minutes. Aliquots of 100 μl were prepared in 1.5 ml eppendorf tubes and kept at -70°C.

2.9. Agarose gel electrophoresis

Agarose gel electrophoresis was used to verify that the PCR reaction was successful, determine the sizes of the DNA fragments against a DNA ladder and assess DNA quality and quantity. A 1% agarose gel was prepared by dissolving 1 g of agarose powder in 100 ml of 0.5x TBE buffer. The mixture was then poured into a specific gel-casting tray and a comb was inserted in order to create several wells. Once solidified, the gel was transferred to a horizontal electrophoresis tank containing 0.5x TBE buffer. Then, 5 μl of sample was mixed with 2 μl of gel loading dye and the samples were loaded into their corresponding wells; a DNA ladder of choice was used to assess fragment size. The run was carried out at ~80 to 90 volts to allow clear separation of the ladder bands.

2.10. Promoters used in this study

Different promoters were used in this study. These promoters were designed to demonstrate different strengths to provide gradual decreases/increases in their activities and fused to the *lac* operon. All changes were applied to the -10, extended -10 and -35 elements. The sequences of these different promoters, from the *Eco*RI site to the -1 base, are shown in Table 6.

2.11. Primers used

Two sets of qPCR primers were used in this study. The first primer set was designed for

Names used in present study	Full name	Promoter sequence	Source
Consensus	Consensus	TTGACA TGnTATAAT	-
BKS101	KAB-TTTG	5' GAATTC**TAGACA**GCTGCATGCATC**TTTG**T**TATGGT**TATTTC 3' *Eco*RI -5 -1	Burr *et al.,* 2000
BKS102	KAB-TTTG-p16G	5' GAATTC**TAGACA**GCTGCATGCATC**TG**TG**TATGGT**TATTTC 3' *Eco*RI -5 -1	Burr *et al.,* 2000
BKS103	KAB-TTTG-p8A-p9A	5' GAATTC**TAGACA**GCTGCATGCATC**TTTG**T**TATAAT**TATTTC 3' *Eco*RI -5 -1	This study
BKS104	KAB-TTTG-p34T	5' GAATTC**TTGACA**GCTGCATGCATC**TTTG**T**TATGGT**TATTTC 3' *Eco*RI -5 -1	This study
BKS105	KAB-TTTG-p12C	5' GAATTC**TAGACA**GCTGCATGCATC**TTTG**T**CATGGT**TATTTC 3' *Eco*RI -5 -1	Chan and busby, 1989
BKS106	KAB-TTTG-p14T	5' GAATTC**TAGACA**GCTGCATGCATC**TTTT**T**TATGGT**TATTTC 3' *Eco*RI -5 -1	This study
BKS107	KAB-TTTG-p14T-p34T	5' GAATTC**TTGACA**GCTGCATGCATC**TTTT**T**TATGGT**TATTTC 3' *Eco*RI -5 -1	Chan and busby, 1989
BKS108	KAB-TTTG-p8A-p9A-p16G-p34T	5' GAATTC**TTGACA**GCTGCATGCATC**TG**TG**TATAAT**TATTTC 3' *Eco*RI -5 -1	This study
BKS109	KAB-TTTG-p8A-p9A-p14T-p30T-p31G-p32T-p33C	5' GAATTC**TACTGT**GCTGCATGCATC**TTTT**T**TATAAT**TATTTC 3' *Eco*RI -5 -1	This study

Table6. **Sequences of the semisynthetic promoters used in this study.**

This table shows the sequence of all the semisynthetic promoters used in this study (5'➔3') from *Eco*RI to base -1. The -35 elements are shown in blue, the extended -10 elements are shown in green (bases -16 and -17) and yellow (bases -14 and -15) and the -10 elements are shown in red. The double-underlined base is the base mutated in the starting promoter to obtain the new promoter.

the first system, which investigated RNAP behavior reflected by different promoters and two different regions (lac1357 and lac2720) of the following *lacZ* gene on pRW50 (Table 7). The promoter and control probes were named according to the DNA region they amplify. However, lac1357 and lac2720 were named according to the distance between the transcription start site and the start of the DNA region they bind to.

The other set of primers was designed to investigate RNAP behavior on the *lac* operon on the chromosome of the *E. coli* MG1655 strain (Table 8). The lac0 and control A primer pairs amplify the promoter region of the *lac* operon and a silent control region on the chromosome, respectively. The lac1, 2, 3, 4 and 5 primer pairs amplify ~300 bp DNA sequences located 518, 1421, 2308, 3691 and 4654 base pairs downstream of the transcript start, respectively.

2.12. Reporter gene assay

The promoters' activities were estimated by measuring β-galactosidase levels as a preliminary test (Miller, 1972). This was done to investigate the differences between the promoters' activities according to the differences between the sequences of their elements. This experiment was performed in triplicate for each sample: three different overnight cultures were prepared for each sample. The next day, 50 μl of each of these overnight cultures was inoculated into tetracycline-supplemented 5 ml 2x LB media in 25 ml flasks. The flasks were then incubated at 37°C with shaking until the OD650 reached 0.3 to 0.6. After that, a drop each of toluene and 1% sodium deoxycholate was

Primer Code	Target	Sequence (5’ to 3’)
D78382DL	Promoter (Fwd)	5’-TCG—CAA—GGA—CGA—GAA—TTT—CC-3’
D78383DL	Promoter (Rev)	5’-GTT—TTT—CAG—CAG—GTC—GTT—G-3’
D78261DL	lac1357 (Fwd)	5’-GAC—AGT—ATC—GGC—CTC—AGG—AA-3’
D78262DL	lac1357 (Rev)	5’-AAC—GTC—GTG—ACT—GGG—AAA—AC-3’
D78263	lac2720 (Fwd)	5’-TCT—CTC—CAG—GTA—GCG—AAA—GC-3’
D78265DL	lac2720 (Rev)	5’-TAA—TCA—CGA—CGC—GCT—GTA—TC-3’
D78254DL	Control A (Fwd)	5’-CAG—TCC—ATC—AGG—TAA—TTG—CCG-3’
D78255DL	Control A (Rev)	5’-GCG—CAA—ACT—GTT—AAT—GCT—GG-3’

Table7. **Probes used in qPCR experiment to measure RNAP flux on the *lac* operon.**

This table shows the codes, targets and sequences (5’➔3’) of the probes used to quantify the immunoprecipitated DNA to test RNAP flux on the *lac* operon in pRW50 vectors.

Primer code	Region	Sequence (5’ to 3’)
D82715BS	lac0 (Fwd)	5’-GCG—TTG—GCC—GAT—TCA—TTA—ATG—CAG—C-3’
D82716BS	lac0 (Rev)	5’-CAA—GGC—GAT—TAA—GTT—GGG—TAA—CGC—CAG-3’
D82717BS	lac1 (Fwd)	5’-GGT—TAC—GGC—CAG—GAC—AGT—CG-3’
D82718BS	lac1 (Rev)	5’-CAC—ATC—TGA—ACT—TCA—GCC—TCC—AGT—AC-3’
D82719BS	lac2 (Fwd)	5’-GAA—TCA—GGC—CAC—GGC—GCT—A-3’
D82720BS	lac2 (Rev)	5’-AAC—CGC—CAA—GAC—TGT—TAC—CCA—TC-3’
D82721BS	lac3 (Fwd)	5’-GCA—ATT—TAA—CCG—CCA—GTC—AGG—C-3’
D82722BS	lac3 (REV)	5’-TGA—TGC—TGC—CAC—GCG—TGA—G-3’
D82767BS	lac4 (Fwd)	5’-TGC—ACT—CAT—CCT—CGC—CGT—TTT—ACT—C-3’
D82768BS	lac4 (Rev)	5’-CCA—AAT—ACC—CGC—GTA—CCC—TGT—TC-3’
D82725BS	lac5 (Fwd)	5’-GTA—GGG—GAA—AAC—GCC—TGG—GTA—G-3’
D82726BS	lac5 (Rev)	5’-CCA—CAT—GAC—TTC—CGA—TCC—AGA—CG-3’
D80465BS	Control bglB (Fwd)	5’-GGA—TCA—AGC—CGC—TGG—TAA—CG-3’
D80466BS	Control bglB (Rev)	5’-CTG—GCA—ACC—AGT—TGA—TGG—TGG—A-3’

Table8. **Probes used in qPCR experiment to measure RNAP flux on the *lac* operon on the chromosome.**

This table shows the codes, targets and sequences (5’➔3’) of the probes used to quantify the immunoprecipitated DNA to test RNAP flux on the *lac* operon on the chromosome. lac0 primer pair probes the promoter region, lac1-5 primer pairs probes the five regions to be tested through the *lac* operon and the bglb primer pair probes a silent control region on the chromosome.

added to lyse the cells. Flasks were vigorously shaken and incubated again at 37°C for 20 minutes with the bung removed. Then, 100 μl of lysed cells were added to 2.5 ml of pre-warmed (37°C) Z buffer containing ONPG at 20 seconds intervals in assay tubes. Cultures were watched, and when their colour turned yellow, the reaction was stopped by adding 1 ml of Na_2CO_3 and the time was noted. OD_{420} measurements were taken for all reactions.

β-galactosidase (β-gal) activity, for each promoter, was calculated using the following equation:

$$\text{β-gal activity} = \frac{1000 \times 2.5 \times 3.6 \times OD_{420}}{OD_{650} \times 4.5 \times t \times v}$$

Where:

2.5 = conversion factor of OD_{650} into bacterial mass, based on an OD_{650} of 1 being equivalent to 0.4 mg/ml bacteria (dry weight).

3.6 = final assay volume (ml).

1000/4.5 = conversion factor of OD_{420} into nmol o-nitrophenyl (ONP), based on 1 nmol ml^{-1} of ONP having an OD_{420} of 0.0045.

t = incubation time (min).

v = volume of lysate added (ml).

2.13. Chromatin immunoprecipitation

The general concept of a ChIP experiment is to crosslink proteins to DNA, shear the DNA by sonication, pull down specific protein-DNA complexes by using specific antibodies and beads, de-crosslink proteins from DNA and purify the resulting DNA. The result of these steps is getting DNA fragments that were attached to our protein of interest (i.e., RNAP) only. The whole ChIP-qPCR experiment was conducted at least three times for each sample as biological repeats.

2.13.1. Cell preparation

A 5 ml culture of the desired strain was grown in 2x LB media supplemented with tetracycline overnight. The OD_{650} was taken and the amount required to be inoculated in 40 ml 2x LB to provide a starting OD_{650} of 0.03 was calculated. Two 40 ml portions of 2x LB media containing tetracycline were inoculated with the bacteria and incubated at 37°C with shaking until the OD_{650} reached ~0.4. Next, 40 µl of rifampicin was added (50 µM final concentration) for 15 minutes to one of the two cultures. However, for analysing RNAP flux through the *lac* operon, the whole ChIP-qPCR experiment was performed twice: one time before and after IPTG addition and another time before and after rifampicin treatment.

2.13.2. Crosslinking

To crosslink RNAPs to DNA, 1.11 ml of 37% formaldehyde (1% final concentration) was added to all cultures and incubated at 37°C with shaking for 20 minutes exactly to avoid excessive crosslinking (Grainger *et al.*, 2005; Herring *et al.*, 2005). After that, 5.33 ml of 3 M glycine (0.5 M final concentration) was added to all cultures and mixed by swirling to stop formaldehyde action. Cultures were then immediately transferred to 50 ml falcon tubes and cells harvested by centrifugation at 4000 rpm and 4°C for 15 minutes. The supernatant was removed from each tube, cells were resuspended in 20 ml of TBS and harvested by centrifugation as before. The supernatant was then removed again and the cells were resuspended in 1 ml of TBS and transferred to 1.5 ml eppendorf tubes. Next, the cells were harvested by centrifugation using a desktop microcentrifuge for 1 minute at maximum speed (~13000 rpm). The supernatant was removed, the pellet was resuspended in 1 ml of FA lysis buffer containing 4 mg/ml lysozyme and transferred to a 15 ml falcon tube. The tubes were then incubated at 37°C for 30 minutes and chilled on ice for >5 minutes. The cells then were sheared through three rounds of sonication. Each round consisted of 10 cycles of 30-second shearing and 30-second resting without shearing. Sonication was performed in ice-cold water to obtain the majority of fragments in the 300 to 400 bp size range. Then, lysates were cleared by spinning in a microcentrifuge for 5 minutes at top speed. After centrifugation, the supernatant containing chromatin (~1 ml) was transferred to a 2 ml microfuge tube and diluted in FA lysis buffer so that there was 1 ml of lysate for each 20 ml of starting culture (for 40 ml, dilute chromatin to a total of 2 ml). These tubes were labeled as input chromatin.

2.13.3. Protein A Sepharose (beads) preparation

Protein A Sepharose (GE Healthcare) beads were washed and a 50% slurry was prepared to immunoprecipitate the resulting chromatin. A total of 300 mg of Protein A Sepharose beads were washed four times in a 15 ml falcon tube by gently re-suspending in 15 ml distilled H_2O, slow spinning at 400 rpm for 5 minutes and carefully discarding the supernatant. After the fourth wash, the beads were resuspended in an equal amount of TBS to achieve a 50% slurry.

2.13.4. Immunoprecipitation

The immunoprecipitation reaction was prepared by mixing 500 µl of lysate (input chromatin), 300 of µl FA lysis buffer, 25µl of freshly washed 50% slurry beads and 1 µl of *E. coli* RNAP beta monoclonal (neoclone) in a clean 1.5 ml eppendorf tube for each sample. The lid was securely closed, wrapped with Parafilm and the tubes were rotated on a wheel for 90 minutes at room temperature. After 90 minutes of rotation, the beads were collected by centrifugation at 3000rpm for 1 minute using a desktop microcentrifuge and the supernatant was discarded carefully. The beads were resuspended gently in 700 µl of FA lysis buffer and transferred to a Spin-X column (Costar). The column was then rotated for 3 minutes at room temperature, centrifuged at 3000 rpm for 1 minute and the flow-through was discarded. The column was washed by adding 750 µl of FA lysis buffer, rotating for 3 minutes, centrifuging as before and the flow-through was again discarded. The wash step was repeated three times with three different buffers: FA lysis buffer 500 mM NaCl, ChIP wash buffer and TE buffer. After washing, the Spin-X column was transferred to a clean eppendorf tube, 100 µl of ChIP

elution buffer was added, the tube was rotated for 3 minutes at room temperature and incubated at 65°C for 10 minutes. After incubation, the clean eppendorf tube was centrifuged at 3000 rpm for 1 minute, the flow-through was kept and the column was discarded. The tubes were then labeled as IP samples.

2.13.5. De-crosslinking

Crosslinking was reversed by incubating the IP samples at 100°C for 10 minutes followed by cooling to 4°C. Then 20 μl of the input chromatin was mixed with 80 μl of FA lysis buffer and de-crosslinked by heating to 100°C for 10 minutes followed by cooling to 4°C. After de-crosslinking, the DNA was purified using a QIAquick PCR Purification Kit (QIAgen) according to the manufacturer's instructions and using the buffers supplied with the kit. The purified DNA was then eluted in either 50 μl of distilled H_2O for IP samples or 200 μl of distilled H_2O for input samples.

2.14. Quantitative Polymerase Chain Reaction

qPCR was the technique of choice to detect the presence of and quantify copies of DNA of interest in our samples after each ChIP experiment. qPCR and conventional PCR share the same principle of amplifying DNA following heating and cooling cycles. However, using a fluorescent reporter (2x Brilliant III SYBR Green qPCR Master Mix from Agilent Technologies in this study) and specific designated primers in qPCR made it easy to monitor and quantify the amount of DNA of interest generated in each qPCR cycle allowing the comparison of DNA amounts before and after rifampicin treatment in order

to calculate occupancies (Wong and Medrano, 2005; VanGuilder *et al.*, 2008; Postollec *et al.*, 2011). Each qPCR reaction was prepared by mixing 10 µl of 2x Brilliant III SYBR Green qPCR Master Mix (Catalogue no. 600882-51), 1 µl of each primer, 1 µl of DNA after the ChIP experiment and 7 µl of distilled H_2O. The reaction mix was prepared in a specific qPCR 96 well plate. The thermal cycling program is shown in Table 9.

2.15. RNAP occupancy unit calculations

Occupancies were calculated depending on the cycle thresholds (Cts) obtained from the qPCR tests for ChIP products. Each qPCR reaction was performed in three technical repeats and averages were taken. In each qPCR run, each region occupancy, was calibrated against the occupancies obtained from the input DNA (chromatin before Immunoprecipitation). Occupancy units were calculated using the comparative C_T method (Livak and Schmittgen, 2001) based on the following formula:

Occupancy unit = $2^{-\Delta\Delta C_T}$

Where:

$\Delta\Delta C_T = \Delta C_{T\ target} - \Delta C_{T\ calibrator}$

ΔC_T = Average $C_{T\ target}$ – Average $C_{T\ control}$

Where:

C_T control = C_T obtained from the control region

Calibrator is the input DNA (chromatin)

Number of cycles	Duration of cycle	Temperature
1	3 minutes	95°C
35	20 seconds	95°C
	30seconds	56°C
	30 seconds	60°C
1	1 minute	95°C
	30 seconds	56°C
	30 seconds	95°C

Table9. **Cycling program used for QPCR experiments**

This table shows the number of cycles and the duration and temperature of each qPCR cycle.

2.16. Calculation of promoter occupancy index

POI was calculated by calculating the percentage of promoter occupancy before rifampicin treatment (measured promoter occupancy) in relation to the promoter occupancy after rifampicin treatment (assumed to be 100% occupied) using the following equation:

$$POI = \frac{\text{Measured promoter occupancy}}{\text{Promoter occupancy after rifampicin treatment}} \times 100$$

2.17. Calculation of promoter escape index

EI was calculated by calculating the percentage of downstream region occupancy in relation to measured promoter occupancy without rifampicin using the following equation:

$$EI = \frac{\text{Region occupancy by RNAP}}{\text{Measured promoter occupancy}} \times 100$$

2.18. Calculation of fragment occupancy percentage

FOP was calculated by calculating the percentage of the measured fragment occupancy from the occupancy after rifampicin treatment of the promoter taken to be 100% occupied as follows:

$$FOP = \frac{\text{Measured fragment occupancy}}{\text{Occupancy after rifampicin treatment of a promoter taken to be 100\% occupied}}$$

2.19. Calculation of polymerases per second

PoPS at a particular gene region was calculated using the following equation:

$$PoPS = \frac{1}{\text{Time interval between RNAPs arriving at the start of the region } (T_{int})}$$

Where:

T_{int} = time taken by RNAP to cross the region of interest (seconds) X n

Where:

The time taken by RNAP to cross the region

= length of the region of interest / RNAP speed (taken as 50 bp/second) (Murakawa *et al.*, 1991; Vogel and Jensen, 1997).

And

$$\frac{1}{n} \times 100 = \text{percentage of region occupancy to measured promoter occupancy}$$

Chapter 3

Measurements of RNAP flux through the *lac* operon on the chromosome

3.1. Introduction

Since its discovery, many aspects of the *lac* operon have been thoroughly studied. These different but related aspects include induction, repression (Müller-Hill *et al.,* 1964; Eron and Block, 1971), regulation (Jacob and Monod, 1961) and even fusions to other genes of interest to understand different biological systems (Silhavy and Beckwith, 1985). The *lac* operon consists of three adjacent genes: *lacZ*, *lacY* and *lacA*. As mentioned previously, these genes are responsible for lactose metabolism as a source of energy. The full length of the *lac* operon is approximately 5300 bp, including the promoter, operator and termination regions (Hediger *et al.,* 1985). The *lacZ* gene is 3069 bases in length and responsible for the production of the β-galactosidase enzyme, which breaks lactose into glucose and galactose (Kalnins *et al.,* 1983); the *lacY* gene is composed of 1251 nucleotides and responsible for the lactose permease production, a membrane enzyme that helps in the lactose uptake (Büchel *et al.,* 1980); and the *lacA* gene is 609 bp long and codes for β-galactoside transacetylase, which helps in bacterial cell detoxification (Andrews and Lin, 1976; Hediger *et al.,* 1985). All of these measurements were confirmed by the Ecocyc website ("*Escherichia Coli* K-12 Substr. MG1655 Chromosome: 360,000 - 370,000").

Transcription of the *lac* operon is negatively controlled by the LacI repressor protein. This protein works by binding to one of the three *lac* operators: operator 1 (O1), O2 and O3 (Fulcrand *et al.*, 2013). It has been recently reported that, when the LacI repressor protein binds to several O1 operators, it can change the topology of the DNA and act as a topological barrier to repress

transcription (Leng *et al.,* 2011). Another regulator for the transcription of the *lac* operon is cAMP-receptor protein (CRP). When glucose is absent and lactose is present, CRP becomes activated by cyclic adenosine monophosphate (cAMP) and activates the transcription of the *lac* operon.

Many previous studies have attempted to measure RNAP flux, velocity and distribution on bacterial operons. In 1998, a study was carried out to address RNAP speed, while progressing along a DNA fragment, under different regimes of force load applied by an optical trap (Wang *et al.,* 1998). Another study, in an attempt to address the regulation of RNA synthesis in *E. coli*, compared the growth rate of ribosomal and messenger RNA (mRNA) chains synthesized by RNAP. In this study, it can be assumed that the growth rate reflects RNAP speed (Dennis and Bremer, 1973). A different study calculated the number of RNAPs on the *lac* operon by calculating the rate of RNA synthesis (Kennel and Riezman, 1977). The researchers compared this number with the same number obtained from ribosomal RNA synthesis. They claimed that, in the *lac* operon, RNAP transcribes 40 nucleotides per second, but this rate increases to 80 nucleotides per second with the ribosomal RNA (rRNA) genes. Condon and colleagues mentioned the same increase in elongation rate between synthesizing mRNA and rRNA chains (Condon *et al.,* 1995).

The *lac* operon is one of the most extensively characterized operons in terms of function, transcription and regulation, making it a model system for bacterial genome studies. In this portion of the current study, I attempted to apply my newly developed method to directly measure RNAP flux through the *lac* operon.

As previously mentioned, transcription of the *lac* operon can be triggered by IPTG, which overcomes the LacI repression. This phenomenon allowed for easy

induction of *lac* operon transcription in this part of the study, which was performed to analyze RNAP flux through the *lac* operon on the chromosome using a ChIP–qPCR experiment and several probes to apply this new and direct method to specific fragments of the *lac* operon.

3.2. Method

To summarize, RNAP was crosslinked to bound DNA targets as a result of formaldehyde treatment of ~0.4 OD_{650} *E. coli* cultures. After crosslinking, DNA was then sheared by sonication and immunoprecipitated using mice monoclonal antibodies against the RNAP β subunit. These pulled down targets were then quantified using qPCR. In this study, we used the well-studied *E. coli* K-12 *lac* operon, which is controlled by a promoter whose activity is repressed by the Lac repressor protein (LacI). Transcription of this operon was induced by IPTG. Seven pairs of probes were designed to analyze RNAP flux through the *lac* operon. One of these pairs of probes (lac0) was designed to sample the *lac* promoter region and another one (bglB) was designed to sample a control region. The lac1-5 pairs of probes were designed to sample five different DNA sequences that are 518, 1421, 2308, 3691 and 4654 bp, respectively, downstream from the transcription start site (Figure 9). All seven pairs of probes were designed to sample ~300 bp fragments, allowing the direct comparison of the signals obtained from different probes. Then, the ChIP-qPCR experiment was applied to *E. coli* strain MG1655, growing to an OD_{650} of ~0.4. This experiment was applied twice to the *lac* operon. The first time was to test RNAP flux in IPTG-induced cells versus non-induced cells. This was performed to ensure that the Lac repressor regulates transcription of the *lac* operon and all signals seen in

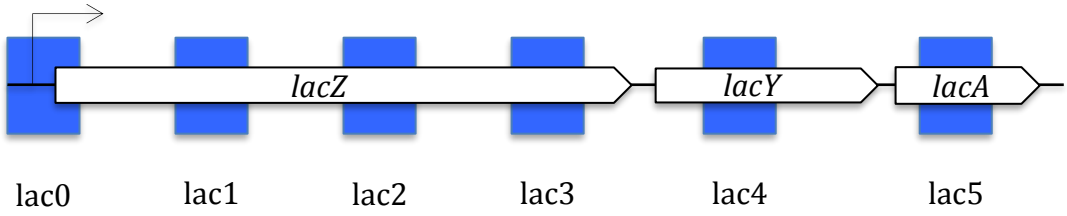

Figure 9. Approximate locations of the lac0-5 probes.

This figure shows the approximate locations of the lac0-5 primer pairs used to probe the five immunoprecipitated target regions of the *lac* operon on the chromosome.

both experiments are due to RNAP genuinely transcribing, and not due to other crosslinked components or proteins, as they are not seen in the control experiments performed without IPTG induction.

In the second instance, the same experiment was done but to test rifampicin-treated cells against non-rifampicin-treated cells where both cultures were treated with 100 μM IPTG to induce the *lac* expression.

3.3. Optimisation of the method

3.3.1. Rifampicin treatment time optimisation:

As weak promoters will not be as competitive as strong ones in recruiting RNAP, the 15 minutes rifampicin treatment time was tested for the weakest promoter in my promoters' hierarchy (BKS105). This will insure that the 15 minutes rifampicin treatment is enough for even weak promoters to be fully occupied with RNAP after rifampicin treatment. This test was carried out by growing the cells in five 40 ml 2x LB cultures to an OD_{650} of ~0.4. These five cultures were treated with rifampicin at a final concentration of 50 μM and incubated with shaking at 37°C. After that, formaldehyde was added to each culture at a time interval of five minutes, so the five cultures will be treated with rifampicin for 5, 10, 15, 20 and 25 minutes. The rest of the ChIP-qPCR experiment was performed normally to all cultures and promoter occupancy of each sample was calculated. The final result of this test confirmed that 15 minutes treatment of rifampicin was enough to block all recruited RNAPs at promoters by showing that promoter occupancy level gradually increased after 5, 10 and 15 minutes rifampicin treatment, however, this gradual increase did not continue with the samples treated for 20 and 25 minutes.

3.3.2. Sonication time optimisation:

This part of the experiment was mainly performed by Dr. Jack Bryant to ensure that the sonication time is suitable to produce the DNA fragments within the desired range of lengths (~300 – 700 bp). Six 1 ml lysates were processed in the Bioruptor sonicator and removed at six cycle numbers (10, 20, 25, 28, 30 and 32 cycle numbers). Sonication was carried out by putting 250 ml beaker of ice in the water bath then the bath filled with cold (4^{o}C) water to the line. After that, samples were applied and sonication was started on HIGH for 10 cycles of 30 seconds ON / 30 seconds OFF. Then, water and ice were replaced with fresh cold ones, as they will have heated up. These sonication steps were repeated three times and reflector bars were cleaned with 70% ethanol in the end. After sonication, samples were spun down, decrosslinked, purified using a QIAquick PCR Purification Kit (QIAgen) according to the manufacturer's instructions and concentrated using SpeedVac concentrator. After that, samples were run on agarose gel to test fragments sizes in comparison to 100 bp ladder loaded in a separate well on the same gel. This experiment suggested that 30 cycles of sonication were best to achieve DNA fragments within the desired length range.

3.4. Measurement of RNAP flux through the *lac* operon before and after induction

The ChIP–qPCR method was applied to the *E. coli* K-12 *lac* operon, with and without IPTG induction, to address the flux of RNAP through this operon. IPTG was added to a final concentration of 100 μM. The data obtained showed that

there was a >10-fold increase in levels of immunoprecipitated DNA after the addition of IPTG (Figure 10). This confirms that the Lac repressor regulates *lac* operon transcription. Recalling that, in this method, the amount of immunoprecipitated DNA reflects the amount of RNAP bound to the DNA in each immunoprecipitated fragment, RNAP levels were similar for approximately the first 2000 bp of the *lac* operon and then declined towards its end. This decline probably resulted from polarity effects. Furthermore, some fragments, such as lac3, showed an increase in the amount of immunoprecipitated DNA. This increase was likely due to some sort of pause sequences (Larson *et al.,* 2014; Vvedenskaya *et al.,* 2014). After this high amount of immunoprecipitated DNA in the lac3 fragment, there was a sizeable decline in fragments lac4 and lac5. This decline may be explained by a previous study that investigated a rho-dependent terminator in the intergenic space between the *lacZ* and *lacY* genes. However, the findings of the study claimed that this terminator is not effective in cells growing in normal conditions (Murakawa *et al.,* 1991).

3.5. Measurement of RNAP flux through the *lac* operon before and after rifampicin treatment

The same experiment was applied again to *E. coli* strain MG1655 under the same conditions, but this time the rifampicin property of blocking RNAPs at promoters was exploited (Grainger *et al.,* 2005). Trapping RNAPs at promoters means that each promoter is occupied by a single RNAP. Carrying out the ChIP-qPCR experiment, with and without rifampicin treatment, after IPTG induction, allowed for the calculation of the absolute numbers of RNAPs bound to each

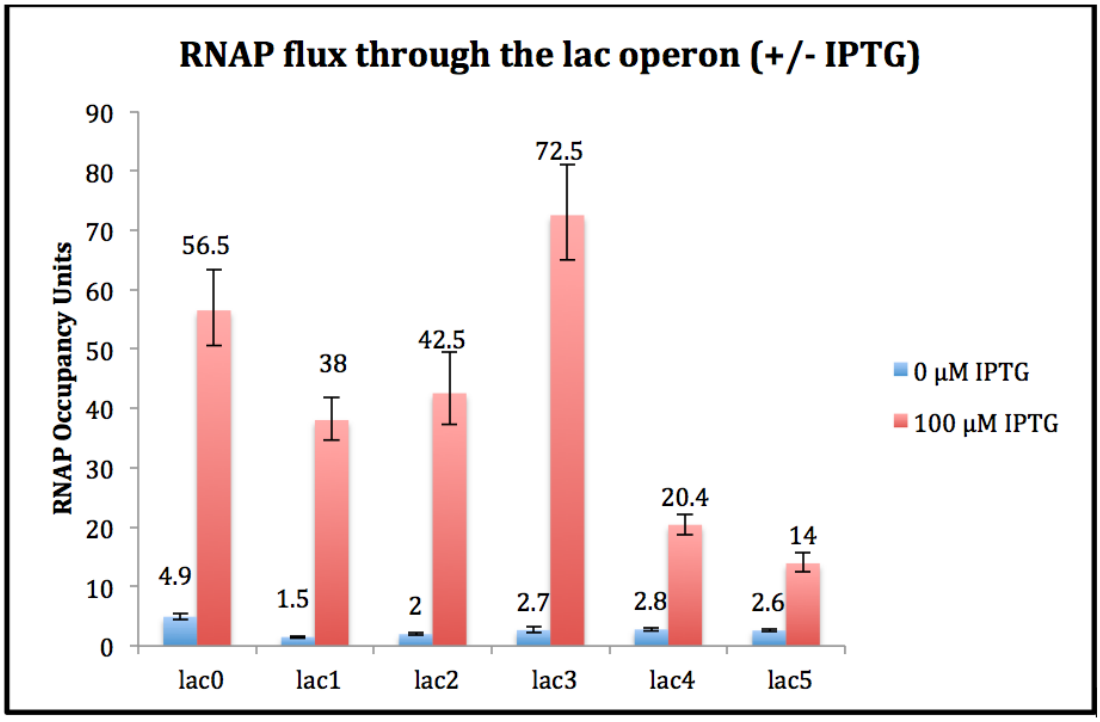

Figure 10. RNAP flux on the *lac* operon on the chromosome before and after IPTG treatment.

This figure shows the RNAP flux on the *lac* operon on the chromosome before (blue bars) and after (red bars) 100 µM IPTG treatment.

immunoprecipitated fragment of the *lac* operon. Figure 11 shows quantification of immunoprecipitated DNA corresponding to the lac0 pair of probes was increased after rifampicin treatment but significantly decreased corresponding to the lac1-5 probes. These data also permitted the calculation of the promoter occupancy index (POI) for the promoter region, the promoter's escape index (EI), fragment occupancy percentage (FOP), the average time interval between RNAPs (T_{int}) and how many RNAPs are passing through each fragment in one second (polymerases per second; PoPS).

As each promoter is occupied by one RNAP after IPTG induction and rifampicin treatment, an assumption can be made that, the promoter is 100% occupied and the POI is ~47.6% ((56.5/118.7) x 100); this signifies that the promoter is approximately 47.6% occupied. This promoter showed an EI value of approximately 67% ((38/56.5) x 100). For the lac1 region, the occupancy value was 38 units, which is ~32% ((38/118.7) x 100) i.e., 1/3.1 of the promoter full occupancy after rifampicin treatment. As RNAP takes roughly 6 seconds to traverse each region (300 bp/50 bp per second), the T_{int} for this region was 18.6 seconds (6 x 3.1) and PoPS for this region was 0.053 (1/18.6). The same calculations were applied to the other regions and all values are shown in Table 10.

3.6. Discussion

The *lac* operon consists of three adjacent genes with related functions: *lacZ*, *lacY* and *lacA*. These genes are responsible for lactose transport and metabolism in many bacteria, including *E. coli*. Expression of the *lacZ* gene allows for the

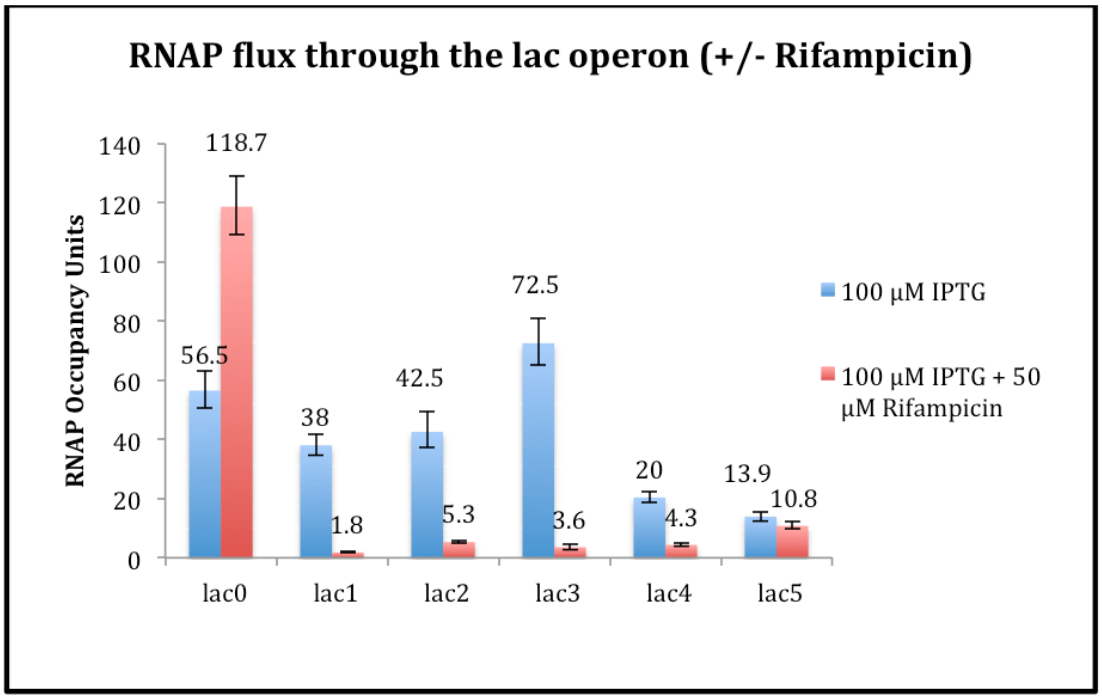

Figure 11. RNAP flux on the *lac* operon on the chromosome before and after rifampicin treatment.

This figure shows the RNAP flux on the *lac* operon on the chromosome with 100 µM IPTG treatment (blue bars) and with 100 µM IPTG + 50 µM rifampicin treatment (red bars).

Region	Occupancy value (unit)	FOP (%)	T_{int} (seconds)	PoPS
lac0	56.5	47.6	-	-
lac1	38	32	18.6	0.053
lac2	42.5	35.8	16.8	0.059
lac3	72.5	61	9.6	0.1
lac4	20	16.8	36	0.027
lac5	14	11.8	51	0.019

Table10. **Occupancy values and percentages, T_{int} and PoPS values for the *lac* operon on the chromosome before and after rifampicin treatment.**

This table illustrates the occupancy values, fragment occupancy percentages (FOP), T_{int} and PoPS values for the lac0-5 regions of the *lac* operon on the chromosome before and after 50 μM rifampicin treatment.

production of the β-galactosidase enzyme, which breaks down lactose into glucose and galactose. Similarly, *lacY* expression produces a membrane protein called the lactose permease, which is responsible for lactose transportation into the cell (Halasz *et al.*, 2007). Finally, the *lacA* gene encodes for the production of the transacetylase enzyme, which helps in cellular detoxification (Andrews and Lin, 1976). The *lac* system has been thoroughly studied over the last few decades and has become a model system for bacterial biological systems including gene expression and regulation.

In this study, a new method was successfully developed to measure RNAP flux through bacterial operons. In this chapter, this method was applied to study RNAP flux through five regions along the well-characterized *E. coli lac* operon. We firstly applied the new method on the *lac* operon before and after IPTG treatment. Then, the same method was applied to IPTG-induced cultures but before and after rifampicin treatment.

More than a 100-fold increase in the amount of immunoprecipitated DNA was recorded after IPTG treatment from the induced versus non-induced cells. This increase confirms that the Lac repressor regulates transcription of the *lac* operon.

The results obtained from the second experiment measuring POI depending on occupancy with and without rifampicin treatment showed that the *lac* promoter was ~47.6% occupied. This confirms the findings of previous studies showing that the *lac* promoter in *E. coli* is limited by RNAP recruitment (Malan *et al.*, 1984; Buckle *et al.*, 1991). This phenomenon might be explained by the fact that the expression of the *lac* operon depends on recruitment of RNAP to the promoter, and this depends on the local concentration of free RNAP. This

dependency was shown in a previous study that tested the expression of the *lac* promoter::*gfp* fusion in different positions on the chromosome; the findings demonstrated different expression rates in different locations on the chromosome (Bryant *et al.,* 2014). This promoter also showed a low EI of approximately 67%, and there is roughly a 18.6-second gap between RNAPs escaping the promoter to the nearest tested region (lac1). This finding supports previous studies claiming that the promoter escape step is complex and may take some time (Chander *et al.,* 2015; Bauer *et al.,* 2016). Several explanations were suggested for promoter clearance delays involving abortive transcription initiation during DNA scrunching (Revyakin *et al.,* 2006) and σ^{70}-dependent transcription pausing (Strobel and Roberts, 2015).

The tested downstream fragments (lac1–lac5) showed different occupancy percentages, ranging from 11.8% in lac5 to 61% in lac3. These ranging occupancy percentages indicate that, during transcription elongation, RNAPs are not evenly spread through the *lac* operon, most likely because of pause sequences (Larson *et al.,* 2014; Vvedenskaya *et al.,* 2014). Furthermore, the occupancy percentage of each fragment is inversely proportional to the time gap between RNAPs appearing at the start of the same region (T_{int}) and parallel to PoPS value of the same region. For instance, the lac5 region has the lowest occupancy percentage of 11.8%, the longest T_{int} value of 51 seconds and the lowest PoPS value of 0.019. Similarly, the lac3 region has the highest occupancy percentage of 61%, the shortest T_{int} value of 9.6 seconds and the highest PoPS value of 0.1. As a result of polarity effects, signals obtained from *lacY* and *lacA* genes are less than those obtained from *lacZ* gene reflecting less transcripts and

less final products. This finding concurs with the findings of Grainger and his coworkers in 2005 (Grainger *et al.*, 2005).

Chapter 4

Manipulating the sequences of promoter elements affects RNAP recruitment and escape in different ways

4.1. Introduction

For DNA-dependent RNAP to transcribe specific genes, it initially needs to bind to specific regions of the DNA located directly before a gene's transcription start site known as promoters (Shultzaberger *et al.*, 2007). These promoters play an important role in transcription as they recruit DNA-dependent RNAP for the transcription of their corresponding genes. Each promoter has four conserved elements that are recognized by the RNAP holoenzyme: the -10, extended -10, -35 and UP elements (Busby and Savery, 2007; Mekler *et al.*, 2012). Promoter conserved elements are the main determinant of RNAP specificity to promoters. A promoter's activity or "strength" is often connected to its ability to recruit RNAPs (i.e., the stronger the promoter, the higher the affinity to RNAPs). This affinity is mainly determined by the sequences of the promoter's conserved elements mentioned above (i.e., the closer the sequence to the consensus, the higher the affinity to RNAPs) (Phan *et al.*, 2012). In this study, the sequences of three promoter elements were manipulated to achieve a gradual increase/decrease in promoter activity. These promoters (Table 6, page 74) were then inserted into the pRW50 plasmid in the K-12 *E. coli* M182 strain just before the well-characterized *lac* operon. In this part of the study, I used the first *lac* gene (*lacZ*) to test different transcriptional activities associated with the designed semisynthetic promoters.

The starting promoter was BKS101, which is a galp1 derivative: a fairly active promoter with a good -35 element holding a sequence of 5' TAGACA 3' (consensus is 5' TTGACA 3'), a fairly good extended -10 element with a sequence of 5' TTTG 3' (consensus is 5' TGTG 3') and a good -10 element with a sequence

of 5’ TATGGT 3’ (consensus is 5’ TATAAT 3’) (Table 11). The promoter sequence was successively manipulated to build a set of promoters with a gradual range of activity ranging from very high to very low. In some promoters, the sequence of each promoter element was changed separately to fully ascertain the effect of each change on the transcription process. Other promoters had two or even three mutations to expand the range of activity between promoters and get them closer to or more distant from consensus promoter sequences. Note that promoters of equal strengths can be built by combinations of different sequence elements.

The promoters’ activities were firstly assessed using a reporter gene assay estimating promoter activity by measuring β-galactosidase levels of the *lacZ* gene of each promoter. Then, the ChIP-qPCR method was applied to get a clear and direct (as in the cells) view of transcriptional behavior, such as POI and RNAP flux in the corresponding gene, EI, promoter competitivity or FOP, associated with each promoter. POI was estimated by calculating the percentage of promoter occupancy in relation to the full occupancy state of the promoter assumed to be 100% occupied after rifampicin treatment. Promoter competitivity was calculated from the percentage of promoter occupancy after rifampicin treatment in relation to the measured occupancy after rifampicin treatment of a strong promoter taken to be 100% occupied. FOP was estimated by calculating the percentage of any downstream fragment occupancy in relation to the full occupancy state of a strong promoter taken to be 100% occupied after rifampicin treatment. However, EI was estimated by calculating the percentage

Promoter	Sequence
BKS101	5' GAATTC**TAGACA**GCTGCATGCATC**TTTG**T**TATGGT**TATTTC 3'

Table11. **Starting promoter BKS101 sequence from the *Eco*R1 site to the -1 base.**

This table shows the sequence of our starting promoter BKS101 (5'➔3') from the *Eco*R1 site to base -1. The -35 element is shown in blue, the extended -10 element is shown in green (bases -16 and -17) and yellow (bases -14 and -15) and the -10 element is shown in red. Note that all promoter sequences are shown in table 6 (page 74).

of RNAP occupancy in the lac1357 region (the number is the distance between the transcription start site and the start of this region) in relation to the promoter occupancy before rifampicin treatment. Being able to calculate the occupancies of the promoter and two downstream regions (lac1357 and lac2720), shown in figure 8 (page 54), allowed the calculation of RNAP flux through each region of the gene, which is represented as PoPS. From now on, promoter occupancy before rifampicin treatment will be referred to as "measured promoter occupancy", and promoter occupancy after rifampicin treatment will be referred to as "promoter full occupancy".

4.2. Reporter gene assay

A reporter gene assay was used in this study to primarily assess the strengths of all promoters *in vitro* prior to the ChIP-qPCR method. β-galactosidase is the final product of the *lacZ* gene, which is activated by our set of promoters. Depending on each promoter's activity, the *lacZ* gene produces β-galactosidase. In other words, the stronger the promoter, the higher the β-galactosidase level observed. All promoters showed different levels of β-galactosidase, ranging from ~30 miller units for weak promoters to ~4000 miller units for strong promoters. An empty vector with no promoter inserted was used as a negative control. The different β-galactosidase activities including the negative control are shown in Figure 12.

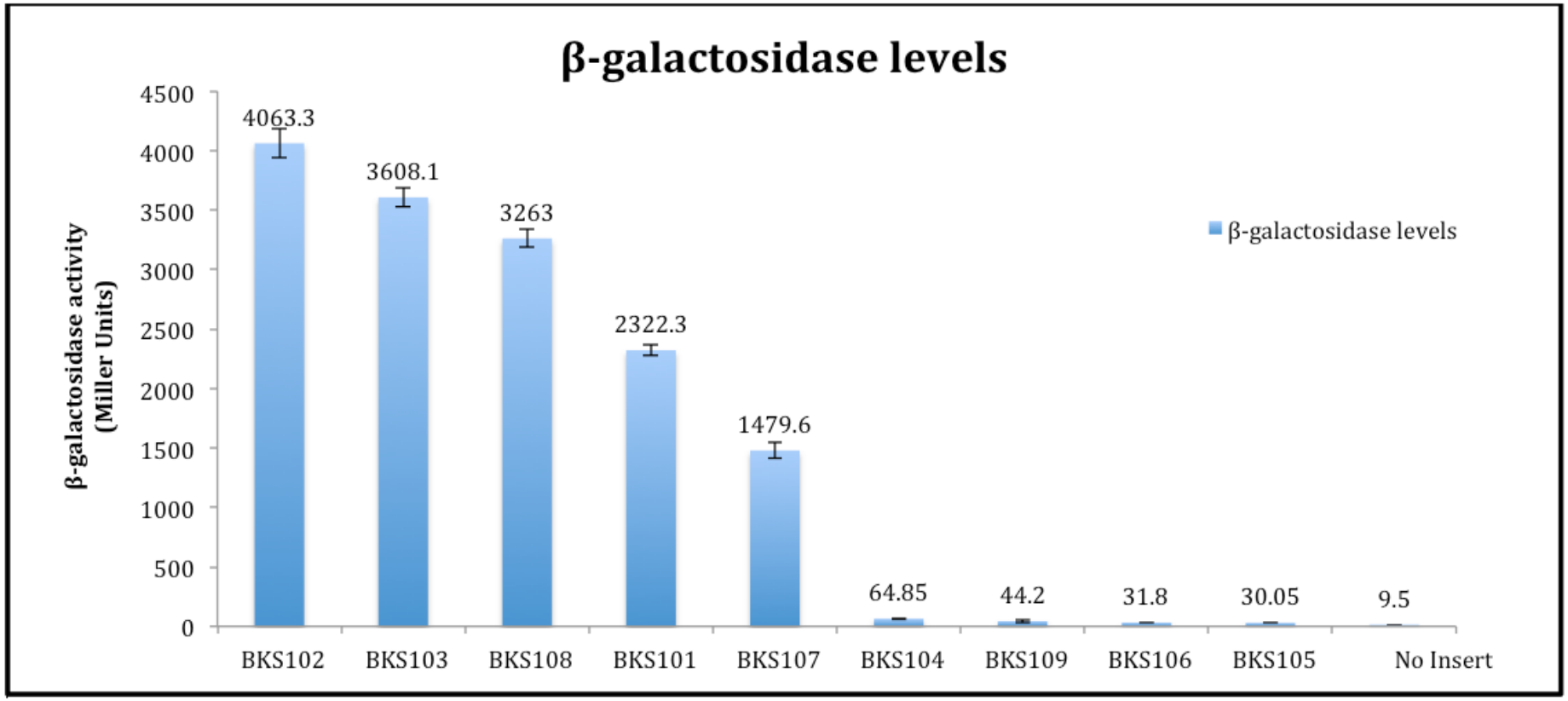

Figure 12. β-galactosidase measured activities for all 9 promoters used in this study.

This bar chart illustrates β-galactosidase activities of the *E. coli* M182 strain carrying the pRW50 plasmid with the 9 promoters used in this study, shown in the X axis, from highest to lowest activity in Miller units. The "No Insert" sample represents the *E. coli* M182 strain carrying the pRW50 plasmid with no promoter inserted as a negative control. Measurements were taken after growing the samples in 2x LB media at 37°C to an OD_{650} of 0.5. These values are the averages of at least three independent assays. The error bars on each bar show the standard deviation for each reading.

4.3. Effect of improving extended -10 element on promoter activity

To address the impact of improving the extended -10 element of the BKS101 promoter on its activity, another promoter carrying a p16G point mutation (promoter BKS102) changing the extended -10 element sequence from 5' TTTG 3' to the consensus sequence of 5' TGTG 3' was designed (Table 12). Unsurprisingly, this mutation has greatly induced overall promoter activity as measured by reporter gene assay. This activity was almost doubled from 2322.342 units in promoter BKS101 to 4063.3 units in promoter BKS102 (Figure 12).

Likewise, our new suggested direct method showed that improving the extended -10 element doubled the overall activity of the promoter (Figure 13). The promoter significantly improved in recruiting RNAP, showing a dramatic increase in POI level from 42.1% ((25.7/61.1) x 100) with promoter BKS101 to approximately 91.3% ((54.8/60) x 100) with the altered promoter BKS102. Both promoters were assumed to be 100% occupied after rifampicin treatment and the promoter competitivity was the same in both promoters showing a value of 100%. However, by comparing EI of both promoters, it is clear that this mutation has no major effect on the RNAP escape rate. The EI only increased by ~7% from 34.2% ((8.8/25.7) x 100) to 41.1% ((22.5/54.8) x 100). β-galactosidase levels, POI and EI values are shown in Table 13.

Promoter	Sequence
BKS101	5' GAATTC**TAGACA**GCTGCATGCATC**TTTG**T**TATGGT**TATTTC 3'
BKS102	5' GAATTC**TAGACA**GCTGCATGCATC**TG**<u>**G**</u>**TG**T**TATGGT**TATTTC 3'

Table12. **Sequences of promoters BKS101 and BKS102.**

This table shows the sequences of the BKS101 and BKS102 promoters (5'➔3') from the *Eco*R1 site to base -1. The -35 elements are shown in blue, the extended -10 elements are shown in green (bases -16 and -17) and yellow (bases -14 and -15) and the -10 elements are shown in red. The double-underlined base is the base mutated in the starting promoter to obtain the new promoter.

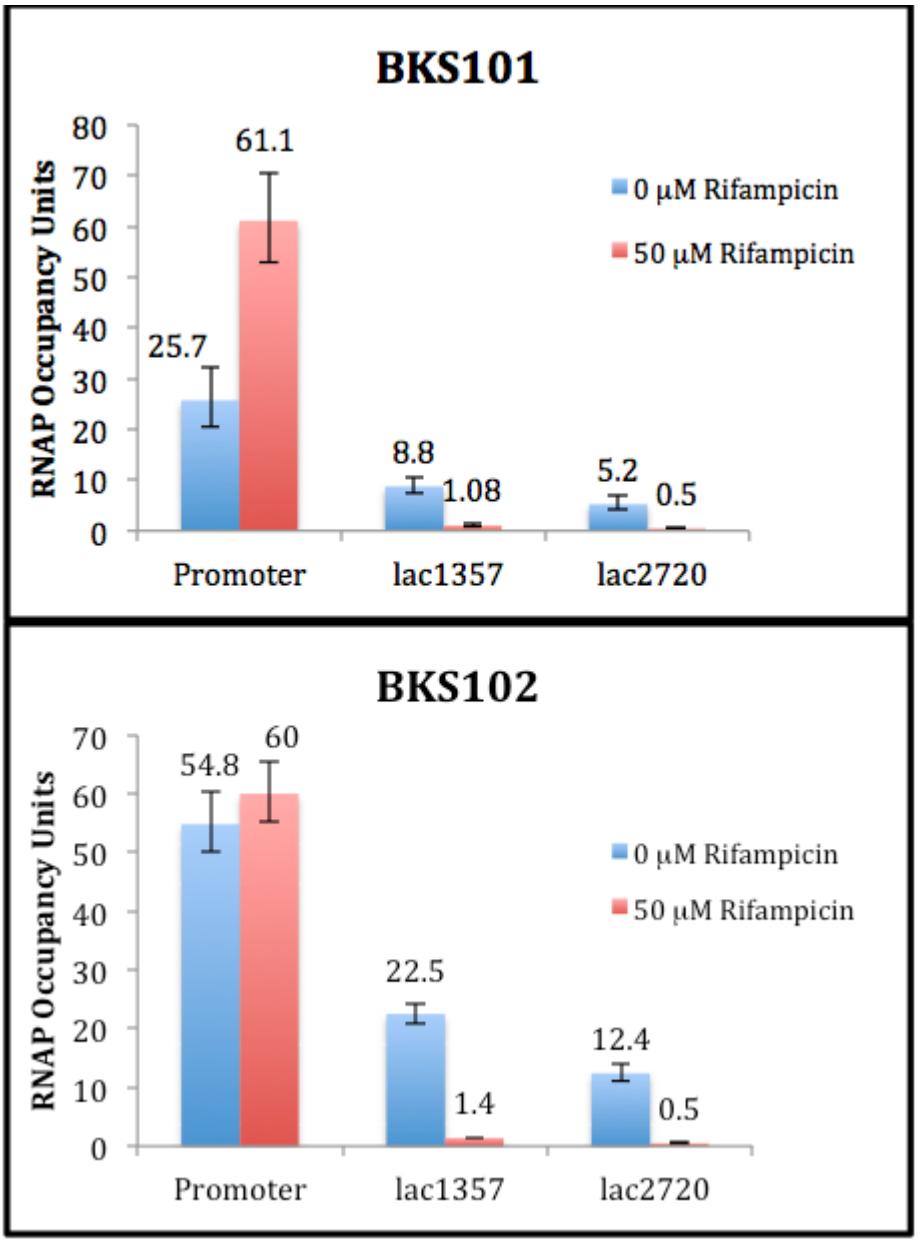

Figure 13. RNAP occupancies of promoters BKS101 and BKS102.

These two bar charts show RNAP occupancies of BKS101 (upper chart) and BKS102 (bottom chart) promoters and their lac1357 and lac2720 regions obtained from the ChIP-qPCR experiment. The blue bars show occupancies before rifampicin treatment and the red bars show occupancies after rifampicin treatment. All occupancies are the average of at least three biological repeats and three technical repeats. Error bars show the standard deviations of these calculated occupancies.

Promoter	β-gal level	Promoter competitivity (%)	POI	EI
BKS101	2322.342	100	42.1%	34.2%
BKS102	4063.302	100	91.3%	41.1%

Table13. **β-galactosidase levels, POI and EI values of promoters BKS101 and BKS102.**

This table summarizes the measured β-galactosidase levels and calculated promoter competitivity, POI and EI values of the BKS101 and BKS102 promoters.

PoPS was also calculated for the lac1357 and lac2720 regions following both promoters depending on the occupancy values of the promoter before and after rifampicin treatment and the measured occupancy values of both downstream regions. However, promoter occupancy after rifampicin treatment for both promoters was considered to be 100% to calculate the percentage of RNAP in downstream regions. The lac1357 and lac2720 regions are approximately 300 bp long. Therefore, considering the fact that RNAP travels at a rate of 50 bp/second, it takes RNAP 6 seconds to cross each region.

For promoter BKS101, the measured occupancy values for promoter, lac1357 and lac2720 were 25.7, 8.8 and 5.2, respectively. Therefore, the FOP of lac1357 region is 14.4% ((8.8/61.1) x 100) so, the region is 1/6.94 of the time occupied. Accordingly, the time interval between RNAPs (T_{int}) crossing this region is 41.6 seconds (6 x 6.94). Therefore, PoPS for this region is 0.024 (1/41.6). Similarly, the FOP of lac2720 region was 8.5% ((5.2/61.1) x 100) so the region is 1/11.76 of the time occupied. Thus, T_{int} is 70.6 seconds (6 x 11.76) and the PoPS value is 0.014 (1/70.6). All these values for promoter BKS101 are illustrated in Table 14.

The same calculations were applied to promoter BKS102 to investigate the FOP, T_{int} and PoPS values after improving the extended -10 element of promoter BKS101. The FOP for the lac1357 region of this promoter was 37.5 (1/2.67 of the time occupied). Thus, T_{int} is 16 (6 x 2.67) and PoPS is 0.06 (1/16). For the lac2720 region, FOP was 20.7 (1/4.83 of the time occupied). Therefore, T_{int} for this region is 29 (6 x 4.83) and PoPS was 0.034 (1/29). These values are shown in Table 15.

Region	Occupancy value (unit)	FOP (%)	T_{int} (seconds)	PoPS
Promoter BKS101	25.7	-	-	-
Lac1357	8.8	14.4	41.6	0.024
Lac2720	5.2	8.5	70.6	0.014

Table14. **PoPS for promoter BKS101.**

This table shows the occupancy value of each region and the FOP for lac1357 and lac2720 regions of the BKS101 promoter::*lac* fusion. It also shows the time interval between RNAPs crossing the lac1357 and lac2720 regions. All of these figures are necessary to calculate PoPS (also shown) for these two *lacZ* regions.

Region	Occupancy value (unit)	FOP (%)	T_{int} (seconds)	PoPS
Promoter BKS102	54.8	-	-	-
Lac1357	22.5	37.5	16	0.06
Lac2720	12.4	20.7	29	0.034

Table15. **PoPS for promoter BKS102.**

This table shows the occupancy value of each region and the FOP for lac1357 and lac2720 regions of the BKS102 promoter::*lac* fusion. It also shows the time interval between RNAPs crossing the lac1357 and lac2720 regions. All of these figures are necessary to calculate PoPS (also shown) for these two *lacZ* regions.

4.4. Effect of downgrading the extended -10 element on promoter activity

The effect of downgrading the extended -10 element of the promoter was tested by introducing a p14T to promoter BKS101 to get promoter BKS106. This mutation altered the extended -10 element sequence from 5' TTTG 3' to 5' TTTT 3' as shown in Table 16. The activity of this newly designed promoter was assessed using a reporter gene assay to measure the β-galactosidase levels of this promoter. Predictably, the promoter showed a great decrease in activity when compared with the original BKS101 promoter. The β-galactosidase level has significantly dropped from 2322.3 in promoter BKS101 to 31.8 in this new promoter (Figure 12).

This drop in activity was also confirmed by the ChIP-qPCR method. The measured promoter occupancy also dropped to 0.52 units from 25.7 units in the original promoter (Figure 14). Considering these occupancy values and promoters' full occupancy values, promoter competitivity, POI and EI were calculated for this promoter. Promoter competitivity sharply dropped from 100% units with promoter BKS101 to 3.24% with the mutated promoter. The POI value also dropped from 42.1% with the original promoter to 0.85% with the mutated promoter. The EI level showed that RNAP escaped the mutated promoter more easily than the original promoter, increasing from 34.2% to 65.4% as shown in Table 17.

Like previous promoters, PoPS was calculated for this promoter depending on the occupancy values obtained from the ChIP-qPCR experiment. These

Promoter	Sequence
BKS101	5’ GAATTC**TAGACA**GCTGCATGCATC**TTTG**T**TATGGT**TATTTC 3’
BKS106	5’ GAATTC**TAGACA**GCTGCATGCATC**TTTT**T**TATGGT**TATTTC 3’

Table16. **Sequences of promoters BKS101 and BKS106.**

This table shows the sequences of the BKS101 and BKS106 promoters (5’➔3’) from the *Eco*R1 site to base -1. The -35 elements are shown in blue, the extended -10 elements are shown in green (bases -16 and -17) and yellow (bases -14 and -15) and the -10 elements are shown in red. The double-underlined base is the base mutated in the starting promoter to obtain the new promoter.

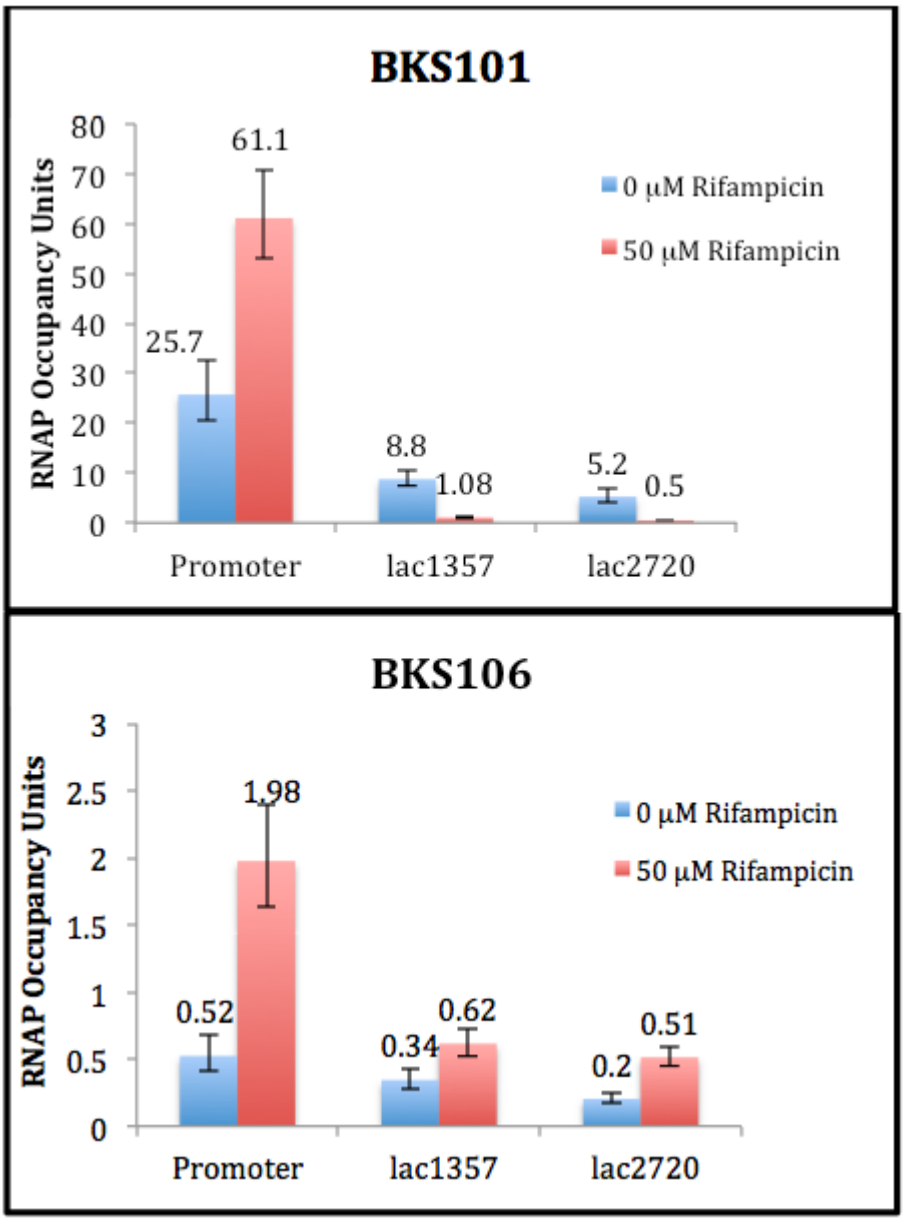

Figure 14. RNAP occupancies of promoters BKS101 and BKS106.

These two bar charts show RNAP occupancies of BKS101 (upper chart) and BKS106 (bottom chart) promoters and their lac1357 and lac2720 regions obtained from the ChIP-qPCR experiment. The blue bars show occupancies before rifampicin treatment and the red bars show occupancies after rifampicin treatment. All occupancies are the average of at least three biological repeats and three technical repeats. Error bars show the standard deviations of these calculated occupancies.

Promoter	β-gal level	Promoter competitivity (%)	POI	EI
BKS101	2322.342	100	42.1%	34.2%
BKS106	31.8	3.24	0.85%	65.4%

Table17. **β-galactosidase levels, POI and EI values of promoters BKS101 and BKS106.**

This table summarizes the measured β-galactosidase levels and calculated promoter competitivity, POI and EI values of the BKS101 and BKS106 promoters.

occupancy values were 0.52 units for the promoter region, 0.34 for the lac1357 region and 0.2 for the lac2720 region of the *lacZ* gene following the promoter. Therefore, the percentage of lac1357 occupancy to promoter BKS101 full occupancy state, after rifampicin treatment, was 0.56% (1/178.6 of the time occupied) and T_{int} was calculated by multiplying the time taken by RNAP to transverse the 300 bp region by 178.6 (6 x 178.6), resulting in 1071.6 seconds. Hence PoPS for this region was 0.0009 (1/1071.6). PoPS was calculated for thelac2720 region following the same method. The percentage of lac2720 occupancy to promoter BKS101 full occupancy state was 0.33 (1/303 of the time occupied). Accordingly, T_{int} for this region was 1818 seconds (6 x 303) and PoPS was 0.0005 (Table 18).

4.5. Effect of improving -35 element on promoter activity

The ChIP-qPCR experiment was repeated using a derivative of the promoter BKS101 having a consensus -35 element (promoter BKS104). The mutation targeted base -34 in particular converting the -35 element sequence from 5' TAGACA 3' to 5' TTGACA 3' and yielding a promoter with a perfect consensus -35 element, good -10 and extended -10 elements (Table 19). Against all expectations, this mutation almost completely disrupted the overall promoter activity, even though there was a consensus -35 element. The reporter gene assay showed that the overall activity decreased markedly from 2322.3 to 64.85 with this promoter (Figure 12). This unexplained decrease was confirmed via ChIP-qPCR method as well. Promoter competitivity was only 6.92% (Figure 15).

Region	Occupancy value (unit)	FOP (%)	T_{int} (seconds)	PoPS
Promoter BKS106	0.52	-	-	-
lac1357	0.34	0.56	1071.6	0.0009
lac2720	0.2	0.33	1818	0.0005

Table18. **PoPS for promoter BKS106.**

This table shows the occupancy value of each region and the FOP for lac1357 and lac2720 regions of the BKS106 promoter::*lac* fusion. It also shows the time interval between RNAPs crossing the lac1357 and lac2720 regions. All of these figures are necessary to calculate PoPS (also shown) for these two *lacZ* regions.

Promoter	Sequence
BKS101	5’ GAATTC**TAGACA**GCTGCATGCATC**TTTG**T**TATGGT**TATTTC 3’
BKS104	5’ GAATTC**TTGACA**GCTGCATGCATC**TTTG**T**TATGGT**TATTTC 3’

Table19. **Sequences of promoters BKS101 and BKS104.**

This table shows the sequences of the BKS101 and BKS104 promoters (5’➔3’) from the *Eco*R1 site to base -1. The -35 elements are shown in blue, the extended -10 elements are shown in green (bases -16 and -17) and yellow (bases -14 and -15) and the -10 elements are shown in red. The double-underlined base is the base mutated in the starting promoter to obtain the new promoter.

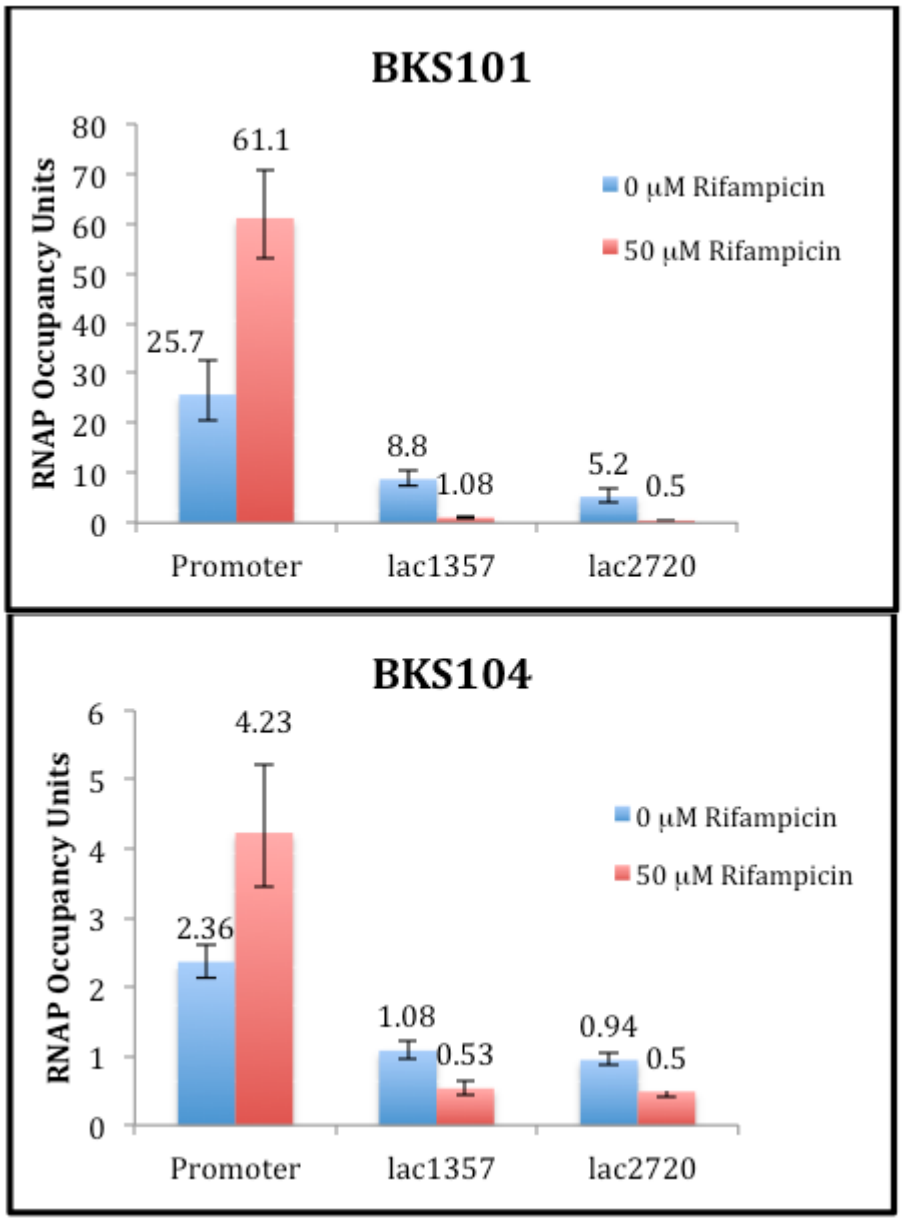

Figure 15. RNAP occupancies of promoters BKS101 and BKS104.

These two bar charts show RNAP occupancies of BKS101 (upper chart) and BKS104 (bottom chart) promoters and their lac1357 and lac2720 regions obtained from the ChIP-qPCR experiment. The blue bars show occupancies before rifampicin treatment and the red bars show occupancies after rifampicin treatment. All occupancies are the average of at least three biological repeats and three technical repeats. Error bars show the standard deviations of these calculated occupancies.

Against all expectations improving the -35 element negatively affected POI level, which decreased from 42.1% in promoter BKS101 to 3.86% in this promoter. Conversely, EI increased from 34.2% to 45.8% (Table 20).

The same method of PoPS calculation was applied to promoter BKS104. Measured occupancy values for the promoter, lac1357 and lac2720 were first calculated. These values were 2.36 for the promoter region, 1.08 for the lac1357 region and 0.94 for the lac2720 region. Therefore, the FOP of lac1357 region was 1.8% (1/55.55 of the time occupied). Accordingly, the T_{int} was 333.3 seconds (6 x 55.55). Therefore, PoPS for this region was 0.003 (1/333.3). The same calculations can be applied to the lac2720 region of this promoter. The FOP of lac2720 region was 1.54% (1/64.9 of the time occupied). Thus, T_{int} was 389.4 seconds and the PoPS value was 0.0026 (1/389.4). All these values for promoter BKS104 are illustrated in Table 21.

4.6. Effect of downgrading -10 element on promoter activity

To test the effect of downgrading the -10 element of promoter BKS101, a single mutation was introduced to base -12 of the promoter, changing it from T to C. This mutation resulted in a promoter (BKS105) with good -35 and extended -10 elements and a corrupted -10 element, which was changed from 5' TATGGT 3' to 5' CATGGT 3' (Table 22). As expected, corrupting the -10 element led to a

Promoter	β-gal level	Promoter competitivity (%)	POI	EI
BKS101	2322.342	100	42.1%	34.2%
BKS104	64.85	6.92	3.86%	45.8%

Table20. **β-galactosidase levels, POI and EI values of promoters BKS101 and BKS104.**

This table summarizes the measured β-galactosidase levels and calculated promoter competitivity, POI and EI values of the BKS101 and BKS104 promoters.

Region	Occupancy value (unit)	FOP (%)	T_{int} (seconds)	PoPS
Promoter BKS104	2.36	-	-	-
lac1357	1.08	1.8	333.3	0.003
lac2720	0.94	1.54	389.4	0.0026

Table21. **PoPS for promoter BKS104.**

This table shows the occupancy value of each region and the FOP for lac1357 and lac2720 regions of the BKS104 promoter::*lac* fusion. It also shows the time interval between RNAPs crossing the lac1357 and lac2720 regions. All of these figures are necessary to calculate PoPS (also shown) for these two *lacZ* regions.

Promoter	Sequence
BKS101	5’ GAATTC**TAGACA**GCTGCATGCATC**TTTG**T**TATGGT**TATTTC 3’
BKS105	5’ GAATTC**TAGACA**GCTGCATGCATC**TTTG**T**CATGGT**TATTTC 3’

Table22. **Sequences of promoters BKS101 and BKS105.**

This table shows the sequences of the BKS101 and BKS105 promoters (5’➔3’) from the *Eco*R1 site to base -1. The -35 elements are shown in blue, the extended -10 elements are shown in green (bases -16 and -17) and yellow (bases -14 and -15) and the -10 elements are shown in red. The double-underlined base is the base mutated in the starting promoter to obtain the new promoter.

dramatic drop in promoter strength. The β-galactosidase level measured with the reporter gene assay dropped from 2322.3 units to around 30 units with this altered promoter (Figure 12).

To have a clear view of how this change affected transcription, the ChIP-qPCR method was applied afterwards on this promoter (BKS105). ChIP-qPCR results agreed with reporter gene assay results showing a great decrease in overall promoter activity. Promoter competitivity dropped to 12.4% after altering the -10 element. Measured promoter occupancy also dropped from 25.7 units to 0.93 units (Figure 16). The POI value dropped from 42.1% with promoter BKS101 to 1.52% with this promoter. However, the EI value greatly increased from 34.2% to 86%, showing that this new promoter let RNAP escape easily (Table 23).

PoPS was also calculated for the lac1357 and lac2720 regions of the BKS105 promoter::*lac* fusion considering the occupancy values of these two regions and the promoter BKS101 full occupancy state. Measured promoter occupancy was 0.93 units, lac1357 region occupancy was 0.8 units and lac2720 region occupancy was 0.64 units. For the lac1357 region, the FOP was 1.3% (1/76.9 of the time occupied). Therefore, T_{int} for this region was 461.4 (6 x 76.9) seconds and PoPS was 0.0022 (1/461.4). The same was applied to the lac2720 region. The FOP was 1.05% (1/95.2 of the time occupied) and the T_{int} was calculated by multiplying the time taken by RNAP to cross the region (6 seconds) by 95.2, resulting in 571.2 seconds. PoPS was calculated by dividing 1 by the time interval between RNAPs arriving at the start of this region (1/571.2), resulting in 0.0017 polymerase per second. These values are illustrated in Table 24.

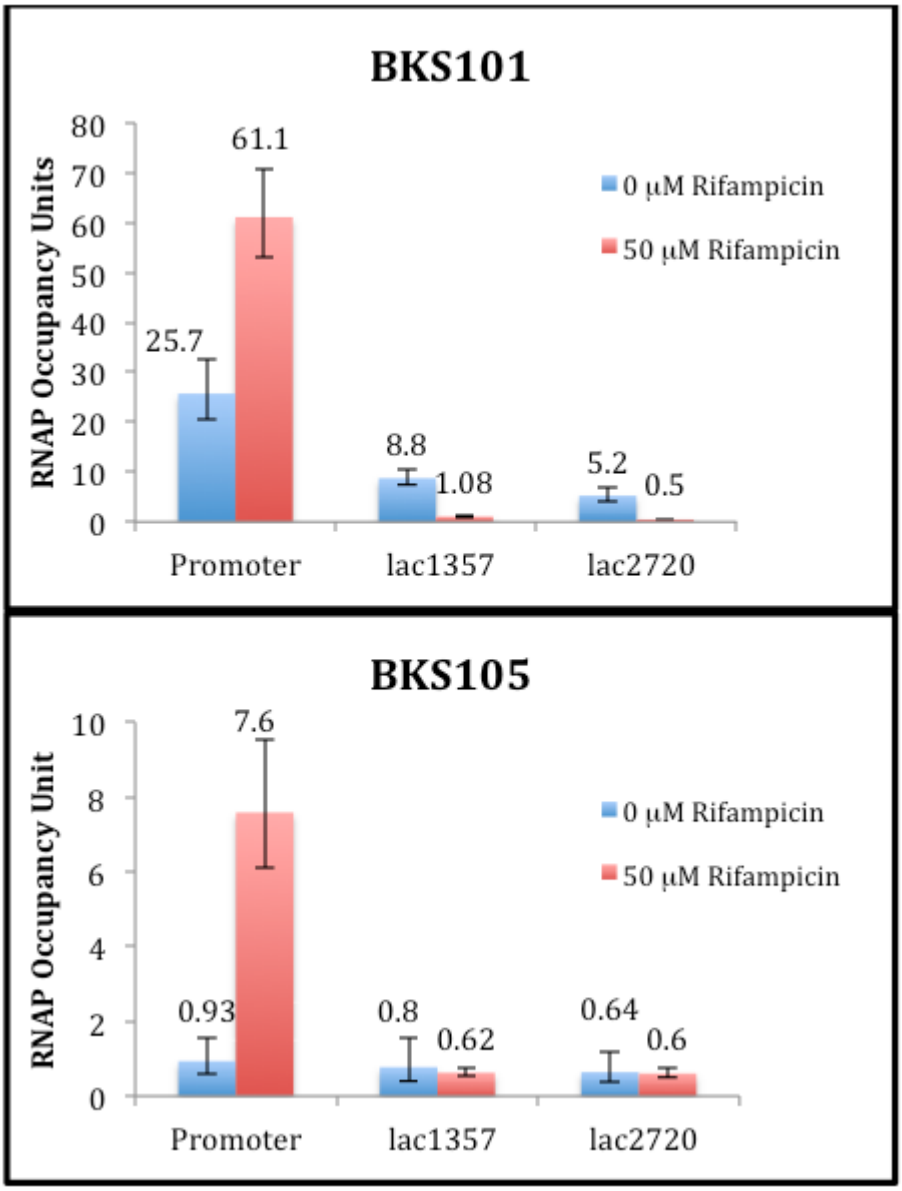

Figure 16. RNAP occupancies of promoters BKS101 and BKS105.

These two bar charts show RNAP occupancies of BKS101 (upper chart) and BKS105 (bottom chart) promoters and their lac1357 and lac2720 regions obtained from the ChIP-qPCR experiment. The blue bars show occupancies before rifampicin treatment and the red bars show occupancies after rifampicin treatment. All occupancies are the average of at least three biological repeats and three technical repeats. Error bars show the standard deviations of these calculated occupancies.

Promoter	β-gal level	Promoter competitivity (%)	POI	EI
BKS101	2322.342	100	42.1%	34.2%
BKS105	30.05	12.4	1.52%	86%

Table23. **β-galactosidase levels, POI and EI values of promoters BKS101 and BKS105.**

This table summarizes the measured β-galactosidase levels and calculated promoter competitivity, POI and EI values of the BKS101 and BKS105 promoters.

Region	Occupancy value (unit)	FOP (%)	T_{int} (seconds)	PoPS
Promoter BKS105	0.93	-	-	-
lac1357	0.8	1.3	461.4	0.0022
lac2720	0.64	1.05	571.2	0.0017

Table24. **PoPS for promoter BKS105.**

This table shows the occupancy value of each region and the FOP for lac1357 and lac2720 regions of the BKS105 promoter::*lac* fusion. It also shows the time interval taken by RNAP to cross the lac1357 and lac2720 regions. All of these figures are necessary to calculate PoPS (also shown) for these two *lacZ* regions.

4.7. Effect of improving the -35 element of a promoter having a corrupted extended -10 element

In this scenario, promoter BKS106 was used as a template to introduce a single mutation to the -35 promoter element in order to have a promoter containing a consensus -35, corrupted extended -10 and a good -10 element (BKS107). This mutation targeted the -34 base of the -35 element, changing it from A to T. Therefore, the -35 element sequence was changed from 5' TAGACA 3' to 5' TTGACA 3' as shown in Table 25.

The effect of this change was first addressed by assaying the *lacZ* gene final product (β-galactosidase). As expected, having a consensus -35 element revived promoter strength. The β-galactosidase level significantly increased from 31.8 units with promoter BKS106 to 1479.6 units with this promoter (Figure 12).

Promoter occupancy values, calculated from the ChIP-qPCR experiment before and after rifampicin treatment, agreed with the reporter gene assay showing that this mutation positively affected promoter strength. After rifampicin treatment, this new promoter showed a full occupancy of 66.3 units with 100% competitivity, whereas it was 1.98 units with 3.24% competitivity in promoter BKS106. Promoter occupancy under normal conditions also increased from 0.52 units with promoter BKS106 to 18 units with this promoter. All of these changes are shown in Figure 17. POI and EI values were calculated considering the occupancies of the promoter and a downstream region of the *lacZ* gene as before. The POI greatly increased from 0.85% to 27.2%. The EI value decreased with this promoter from 65.4% to 24.4% (Table 26).

Promoter	Sequence
BKS106	5' GAATTC**TAGACA**GCTGCATGCATC**TTTTT**T**TATGGT**TATTTC 3'
BKS107	5' GAATTC**TTGACA**GCTGCATGCATC**TTTTT**T**TATGGT**TATTTC 3'

Table25. **Sequences of promoters BKS106 and BKS107.**

This table shows the sequences of the BKS106 and BKS107 promoters (5'➔3') from the *Eco*R1 site to base -1. The -35 elements are shown in blue, the extended -10 elements are shown in green (bases -16 and -17) and yellow (bases -14 and -15) and the -10 elements are shown in red. The double-underlined base is the base mutated in promoter BKS106 to obtain promoter BKS107.

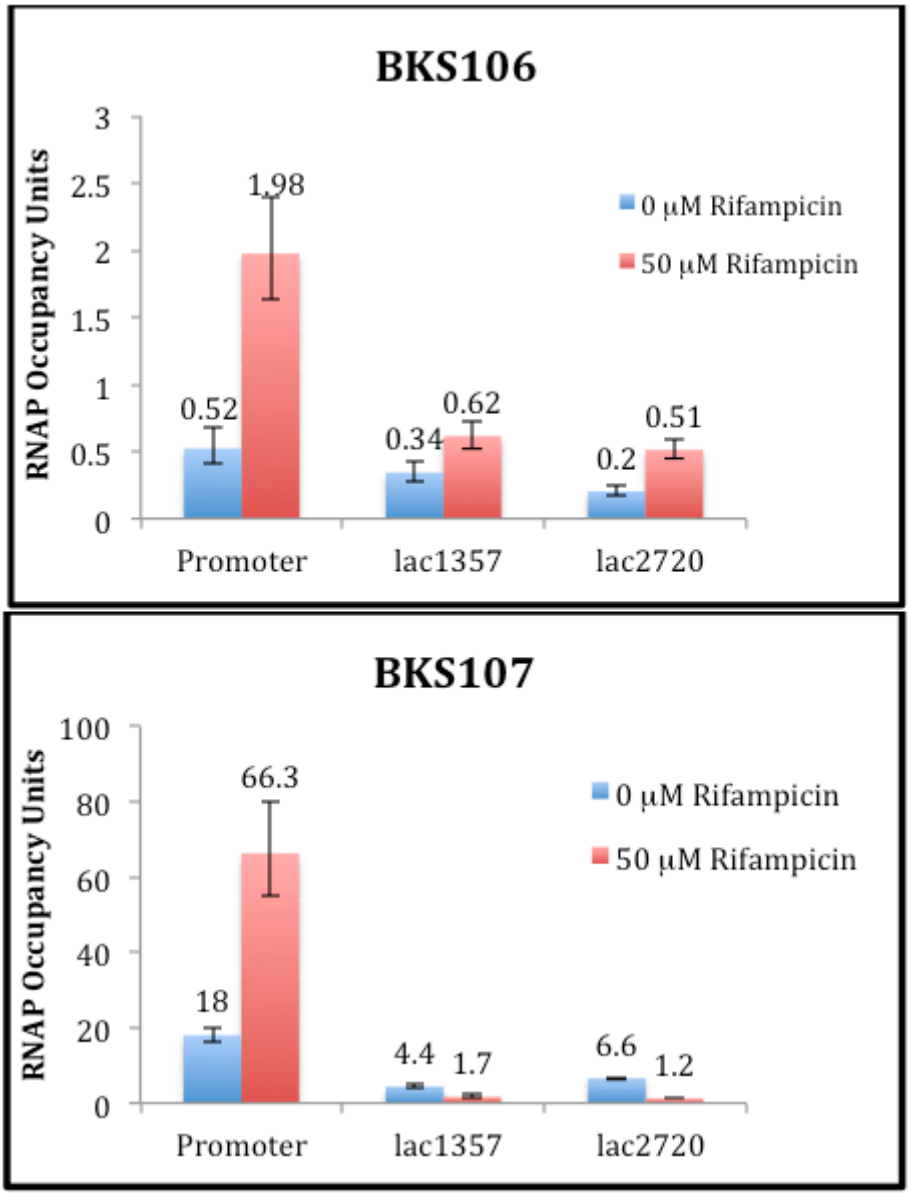

Figure 17. RNAP occupancies of promoters BKS106 and BKS107.

These two bar charts show RNAP occupancies of BKS106 (upper chart) and BKS107 (bottom chart) promoters and their lac1357 and lac2720 regions obtained from the ChIP-qPCR experiment. The blue bars show occupancies before rifampicin treatment and the red bars show occupancies after rifampicin treatment. All occupancies are the average of at least three biological repeats and three technical repeats. Error bars show the standard deviations of these calculated occupancies.

Promoter	β-gal level	Promoter competitivity (%)	POI	EI
BKS106	31.8	3.24	0.85%	65.4%
BKS107	1479.6	100	27.2%	24.4%

Table26. **β-galactosidase levels, POI and EI values of promoters BKS106 and BKS107.**

This table summarizes the measured β-galactosidase levels and calculated promoter competitivity, POI and EI values of the BKS106 and BKS107 promoters.

PoPS was calculated for the two downstream regions lac1357 and lac2720 depending on occupancies before and after rifampicin treatment as before. Occupancies of the promoter, lac1357 and lac2720 regions were 18, 4.4 and 6.6, respectively. Therefore, PoPS for the lac1357 region was calculated by first calculating the FOP, which yielded 6.64% (1/15 of the time occupied). Therefore, T_{int} for this region is 90 seconds (6 x 15). PoPS was calculated by dividing 1 by T_{int}, resulting in a value of 0.011.

For the lac2720 region, the FOP was 10% (1/10 of the time occupied). T_{int} for this region was 60 seconds (6 x 10). Finally, PoPS was calculated by dividing 1 by T_{int}, yielding a PoPS of 0.017. These calculations can be seen in Table 27.

4.8. Effect of improving -10 element on promoter activity

The effect of having a consensus -10 element in the starting promoter BKS101 on promoter activity was also tested. The sequence of this element was altered by introducing a double mutation to bases -7 and -8, thereby changing the element sequence from 5' TATGGT 3' to the consensus sequence 5' TATAAT 3'. Therefore, this altered promoter (BKS103) has a good -35, a good extended -10 and a consensus -10 hexamer as shown in Table 28. This change induced promoter activity as seen by reporter gene assay. The β-galactosidase level increased from 2322.3 in promoter BKS101 to 3608.1 with this promoter having a consensus -10 hexamer. The reporter gene assay results can be seen and compared in Figure 12.

Region	Occupancy value (unit)	FOP (%)	T_{int} (seconds)	PoPS
Promoter BKS107	18	-	-	-
lac1357	4.4	6.64	90	0.011
lac2720	6.6	10	60	0.017

Table27. **PoPS for promoter BKS107.**

This table shows the occupancy value of each region and the for lac1357 and lac2720 regions of the BKS107 promoter::*lac* fusion. It also shows the time interval between RNAPs crossing the lac1357 and lac2720 regions. All of these figures are necessary to calculate PoPS (also shown) for these two *lacZ* regions.

Promoter	Sequence
BKS101	5' GAATTC**TAGACA**GCTGCATGCATC**TTTG**T**TATGGT**TATTTC 3'
BKS103	5' GAATTC**TAGACA**GCTGCATGCATC**TTTG**T**TATAAT**TATTTC 3'

Table28. **Sequences of promoters BKS101 and BKS103.**

This table shows the sequences of the BKS101 and BKS103 promoters (5'➔3') from the *Eco*R1 site to base -1. The -35 elements are shown in blue, the extended -10 elements are shown in green (bases -16 and -17) and yellow (bases -14 and -15) and the -10 elements are shown in red. The double underlined bases are the bases mutated in the starting promoter to obtain the new promoter.

Promoter occupancy values obtained from the promoter region before and after rifampicin treatment using the ChIP-qPCR method also showed that the activity increased as a result of improving the -10 element of the promoter. Its full occupancy after rifampicin treatment increased from 61.1 units in promoter BKS101 to 83.9 units in this promoter giving 100% promoter competitivity. The measured promoter occupancy also increased from 25.7 units to 41.9 units as shown in Figure 18. POI slightly increased from 42.1% to 49.9% by having a consensus -10 element. However, the EI value decreased from 34.2% to 21%. These values are shown in Table 29.

PoPS values for lac1357 and lac2720 downstream of this new promoter were also calculated using measured occupancies obtained from the ChIP-qPCR experiment. These occupancy values were 41.9, 8.8 and 4.3 units for the promoter, lac1357 and lac2720, respectively. PoPS was calculated for each region separately. For the lac1357 region, FOP was 10.5% (1/9.52 of the time occupied). Therefore, T_{int} for this region was 57.12 seconds (6 x 9.52). This leads to a PoPS value of 0.017 (1/57.12). Likewise, for the lac2720 region, FOP was 5.1% (1/19.6 of the time occupied). Therefore, T_{int} for this region was 117.6 seconds and the PoPS was 0.008 as shown in Table 30.

4.9. Effect of having fully consensus promoter sequences on promoter activity

To get a promoter containing consensus sequences in the -35, extended -10 an

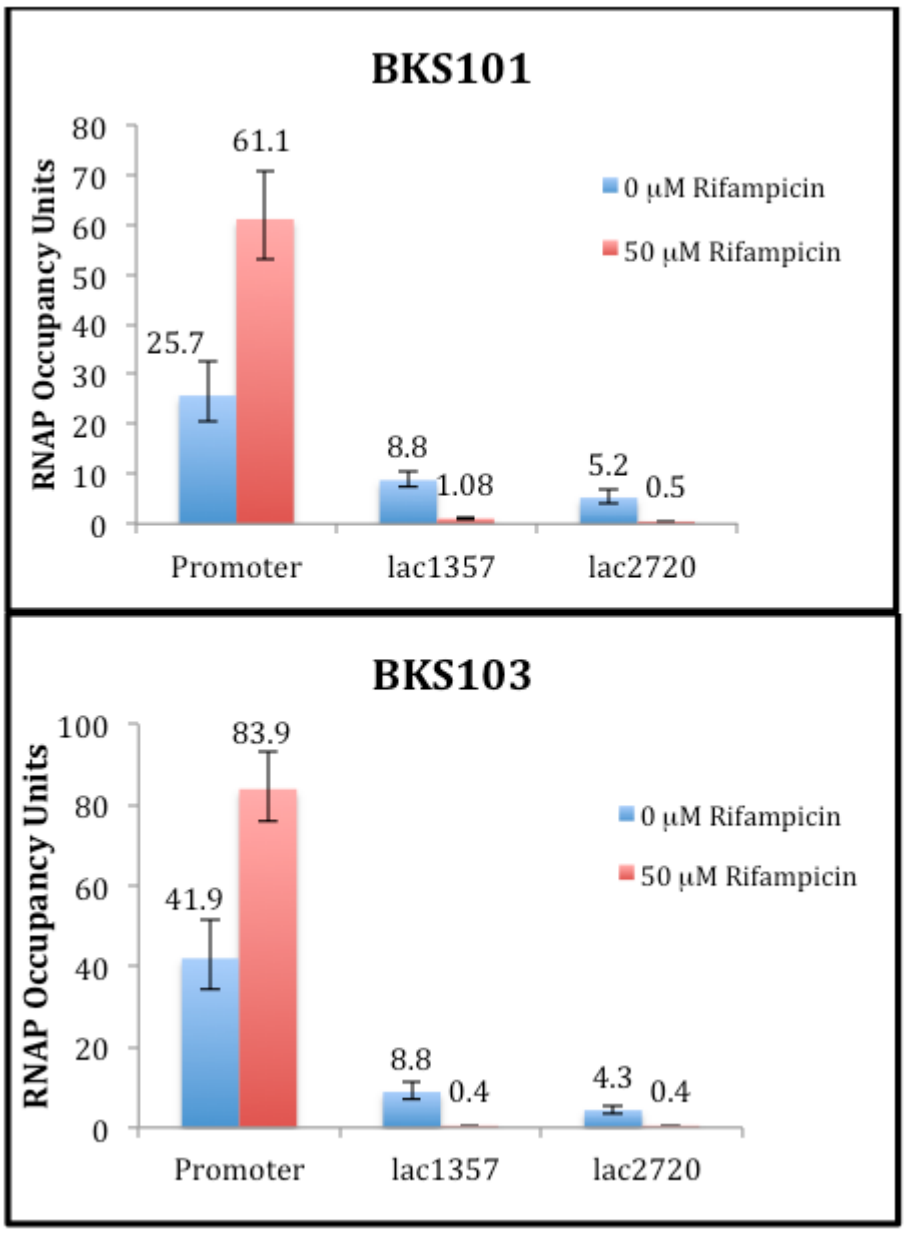

Figure 18. RNAP occupancies of promoters BKS101 and BKS103.

These two bar charts show RNAP occupancies of BKS101 (upper chart) and BKS103 (bottom chart) promoters and their lac1357 and lac2720 regions obtained from the ChIP-qPCR experiment. The blue bars show occupancies before rifampicin treatment and the red bars show occupancies after rifampicin treatment. All occupancies are the average of at least three biological repeats and three technical repeats. Error bars show the standard deviations of these calculated occupancies.

Promoter	β-gal level	Promoter competitivity (%)	POI	EI
BKS101	2322.342	100	42.1%	34.2%
BKS103	3608.1	100	49.9%	21%

Table29. **β-galactosidase levels, POI and EI values of promoters BKS101 and BKS103.**

This table summarizes the measured β-galactosidase levels and calculated promoter competitivity, POI and EI values of the BKS101 and BKS103 promoters.

Region	Occupancy value (unit)	FOP (%)	T_{int} (seconds)	PoPS
Promoter BKS103	41.9	100	-	-
lac1357	8.8	10.5	57.12	0.017
lac2720	4.3	5.1	117.6	0.008

Table30. **PoPS for promoter BKS103.**

This table shows the occupancy value of each region and the FOP for lac1357 and lac2720 regions of the BKS103 promoter::*lac* fusion. It also shows the time interval between RNAPs crossing the lac1357 and lac2720 regions. All of these figures are necessary to calculate PoPS (also shown) for these two *lacZ* regions.

-10 elements, two single mutations (p34T and p16G) were introduced to the previously designed promoter BKS103. These two mutations aimed to provide the promoter with a consensus -35 element by changing base -34 from A to T and a consensus extended -10 element by changing base -16 from T to G. Therefore, the new promoter (BKS108) had a consensus -35 element sequence of 5' TTGACA 3', a consensus extended -10 element sequence of 5' TGTG 3' and a consensus -10 element sequence of 5' TATAAT 3' as shown in Table 31.

When the β-galactosidase level was assessed using reporter gene assay, this promoter showed a lower overall activity than promoter BKS103, even though all three elements consisted of consensus sequences. Specifically, the β-galactosidase level decreased from 3608.1 units with promoter BKS103 to 3263 units with this promoter. These findings are shown in Figure 12.

However, this promoter showed higher promoter occupancy values than promoter BKS103 according to the ChIP-qPCR method. Promoter competitivity dramatically increased showing occupancy after rifampicin treatment of 294 units, while it was 83.9 units for promoter BKS103 (both assumed as 100% occupancy). Similarly, the occupancy before rifampicin treatment also increased from 41.9 units with promoter BKS103 to 83.1 units with this consensus promoter (Figure 19). Occupancy values of the promoter and the lac1357 region were used to calculate promoter competitivity, POI and EI values for this promoter (Table 32). This promoter, with its consensus sequences, showed lower POI (28.26%) and EI (10.1%) values than promoter BKS103's POI (49.9%) and EI (21%) values.

Promoter	Sequence
BKS103	GAATTC**TAGACA**GCTGCATGCATC**TTTG**T**TATAAT**TATTTC
BKS108	GAATTC**TTGACA**GCTGCATGCATC**TGTG**T**TATAAT**TATTTC

Table31. **Sequences of promoters BKS103 and BKS108.**

This table shows the sequences of the BKS103 and BKS108 promoters (5'➔3') from the *Eco*R1 site to base -1. The -35 elements are shown in blue, the extended -10 elements are shown in green (bases -16 and -17) and yellow (bases -14 and -15) and the -10 elements are shown in red. The double-underlined bases are the bases mutated in promoter BKS103 to obtain the promoter BKS108.

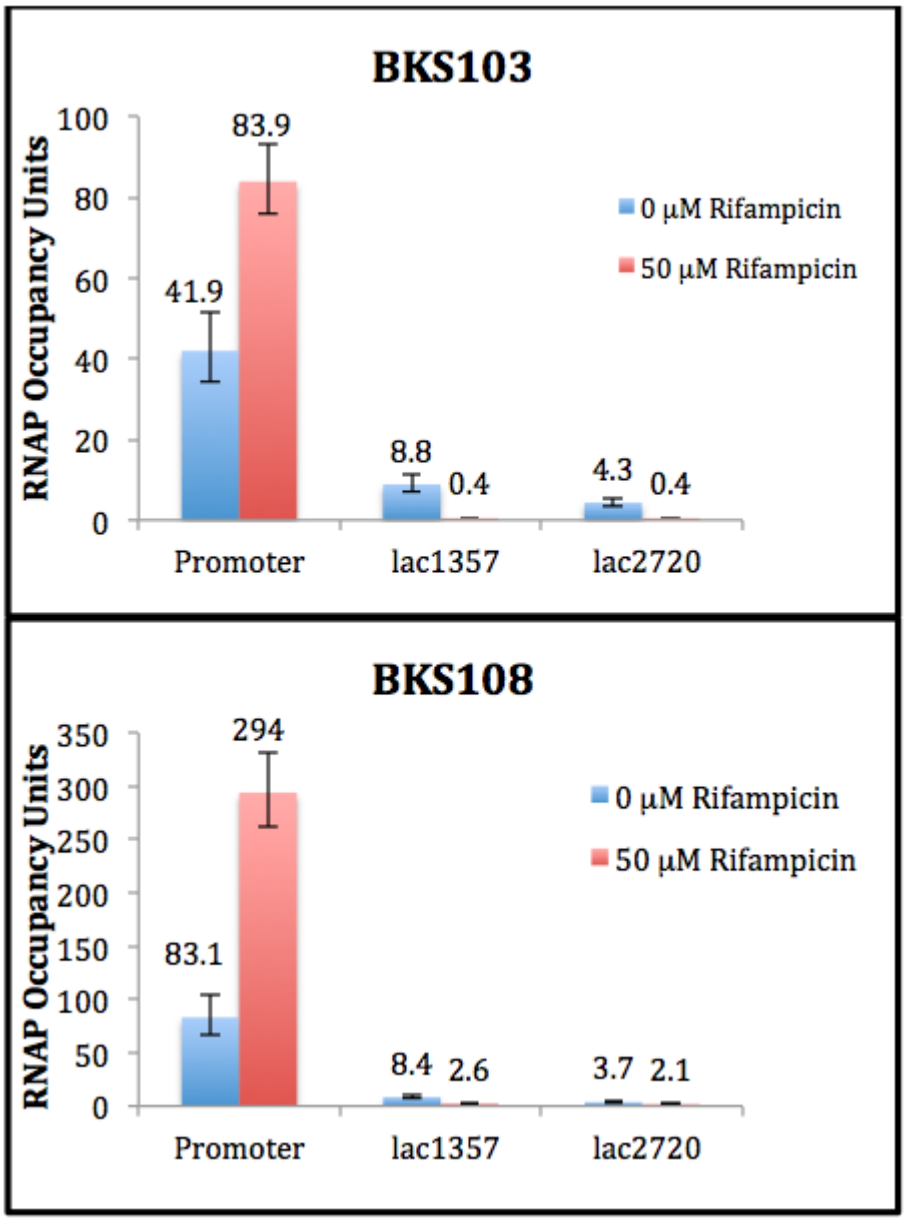

Figure 19. RNAP occupancies of promoters BKS103 and BKS108.

These two bar charts show RNAP occupancies of BKS103 (upper chart) and BKS108 (bottom chart) promoters and their lac1357 and lac2720 regions obtained from the ChIP-qPCR experiment. The blue bars show occupancies before rifampicin treatment and the red bars show occupancies after rifampicin treatment. All occupancies are the average of at least three biological repeats and three technical repeats. Error bars show the standard deviations of these calculated occupancies.

Promoter	β-gal level	Promoter competitivity (%)	POI	EI
BKS103	3608.1	100	49.9%	21%
BKS108	3263	100	28.26%	10.1%

Table32. **β-galactosidase levels, POI and EI values of promoters BKS103 and BKS108.**

This table summarizes the measured β-galactosidase levels and calculated promoter competitivity, POI and EI values of the BKS103 and BKS108 promoters.

Regarding PoPS calculations, the full occupancy state of the promoter and the measured occupancies of all regions were used. The FOP of lac1357 region was 2.86% (1/35 of the time occupied). Therefore, this region was 1/35 of the time occupied by RNAP. Accordingly, T_{int} was calculated by multiplying the time taken by RNAP to traverse this region (6 seconds) by 35, which yielded 210 seconds. Consequently, the PoPS value was 0.0048 (1/210). Similarly, the FOP for the lac2720 region was 1.26% (1/79.4 of the time occupied). Therefore, T_{int} for this region was 476.4 seconds (6 x 79.4) meaning that a new RNAP arrives at the start of this region every 476.4 seconds. Therefore, PoPS was calculated by dividing 1 by 476.4, which yielded a value of 0.002. These calculated values are illustrated in Table 33.

4.10.Effect of having a consensus -10 hexamer only on promoter activity

The effect of having a consensus -10 hexamer only in a promoter was also tested by the new method developed in this study. The promoter (BKS109) was designed by changing 4 bases (-30 to -33) of the -35 element and one base (-14) of the extended -10 element of promoter BKS103. The activity of the -35 element was greatly reduced by changing its sequence from 5' TAGACA 3' (5/6 fit to the consensus) to 5' TACTGT 3' (1/6 fit to the consensus). The sequence of the extended -10 element was changed from 5' TTTG 3' to 5' TTTT 3' as shown in Table 34. After designing the promoter, promoter activity was primarily

Region	Occupancy value (unit)	FOP (%)	T_{int} (seconds)	PoPS
Promoter BKS108	83.1	-	-	-
lac1357	8.4	2.86	210	0.0048
lac2720	3.7	1.26	476.4	0.002

Table33. **PoPS for promoter BKS108.**

This table shows the occupancy value of each region and the FOP for lac1357 and lac2720 regions of the BKS108 promoter::*lac* fusion. It also shows the time interval between RNAPs crossing the lac1357 and lac2720 regions. All of these figures are necessary to calculate PoPS (also shown) for these two *lacZ* regions.

Promoter	Sequence
BKS103	5’ GAATTC**TAGACA**GCTGCATGCATC**TTTG**T**TATAAT**TATTTC 3’
BKS109	5’ GAATTC**TA<u>CTGT</u>**GCTGCATGCATC**TTT<u>T</u>**T**TATAAT**TATTTC 3’

Table34. **Sequences of promoters BKS103 and BKS109.**

This table shows the sequences of the BKS103 and BKS109 promoters (5’➔3’) from the *Eco*R1 site to base -1. The -35 elements are shown in blue, the extended -10 elements are shown in green (bases -16 and -17) and yellow (bases -14 and -15) and the -10 elements are shown in red. The double-underlined bases are the bases mutated in promoter BKS103 to obtain BKS109 promoter.

assessed using the reporter gene assay to measure the β-galactosidase level with this promoter. The β-galactosidase level immensely declined from 3608.1 units with promoter BKS103 to 44.2 units with this promoter containing a consensus -10 element only, as shown in Figure 12.

After primary promoter activity assessment with the reporter gene assay, the ChIP-qPCR method was applied to gain a deeper perspective on how promoter activity had changed. The promoter showed lower RNAP recruitment activity by demonstrating lower occupancy values both before and after rifampicin treatment. Promoter occupancy after rifampicin treatment decreased from 83.9 units to 20.5 units showing promoter competitivity of 24.4%. Similarly, the measured promoter occupancy before rifampicin treatment decreased from 41.9 units to 3.2 units as shown in Figure 20. The POI of this promoter greatly decreased from 49.9% to 3.8% after altering the -35 and extended -10 promoter elements; however, the EI of this promoter slightly improved, increasing from 21% with promoter BKS103 to 21.9% with this promoter, as shown in Table 35.

PoPS values were also calculated for the lac1357 and lac2720 regions of the BKS109 promoter::*lac* fusion using occupancy values obtained from the ChIP-qPCR experiment. These occupancies were 3.2 units for the promoter region and 0.7 units for both the lac1357 and lac2720 regions. For both *lacZ* regions, the FOP was 0.83% (1/120.5 of the time occupied). Therefore, the T_{int} for both segments was 723 seconds (6 x 120.5). Finally, PoPS was calculated by dividing 1 by T_{int} resulting in a value of 0.0014 for both regions, as shown in Table 36.

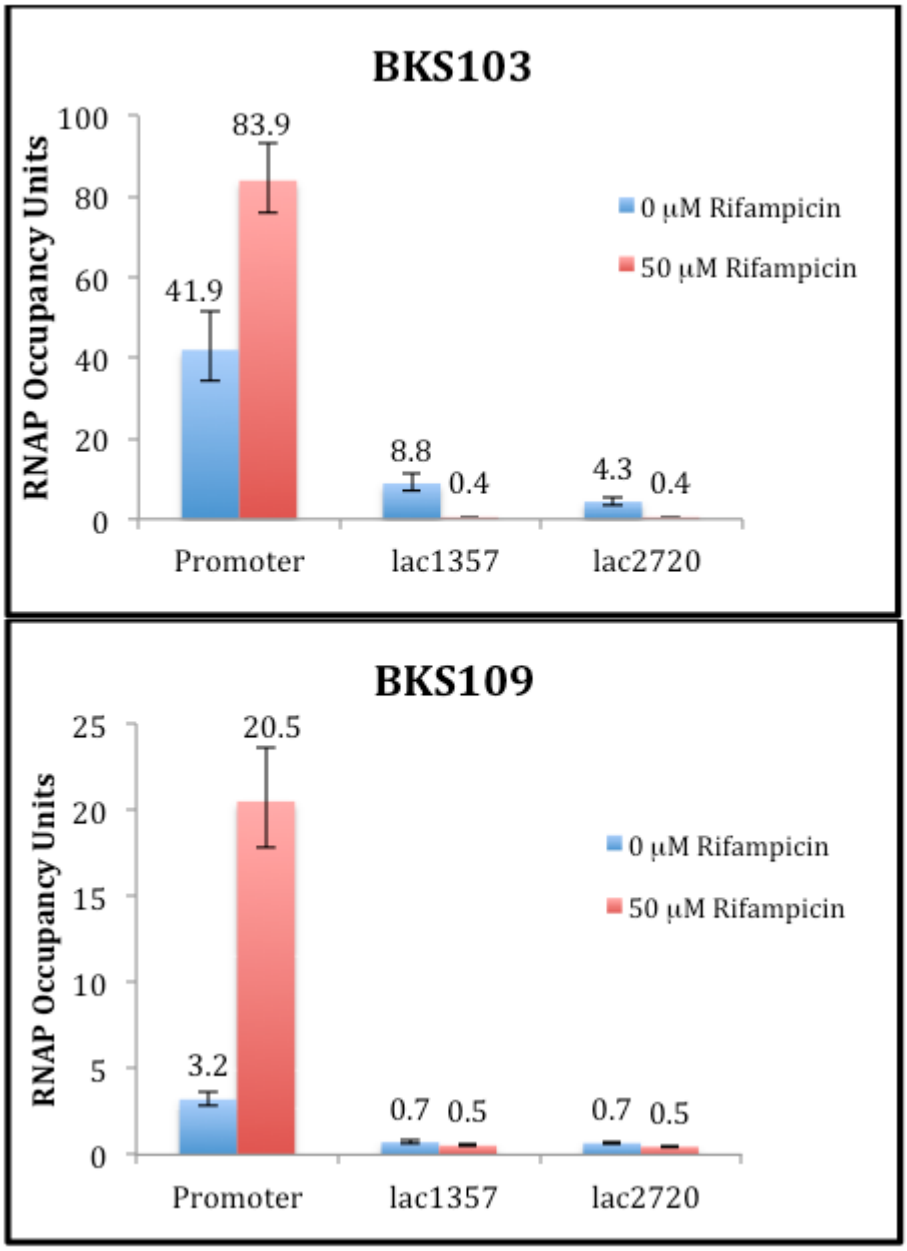

Figure 20. RNAP occupancies of promoters BKS103 and BKS109.

These two bar charts show RNAP occupancies of BKS103 (upper chart) and BKS109 (bottom chart) promoters and their lac1357 and lac2720 regions obtained from the ChIP-qPCR experiment. The blue bars show occupancies before rifampicin treatment and the red bars show occupancies after rifampicin treatment. All occupancies are the average of at least three biological repeats and three technical repeats. Error bars show the standard deviations of these calculated occupancies.

Promoter	β-gal level	Promoter competitivity (%)	POI	EI
BKS103	3608.1	100	49.9%	21%
BKS109	44.2	24.4	3.8%	21.9%

Table35. **β-galactosidase levels, POI and EI values of promoters BKS103 and BKS109.**

This table summarizes the measured β-galactosidase levels and calculated promoter competitivity, POI and EI values of the BKS103 and BKS109 promoters.

Region	Occupancy value (unit)	FOP (%)	T_{int} (seconds)	PoPS
Promoter BKS109	3.2	-	-	-
lac1357	0.7	0.83	723	0.0014
lac2720	0.7	0.83	723	0.0014

Table36. **PoPS for promoter BKS109.**

This table shows the occupancy values and the FOP lac1357 and lac2720 regions of the BKS109 promoter::*lac* fusion. It also shows the time interval between RNAPs crossing the lac1357 and lac2720 regions. All of these figures are necessary to calculate PoPS (also shown) for these two *lacZ* regions.

4.11.Discussion

Promoters are components of the DNA that participate in controlling transcription specificity by recruiting RNAP for the transcription of specific genes. This specificity depends solely on four specific promoter hexamer sequence elements and their sequences: the -10, extended -10, -35 and UP elements. In this study, a new method was developed to directly measure promoter strength and correlate it to different combinations of promoter elements. This method was optimized by testing promoters with different combinations of elements that was inserted into pRW50 plasmid carried by *E. coli*. Although there are several promoter elements known to effect promoter strength, we concentrated on testing the effect of three of these elements: the -10, extended -10 and -35 promoter elements. This effect was assessed by comparing the measured activities of two promoters at a time.

The starting promoter, BKS101, was obtained from the work of Burr and colleagues (2000). This promoter has a favorable combination of promoter elements and showed a β-galactosidase activity of 2322.3 units and a 100% promoter competitivity. The POI of this promoter was 42.1%, meaning that it recruits RNAPs to a fairly good extent. Looking at the EI of this promoter, it is clear that not all recruited RNAPs are escaping easily. In fact, only 34.2% of recruited RNAPs are escaping downstream to transcribe subsequent genes. These escaping RNAPs showed PoPS values of 0.024 and 0.014 for downstream regions lac1357 and lac2720 regions, respectively. In other words, this promoter is limited in terms of the escape level of RNAPs. These findings were aligned with

our expectations, as robust promoter-DNA interactions bind RNAPs tightly and do not let the RNAPs escape easily.

Improving the extended -10 region of promoter BKS101 by changing the extended -10 sequence form 5' TTTG 3' to 5' TGTG 3' notably boosted overall promoter activity. There was an approximate 57% increase in β-galactosidase levels, implying that the extended -10 element plays a vital role in transcription activity. However, the promoter competitivity was the same as in the original promoter BKS101 showing a value of 100%. Burr and colleagues (2000) tested the same mutation and recorded approximately the same increase in β-galactosidase levels. Yielding a value of approximately 91%, POI increased by more than two-folds as well, suggesting that the promoter's affinity to RNAPs has increased after improving the extended -10 element. Such a strong promoter will bind RNAPs very well and will not let them escape easily. The EI slightly increased by ~7% more than promoter BKS101 EI, meaning that this new promoter is also limited in terms of escape of RNAPs. However, considering the amount of recruited RNAPs, the escaping RNAPs showed higher PoPS values than PoPS values of the starting promoter BKS101 in both downstream regions. These values were 0.06 and 0.034 for lac1357 and lac2720 regions, respectively.

The effect of demoting the extended -10 element of promoter BKS101 was also tested. The sequence of this element was changed from 5' TTTG 3' to 5' TTTT 3', and this change lowered promoter activity as measured by reporter gene assay and promoter full occupancy with rifampicin. β-galactosidase levels decreased from 2322.3 units to approximately 31.8 units and the full occupancy state dropped from 61.1 units to approximately 2 units showing a 3.24 promoter

competitivity. The POI dropped to 0.85% only. However, the EI showed almost a two-fold increase from 34.2% to 65.4%. These results show that knocking down the extended -10 element severely impairs the promoter's ability to recruit and hold RNAPs. However, there was a very long time gap between RNAPs passing through downstream regions. For lac1357, T_{int} was around 18 minutes resulting in a PoPS value of 0.0009 and for lac2720, T_{int} was around 30 minutes giving a PoPS value of 0.0005. The same combination of promoter elements was tested by Burr and colleagues (2000) and showed the same diminishing promoter activity.

Conversely, giving promoter BKS101 a perfect -35 element did not increase promoter activity. Instead, the overall activity decreased dramatically, yielding a β-galactosidase level of only 64.85 units. The promoter competitivity greatly decreased to 6.92, as well. POI greatly decreased from 42.1% in promoter BKS101 to 3.86% in this promoter. However, EI value was slightly higher than promoter BKS101 showing a value of 45.8%, which means the promoter is still limited in terms of escape similar to promoter BKS101. The amount of RNAPs escaping the promoter showed PoPS values lower than promoter BKS101. The T_{int} for lac1357 and lac2720 regions were around 5.5 and 6.5 minutes and yielded PoPS values of 0.003 and 0.0026, respectively. These findings were in accordance with expectations and supported the findings of previous studies. A study carried out by Ellinger and colleagues (1994) confirmed that promoters containing elements with consensus sequences exhibited RNAP stalling in the +6 to +12 region followed by slow release from this state. Monsalve and colleagues (1997) studied transcription repression by phage Φ29 protein p4 and revealed that this repression occurs in the presence of the -35 element of the promoter;

furthermore, removing this element leads to transcription activation. This same finding was also presented in another study conducted by Hsu (2002). The results of these three studies support the notion that repression results from the stalling of RNAP at the promoter, which is caused by the strong promoter-DNA interactions of consensus promoter elements.

The effect of weakening the promoter -10 element was also tested in this study. A promoter containing the same sequence as BKS101 with a single mutation to the -10 element was used (BKS105). This promoter was designed by Chan and Busby (1989) as a control in their study. The researchers found that changing the -12 base of the -10 element from T to C lowered promoter activity by more than 50-fold. We assessed the same activity by reporter gene assay and found that β-galactosidase levels drastically decreased by around 77-fold to only approximately 30.05 units. Even more interestingly, promoter competitivity decreased to only 12.4%. POI dropped to around 1.52%, demonstrating that the -10 element plays a decisive role in recruiting RNAPs (i.e., the weaker the -10 element, the fewer RNAPs are recruited). On the contrary, EI dramatically increased to approximately 86% and almost all recruited RNAPs escaped easily. This implies that demoting the -10 element significantly facilitated open complex formation and, accordingly, RNAP escape, resulting in a promoter that is limited at the recruitment step. This increased level of promoter escape resulted in a T_{int} of around 7.7 minutes and a PoPS value of 0.0022 in lac1357 region. For lac2720 region, the T_{int} was around 9.5 minutes resulting in a PoPS of 0.0017.

To test the effect of downgrading the extended -10 element, the ChIP-qPCR method was applied to another promoter (BKS107), which was obtained from

Chan and busby (1989). This promoter differs from promoter BKS101 at bases -14 and -34 of the extended -10 and -35 elements, respectively. The sequence of the extended -10 element was 5' TTTT 3' instead of 5' TTTG 3' in promoter BKS101, while the -35 element contains a consensus sequence of 5' TTGACA 3'. Demoting the extended -10 element alone, in promoter BKS106 tested previously, decreased promoter activity. Nevertheless, combining this demotion with a perfect consensus -35 element revived promoter activity to some extent. Reporter gene assay for this new promoter showed a β-galactosidase level of 1479.6 units, which is higher than the promoter with a demoted extended -10 element only (BKS106). The full occupancy increased as well, to approximately 66 units showing a 100% competitive promoter. POI and EI were calculable and reached 27.2% and 24.4%, respectively, which are still lower than those of promoter BKS101 that has a good extended -10 element (5' TTTG 3'). However, by comparing POI and EI values of this promoter to the promoter BKS106 that has the same sequence with a mutated -35 element (shown above), the POI greatly increased from 0.85% in promoter BKS106 to 27.2% in this promoter. The EI value decreased with the new promoter, containing a consensus -35 element sequence, from 65.4% with promoter BKS106 to 24.4% with this promoter; this supports the idea that increasing promoter-DNA interactions with consensus sequences stalls RNAP at promoters and makes promoter escape slower and more difficult. This new promoter showed T_{int} values of around 1.5 and 1 minutes and PoPS values of 0.011 and 0.017 for lac1357 and lac2720, respectively. PoPS for lac2720 was slightly higher than in lac1357 of the same promoter, which might be as a result of RNAP bursting.

Two of the designed promoters, apart from the fully consensus promoter BKS108, contained a consensus -35 element, which are BKS104 and BKS107. In promoter BKS104, this consensus element was combined with a good extended -10 element (5' TTTG 3'). Whereas, in promoter BKS107, this consensus -35 element was combined with a weak extended -10 element (5' TTTT 3'). Surprisingly, promoter activity was extremely less in promoter BKS104 than promoter BKS107. Furthermore, the consensus -35 element, when combined with a good extended -10 element in promoter BKS104, showed more T_{int} and less PoPS levels than when it was combined with a weak extended -10 element in promoter BKS107.

Unlike BKS105, which has the same elements as promoter BKS101 with the -10 element demoted, the BKS103 promoter has been designed to have the same sequence with the -10 element improved to a consensus sequence (5' TATAAT 3'). This mutation positively affected β-galactosidase level, full occupancy state and POI for this new promoter. This promoter showed a 100% competitivity, a β-galactosidase level of 3608.1, a POI of 49.9% and an EI of 21%. Improving the -10 element of the promoter enhanced its ability to recruit RNAP but lowered RNAP escape rate. This decreased escape rate resulted in more T_{int} levels and less PoPS levels in downstream regions. The fact that improving the -10 element lowered the RNAP escape rate confirms that improving promoter sequence with consensus sequences increases the affinity of the promoter to RNAP and reduces the opportunity for RNAP escape.

As demonstrated previously, the perfect promoter with consensus sequences in all elements has lower β-galactosidase level, POI and EI than promoters BKS102

and BKS103. Essentially, this promoter showed the lowest EI among all tested promoters. The lower β-galactosidase level and EI are understandable, because the promoter has a very high affinity to RNAP and binds it very tightly; this negatively affects the escape rate and, consequently, the transcription rate. The fact that this consensus promoter recorded the highest occupancy after rifampicin treatment among all tested promoters is noteworthy. However, even though promoter occupancy under normal conditions was higher than all tested promoters, the POI value was lower than promoters BKS101, BKS102 and BKS103. The same consensus promoter was tested in a previous study conducted by Vo and colleagues (2003). They confirmed that consensus promoters recruit RNAPs but bind them more tightly than usual, leading to higher abortive initiation.

Improving the -10 element only, without good -35 and extended -10 elements in promoter BKS109, did not increase the promoter activity. The β-galactosidase level was very low according to reporter gene assay. Full promoter occupancy with rifampicin declined drastically, showing promoter competitivity of 24.4%, and indicated that this promoter was not recruiting RNAPs effectively. The low POI of 3.8% further indicates that the promoter is not recruiting RNAPs efficiently in normal conditions. The low EI of 21.9% shows that recruited RNAPs, although they are very few in number, are not escaping easily. Therefore, the -10 element cannot initiate successful transcription without the help of the -35 and extended -10 elements.

All figures obtained for different promoters and their downstream elements are shown in Table 37. For the promoter region, these figures include

β-galactosidase activity, promoter competitivity, POI and EI. While, for the downstream regions lac1357 and lac2720, these figures include FOP, Tint and PoPS

The new method developed in this study provides sound and direct insight on how promoter elements affect transcription activity. Generally, the closer the element to the consensus sequence, the more active the promoter. However, providing a promoter with consensus -35, extended -10 and -10 elements will increase the promoter's affinity to RNAPs and lower the chance of successful transcription. The promoter will hold RNAPs tightly, which makes RNAP escape very difficult. Moreover, initiating successful transcription is a collaborative task, which means that no matter how strong is the element; it cannot begin transcription without the help of other elements

Promoter	β-gal	Promoter competitivity (%)	POI (%)	EI (%)	FOP (%) (lac1357)	T_{int} (sec) (lac1357)	PoPS (lac1357)	FOP (%) (lac2720)	T_{int} (sec) (lac2720)	PoPS (lac2720)
BKS101	2322.3	100	42.1	34.2	14.4	41.6	0.024	8.5	70.6	0.014
BKS102	4063.3	100	91.3	41.1	37.5	16	0.06	20.7	29	0.034
BKS103	3608.1	100	49.9	21	10.5	57.12	0.017	5.1	117.6	0.008
BKS104	64.85	6.92	3.86	45.8	1.8	333.3	0.003	1.54	389.4	0.0026
BKS105	30.05	12.4	1.52	86	1.3	461.4	0.0022	1.05	571.2	0.0017
BKS106	31.8	3.24	0.85	65.4	0.56	1071.6	0.0009	0.33	1818	0.0005
BKS107	1479.6	100	27.2	24.4	6.64	90	0.011	9.95	60	0.017
BKS108	3263	100	28.26	10.1	2.86	210	0.0048	1.26	476.4	0.0021
BKS109	44.2	24.4	3.8	21.9	0.83	723	0.0014	0.83	723	0.0014

Table37. **Summary of calculated parameters for all promoters used in this study.**

This table summarizes all calculated parameters for all semisynthetic promoters used in this study. It shows β-galactosidase levels, POI and EI for each promoter. It also illustrates FOP, T_{int} and PoPS values for the two tested downstream regions the of *lacZ* gene following each promoter.

Chapter 5

Final Discussion

It has been well established that the gene expression pattern of any bacterium depends on different environmental conditions that consequently affect the number of RNAPs transcribing different bacterial genes. While these differences have been observed with different laboratory techniques, such as reporter gene assays, most of these techniques rely on estimating the gene product, which indirectly measures gene activity. In this study, a new simple and direct method, consisting of ChIP followed by qPCR, was developed to offer a direct insight into RNAP behavior *in vivo* at any selected DNA fragment. This technique works simply by crosslinking RNAP to DNA, which is then sheared by sonication and pulled down using specific beads and antibodies against RNAP. These immunoprecipitated DNA fragments are then quantified by qPCR using specifically designed probes to quantify only DNA fragments of interest. Exploiting the ability of rifampicin to block RNAPs at promoters allowed the calculation of promoter occupancy index (POI) and promoter competitivity. Quantifying immunoprecipitated DNA fragments of the same lengths allowed the calculation of EI (Escape Index) and RNAP flux through downstream transcription units. Two aspects of RNAP flux, through transcription units, were calculated: time interval between RNAPs coming to a specific single point of the transcribed fragment (T_{int}) and, subsequently, how many RNAPs pass through this point every second (Polymerase per Second; PoPS). Knowing the T_{int} allowed for the calculation of PoPS for each fragment of interest. These calculations were mainly dependent on the lengths of DNA fragments and how fast RNAP traverses these fragments. In present study, some of the calculated parameters, which are percentage of measured fragment occupancy to measured promoter occupancy; and T_{int}, were used to calculate PoPS for downstream fragments. However, there are some other derived parameters, which are promoter competitivity, POI, EI and FOP, which were

used to compare promoter strengths after manipulating the sequences of the promoter different elements to address the effect of introducing particular mutations on general promoter strength.

In a previous study, PoPS was calculated indirectly for different promoters under specific conditions (Kelly *et al.*, 2009). At the time of that study and to the researcher's best knowledge, there were no direct *in vivo* methods to measure PoPS. However, the researchers measured PoPS, indirectly for their promoters by fusing them to GFP coding genes. The amount of synthesized GFP can reflect promoter activity and be used to calculate PoPS using a specific quantitative model. Interestingly, Kelly and colleagues found that transcribing RNAPs are separated and in some of their promoters they reported PoPS values of 0.03 and 0.015, which are very similar to our findings using the developed direct method on the *lac* operon. Recently, Nielsen and colleagues (2016) designed genetic circuits with known functions using computational approaches. These circuits were then inserted into different genetic positions and transformed into specific bacterial strains. The researchers used one of the promoters as a standard with an output value of 1 relative promoter unit (RPU), which was then measured and found to be approximately 0.02 PoPS per promoter. This figure is in the same window of PoPS values measured in the present study using my direct method.

Although my developed technique is direct and provides a good *in vivo* perspective on the transcription process, it also has disadvantages. One of the major drawbacks of the method is that it is a blunt tool and might not be terribly accurate, especially when promoter activity goes really low and close to the background signal. In addition to this, it is an "ensemble" measurement, and ignores stochasticity between cells (i.e., it does not analyze single RNAP molecules). Another disadvantage of this method derives from the

significant background signals, which make the technique more plausible with strong promoters than weak ones; weak promoters cannot compete with strong ones in recruiting RNAPs even after rifampicin treatment. The final drawback is the result of the pause sequences, which are not detectable and highly affect the parameters measured with this technique, such as T_{int} and PoPS. The high background levels and the pause sequences issues can be precluded by a technique, that was developed by Churchman and Weissman (2011) whilst I have been doing my work, known as the nascent elongating template sequencing (NET-seq). This method was also used by Larson and his colleagues to identify a consensus pause sequence responsible for over 20,000 regulatory pause sites in *E. coli* (Larson *et al.*, 2014).

The calculated POI, EI, T_{int} and PoPS of transcription units having promoters with different combinations of elements revealed that changing promoter elements have a non-negligible effect on promoter ability to recruit RNAPs and on RNAP escape efficiency. While some promoter elements are more involved in RNAP recruitment, the same elements can affect RNAP escape rate and vice versa. Promoter strength can consequently affect the transcription of downstream genes by decreasing or increasing the RNAP supply to these genes.

Calculating the same parameters of the *lac* operon on the chromosome, surprisingly, showed that transcribing RNAPs are well separated along the *lac* operon by at least 9.6 seconds in the lac3 DNA fragment. Accordingly, recalling that RNAP crosses ~50 bases each second, RNAPs in this region will be separated by ~480 bases. In the lac5 fragment, RNAPs were separated by ~51 seconds, which indicates that they are separated by ~2550 bases. This unexpected finding has raised the question of why RNAPs are well separated during transcription of the *lac* operon. One possible reason is that, during

transcription elongation, RNAP causes local topological changes to the DNA by introducing negative supercoils behind and positive supercoils ahead (Liu and Wang, 1987); this may lead to limited access to the gene being transcribed for the following RNAPs, resulting in the noted gap between RNAPs. Transcriptional pausing is another possible cause of this separation between RNAPs. As has been noted in a previous study, RNAP pausing because of backtracking results in an extended elongation time (Voliotis *et al.*, 2008). This represents a logical concept, as RNAP involved in backtracking will be delayed for some time and the distance to the RNAP ahead will increase. Interestingly, regardless of the differences in PoPS values, PoPS values calculated in linear DNA and on the plasmid are so close to each other showing that transcribing RNAPs are well separated on both types of DNA.

The direct method developed in the present study is of a beneficial use in synthetic biology, as it can be used to establish a standardized hierarchy of core promoters based on the number of polymerase molecules passing through the gene per second (i.e., PoPS). This hierarchy can be built by correlating the ability of RNAP to transcribe a gene with the strength of core promoter, which is achievable by counting the number of RNAP enzymes traveling through a gene that are inhibited at different strengths of core promoters. This information can be utilized in synthetic biology systems where precise amounts of gene expression can be tailored to fit specific applications.

This developed direct method can contribute to the improvement of engineering biological parts and consequently the improvement of synthetic biology. This contribution could potentially result from obtaining previously estimated final products of constructed genes and the ability to understand RNAP distribution through a given gene. The idea of building biological parts was proposed by Knight and colleagues

(2003) in an attempt to improve engineering mechanisms for assembling constructed biological components into cellular systems. These biological parts are used as "biobricks" to build larger systems with known functions to be incorporated into organisms, such as bacteria.

The present study provides substantial information about the RNAP behavior at certain DNA sequences by measuring RNAP recruitment and flux directly instead of measuring final gene products. However, to gain a holistic understanding of the transcription system, many other methods can be applied to enhance knowledge of different aspects of this work and to confirm its findings. These methods include the following: electron tomography (ET); cell modelling, such as virtual electronic cell (e-cell); high-resolution imaging, such as photo-activated localisation microscopy (PALM) and high-throughput sequencing.

ET can be used to confirm my findings by creating a three-dimensional structure of the cellular contents. If possible, ET will allow the visualisation of bacterial DNA and all DNA-attached proteins including RNAPs, which will be beneficial in the study of all the parameters calculated in the present research for any given gene and RNAP distribution along the DNA. In an ideal experiment, it may be possible to follow populations in time as well. A study involving virtual cell modeling using computations, mathematical modeling and simulations of transcription would produce a model of RNAP recruitment, open promoter complex formation, RNAP escape and PoPS. Such a study would be of use for validating the findings of the present study by clarifying the differential access of RNAP to certain genes in relation to its location on the chromosome. Using tomographic techniques like photo activated-localisation microscopy (PALM), cryo-electron microscopy and single-molecule tracking by RNAP tagging can be used to visualize

RNAPs. This approach will validate the results of the present study and overcome the study's drawbacks, such as an inability to visualize RNAP bursting, DNA scrunching and stalled RNAPs. Applying the direct method developed in the present study on a large population of different promoters and using different host cells can verify RNAP recruitment, escape and flux in virtual cell models as well. High-throughput sequencing could also assist in confirming my findings by sequencing RNA chains produced by the constructed biobricks. This approach would provide beneficial, high-resolution information about pause sequences that essentially affect RNAP flux through transcription units (Larson *et al.,* 2014). These aforementioned research projects were beyond the scope of present study, but may be important in further elucidating the processes that were researched.

The wider conclusion of this study is that the density of RNAP, on both the plasmid and the chromosome, is a lot less than we think. What is more interesting is that PoPS values on the plasmid and on the chromosome are roughly the same. The direct method proposed in this study has a potential to directly measure RNAP density on any gene other than the *lac* operon. It also has a potential in situations other than studying RNAP density on the *lac* operon performed in present study. This method is also has a potential for use in any coding region.

Bibliography

Adelman, K. and Lis, J.T., 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews Genetics*, 13(10), pp.720-731.

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

Altieri, A.S., Mazzulla, M.J., Horita, D.A., Coats, R.H., Wingfield, P.T., Das, A., Court, D.L. and Byrd, R.A., 2000. The structure of the transcriptional antiterminator NusB from *Escherichia coli*. *Nature Structural & Molecular Biology,* 7(6), pp.470-474.

Andrews, K.J. and Lin, E.C., 1976. Thiogalactoside transacetylase of the lactose operon as an enzyme for detoxification. *Journal of bacteriology,* 128(1), p.510.

Archambault, J. and Friesen, J.D., 1993. Genetics of eukaryotic RNA polymerases I, II, and III. *Microbiological reviews*, 57(3), pp.703-724.

Artsimovitch, I. and Henkin, T.M., 2009. *In vitro* approaches to analysis of transcription termination. *Methods*, 47(1), pp.37-43.

Artsimovitch, I., Vassylyeva, M.N., Svetlov, D., Svetlov, V., Perederina, A., Igarashi, N., Matsugaki, N., Wakatsuki, S., Tahirov, T.H. and Vassylyev, D.G., 2005. Allosteric modulation of the RNA polymerase catalytic reaction is an essential component of transcription control by rifamycins. *Cell*, *122*(3), pp.351-363.

Avery, O.T., MacLeod, C.M. and McCarty, M., 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *The Journal of experimental medicine,* 79(2), pp.137-158.

Bauer, D.L., Duchi, D. and Kapanidis, A.N., 2016. *E. coli* RNA Polymerase Pauses during Initial Transcription. *Biophysical Journal,* 110, p.21a.

Bell, C.E. and Lewis, M., 2001. The Lac repressor: a second generation of structural and functional studies. *Current opinion in structural biology*, 11(1), pp.19-25.

Belogurov, G.A., Mooney, R.A., Svetlov, V., Landick, R. and Artsimovitch, I., 2009. Functional specialization of transcription elongation factors. *The EMBO journal*, 28(2), pp.112-122.

Berg, H.C., 2008. *E. coli* in Motion. New York: Springer. pp.1-3.

Blatter, E.E., Ross, W., Tang, H., Gourse, R.L. and Ebright, R.H., 1994. Domain organization of RNA polymerase α subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell,* 78(5), pp.889-896.

Block, D.H., Hussein, R., Liang, L.W. and Lim, H.N., 2012. Regulatory consequences of gene translocation in bacteria. *Nucleic acids research*, 40(18), pp.8979-8992.

Brennan, C.A., Dombroski, A.J. and Platt, T., 1987. Transcription termination factor rho is an RNA-DNA helicase. *Cell,* 48(6), pp.945-952.

Brown, D.R., Barton, G., Pan, Z., Buck, M. and Wigneshweraraj, S., 2014. Nitrogen stress response and stringent response are coupled in *Escherichia coli*. *Nature communications,* 5.

Brown, N.L., Stoyanov, J.V., Kidd, S.P. and Hobman, J.L., 2003. The MerR family of transcriptional regulators. *FEMS microbiology reviews,* 27(2-3), pp.145-163.

Browning, D.F. and Busby, S.J., 2004. The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, 2(1), pp.57-65.

Browning, D.F. and Busby, S.J., 2016. The regulation of bacterial transcription initiation. *Nature Reviews Microbiology,* (in press).

Browning, D.F., Grainger, D.C. and Busby, S.J., 2010. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Current opinion in microbiology,* 13(6), pp.773-780.

Bryant, J.A., Sellars, L.E., Busby, S.J. and Lee, D.J., 2014. Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic acids research,* 42(18), pp.11383-11392.

Büchel, D.E., Gronenborn, B. and Müller-Hill, B., 1980. Sequence of the lactose permease gene. *Nature*, 283(5747), pp.541-545.

Buckle, M., Fritsch, A., Roux, P., Geiselmann, J. and Buc, H., 1991. [14] Kinetic studies on promoter-RNA polymerase complexes. *Methods in enzymology*, *208*, pp.236-258.

Burgess, R.R. and Anthony, L., 2001. How sigma docks to RNA polymerase and what sigma does. *Current opinion in microbiology*, 4(2), pp.126-131.

Burmann, B.M. and Rösch, P., 2011. The role of *E. coli* Nus-factors in transcription regulation and transcription: translation coupling: From structure to mechanism. *Transcription,* 2(3), pp.130-134.

Burmann, B.M., Luo, X., Rösch, P., Wahl, M.C. and Gottesman, M.E., 2009. Fine tuning of the *E. coli* NusB: NusE complex affinity to BoxA RNA is required for processive antitermination. *Nucleic acids research*, p.gkp736.

Burmann, B.M., Scheckenhofer, U., Schweimer, K. and Rösch, P., 2011. Domain interactions of the transcription–translation coupling factor *Escherichia coli* NusG are intermolecular and transient. *Biochemical Journal,* 435(3), pp.783-789.

Burmann, B.M., Schweimer, K., Luo, X., Wahl, M.C., Stitt, B.L., Gottesman, M.E. and Rösch, P., 2010. A NusE: NusG complex links transcription and translation. *Science,* 328(5977), pp.501-504.

Burr, T., Mitchell, J., Kolb, A., Minchin, S. and Busby, S., 2000. DNA sequence elements located immediately upstream of the–10 hexamer in *Escherichia coli* promoters: a systematic study. *Nucleic acids research,* 28(9), pp.1864-1870.

Busby, S.J. and Savery, N.J., 2007. Transcription activation at bacterial promoters. *eLS.*pp.1-9.

Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A. and Darst, S.A., 2001. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell,* 104(6), pp.901-912.

Carafa, Y.D.A., Brody, E. and Thermes, C., 1990. Prediction of rho-independent *Escherichia coli* transcription terminators: A statistical analysis of their RNA stem-loop structures. *Journal of molecular biology*, 216(4), pp.835-858.

Cerdeño-Tárraga, A.M., Patrick, S., Crossman, L.C., Blakely, G., Abratt, V., Lennard, N., Poxton, I., Duerden, B., Harris, B., Quail, M.A. and Barron, A., 2005. Extensive DNA inversions in the B. fragilis genome control variable gene expression. *Science,* 307(5714), pp.1463-1465.

Chakraborty, A., Wang, D., Ebright, Y.W., Korlann, Y., Kortkhonjia, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Jansen, R. and Nixon, B.T., 2012. Opening and closing of the bacterial RNA polymerase clamp. *Science,* 337(6094), pp.591-595.

Chan, B. and Busby, S., 1989. Recognition of nucleotide sequences at the *Escherichia coli* galactose operon P1 promoter by RNA polymerase. *Gene*, 84(2), pp.227-236.

Chander, M., Lee, A., Vallery, T.K., Thandar, M., Jiang, Y. and Hsu, L.M., 2015. Mechanisms of Very Long Abortive Transcript Release during Promoter Escape. *Biochemistry*, 54(50), pp.7393-7408.

Cho, B.K., Kim, D., Knight, E.M., Zengler, K. and Palsson, B.O., 2014. Genome-scale reconstruction of the sigma factor network in *Escherichia coli*: topology and functional states. *BMC biology*, 12(1), p.1.

Cho, B.K., Zengler, K., Qiu, Y., Park, Y.S., Knight, E.M., Barrett, C.L., Gao, Y. and Palsson, B.Ø., 2009. The transcription unit architecture of the *Escherichia coli* genome. *Nature biotechnology,* 27(11), pp.1043-1049.

Chung, H., Pamp, S.J., Hill, J.A., Surana, N.K., Edelman, S.M., Troy, E.B., Reading, N.C., Villablanca, E.J., Wang, S., Mora, J.R. and Umesaki, Y., 2012. Gut immune maturation depends on colonization with a host-specific microbiota. *Cell,* 149(7), pp.1578-1593.

Churchman, L.S. and Weissman, J.S., 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, *469*(7330), pp.368-373.

Cohan, F.M., 2001. Bacterial species and speciation. *Systematic biology*, 50(4), pp.513-524.

Condon, C., Squires, C. and Squires, C.L., 1995. Control of rRNA transcription in *Escherichia coli*. *Microbiological reviews*, 59(4), pp.623-645.

Cook, H. and Ussery, D.W., 2013. Sigma factors in a thousand *E. coli* genomes. *Environmental microbiology,* 15(12), pp.3121-3129.

Cozzone, A.J., 1988. Protein phosphorylation in prokaryotes. *Annual Reviews in Microbiology,* 42(1), pp.97-125.

Cramer, P., 2002. Multisubunit RNA polymerases. *Current opinion in structural biology,* 12(1), pp.89-97.

Cukras, A.R. and Green, R., 2005. Multiple effects of S13 in modulating the strength of intersubunit interactions in the ribosome during translation. *Journal of molecular biology,* 349(1), pp.47-59.

Daegelen, P., Studier, F.W., Lenski, R.E., Cure, S. and Kim, J.F., 2009. Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21 (DE3). *Journal of molecular biology*, 394(4), pp.634-643.

Dahm, R., 2005. Friedrich Miescher and the discovery of DNA. *Developmental biology,* 278(2), pp.274-288.

Darst, S.A., 2001. Bacterial RNA polymerase. *Current opinion in structural biology,* 11(2), pp.155-162.

de Lorenzo, V., Fernández, S., Herrero, M., Jakubzik, U. and Timmis, K.N., 1993. Engineering of alkyl-and haloaromatic-responsive gene expression with mini-transposons containing regulated promoters of biodegradative pathways of Pseudomonas. *Gene,* 130(1), pp.41-46.

Dennis, P.P. and Bremer, H., 1973. Regulation of ribonucleic acid synthesis in *Escherichia coli*Br: An analysis of a shift-up: 1. Ribosomal RNA chain growth rates. *Journal of molecular biology,* 75(1), pp.145-159.

Dethlefsen, L., Eckburg, P.B., Bik, E.M. and Relman, D.A., 2006. Assembly of the human intestinal microbiota. *Trends in Ecology & Evolution*, 21(9), pp.517-523.

DiBaise, J.K., Zhang, H., Crowell, M.D., Krajmalnik-Brown, R., Decker, G.A. and Rittmann, B.E., 2008. Gut microbiota and its possible relationship with obesity. *Mayo Clinic Proceedings*, 83(4), pp.460-469.

Dillon, R.J., Vennard, C.T., Buckling, A. and Charnley, A.K., 2005. Diversity of locust gut bacteria protects against pathogen invasion. *Ecology Letters,* 8(12), pp.1291-1298.

Dobzhansky, T., 1936. Position effects on genes. *Biological Reviews,* 11(3), pp.364-384.

Dolan, J.W., Marshall, N.F. and Richardson, J.P., 1990. Transcription termination factor rho has three distinct structural domains. *Journal of Biological Chemistry*, 265(10), pp.5747-5754.

Ebright, R.H., 1993. Transcription activation at class I CAP-dependent promoters. *Molecular microbiology,* 8(5), pp.797-802.

Ellinger, T., Behnke, D., Bujard, H. and Gralla, J.D., 1994. Stalling of *Escherichia coli* RNA polymerase in the+ 6 to+ 12 region *in vivo* is associated with tight binding to consensus promoter elements. *Journal of molecular biology*, *239*(4), pp.455-465.

El-Mowafi, S.A., Sineva, E., Alumasa, J.N., Nicoloff, H., Tomsho, J.W., Ades, S.E. and Keiler, K.C., 2015. Identification of inhibitors of a bacterial sigma factor using a new high-throughput screening assay. *Antimicrobial agents and chemotherapy*, *59*(1), pp.193-205.

Eron, L. and Block, R., 1971. Mechanism of Initiation of Repression of *In Vitro* Transcription of the *Lac* Operon of *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 68(8), pp.1828-1832.

Escherichia coli K-12 Substr. MG1655 Chromosome: 360,000 - 370,000". Ecocyc.org. N.p., 2016. Web. 30 Apr. 2016.

Feklistov, A. and Darst, S.A., 2011. Structural basis for promoter− 10 element recognition by the bacterial RNA polymerase σ subunit. *Cell,* 147(6), pp.1257-1269.

Feklistov, A., 2013. RNA polymerase: in search of promoters. *Annals of the New York Academy of Sciences*, 1293(1), pp.25-32.

Feklístov, A., Sharon, B.D., Darst, S.A. and Gross, C.A., 2014. Bacterial sigma factors: a historical, structural, and genomic perspective. *Annual review of microbiology*, *68*, pp.357-376.

Feng, Y., Zhang, Y., and Ebright, R.H., 2016. Structural basis of transcription activation.

Fenton, M.S., Lee, S.J. and Gralla, J.D., 2000. *Escherichia coli* promoter opening and− 10 recognition: mutational analysis of σ70. *The EMBO Journal,* 19(5), pp.1130-1137.

Friedman, D.I. and Court, D.L., 1995. Transcription antitermination: the λ paradigm updated. *Molecular microbiology,* 18(2), pp.191-200.

Fulcrand, G., Eichelbaum, S. and Leng, F., 2013. Lactose Repressor Functions as a DNA Topological Barrier in *Escherichia Coli* Lactose Operon. *Biophysical Journal*, *104*(2), pp.417a-418a.

Gatti-Lafranconi, P., Dijkman, W.P., Devenish, S.R. and Hollfelder, F., 2013. A single mutation in the core domain of the Lac repressor reduces leakiness. *Microbial cell factories,* 12(1), p.1.

Geiselmann, J., Yager, T.D., Gill, S.C., Calmettes, P. and Von Hippel, P.H., 1992. Physical properties of the *Escherichia coli* transcription termination factor rho. 1. Association states and geometry of the rho hexamer. *Biochemistry,* 31(1), pp.111-121.

Geszvain, K. and Landick, R., 2005. The structure of bacterial RNA polymerase. The bacterial chromosome. *ASM, Washington, DC*, pp.283-296.

Grainger, D.C., Hurd, D., Harrison, M., Holdstock, J. and Busby, S.J., 2005. Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proceedings of the National Academy of Sciences of the United States of America*, 102(49), pp.17693-17698.

Gries, T.J., Kontur, W.S., Capp, M.W., Saecker, R.M. and Record, M.T., 2010. One-step DNA melting in the RNA polymerase cleft opens the initiation bubble to form an unstable open complex. *Proceedings of the National Academy of Sciences,* 107(23), pp.10418-10423.

Gruber, T.M. and Gross, C.A., 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annual Reviews in Microbiology*, 57(1), pp.441-466.

Gusarov, I. and Nudler, E., 1999. The mechanism of intrinsic transcription termination. *Molecular cell,* 3(4), pp.495-504.

Gusarov, I. and Nudler, E., 2001. Control of intrinsic transcription termination by N and NusA: the basic mechanisms. *Cell,* 107(4), pp.437-449.

Ha, K.S., Toulokhonov, I., Vassylyev, D.G. and Landick, R., 2010. The NusA N-terminal domain is necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase. *Journal of molecular biology,* 401(5), pp.708-725.

Halasz, A., Kumar, V., Imieliński, M., Belta, C., Sokolsky, O., Pathak, S. and Rubin, H., 2007. Analysis of lactose metabolism in *E. coli* using reachability analysis of hybrid systems. *Systems Biology, IET*, 1(2), pp.130-148.

Hediger, M.A., Johnson, D.F., Nierlich, D.P. and Zabin, I., 1985. DNA sequence of the lactose operon: the *lacA* gene and the transcriptional termination region. *Proceedings of the National Academy of Sciences,* 82(19), pp.6414-6418.

Heldwein, E.E.Z. and Brennan, R.G., 2001. Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature*, 409(6818), pp.378-382.

Henkin, T.M. and Yanofsky, C., 2002. Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *Bioessays,* 24(8), pp.700-707.

Henkin, T.M., 2000. Transcription termination control in bacteria. *Current opinion in microbiology*, 3(2), pp.149-153.

Herbert, K.M., Zhou, J., Mooney, R.A., La Porta, A., Landick, R. and Block, S.M., 2010. *E. coli* NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. *Journal of molecular biology,* 399(1), pp.17-30.

Herring, C.D., Raffaelle, M., Allen, T.E., Kanin, E.I., Landick, R., Ansari, A.Z. and Palsson, B.Ø., 2005. Immobilization of *Escherichia coli* RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. *Journal of bacteriology*, *187*(17), pp.6166-6174.

Hinkle, D.C., Mangel, W.F. and Chamberlin, M.J., 1972. Studies of the binding of *Escherichia coli* RNA polymerase to DNA: IV. The effect of rifampicin on binding and on RNA chain initiation. *Journal of molecular biology*, 70(2), pp.209-220.

Hollands, K., Sevostiyanova, A. and Groisman, E.A., 2014. Unusually long-lived pause required for regulation of a Rho-dependent transcription terminator. *Proceedings of the National Academy of Sciences,* 111(19), pp.E1999-E2007.

Hooper, L.V. and Gordon, J.I., 2001. Commensal host-bacterial relationships in the gut. *Science,* 292(5519), pp.1115-1118.

Hsu, L.M., 2002. Promoter clearance and escape in prokaryotes. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 1577(2), pp.191-207.

Jacob, F. and Monod, J., 1961, January. On the regulation of gene activity. *Cold Spring Harbor Symposia on Quantitative Biology,* 26, pp 193-211.

Jacob, F. and Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology,* 3(3), pp.318-356.

Kalnins, A., Otto, K., Rüther, U. and Müller-Hill, B., 1983. Sequence of the *lacZ* gene of *Escherichia coli*. *The EMBO journal,* 2(4), p.593.

Kaper, J.B., Nataro, J.P. and Mobley, H.L., 2004. Pathogenic *Escherichia coli*. *Nature Reviews Microbiology,* 2(2), pp.123-140.

Kelly, J.R., Rubin, A.J., Davis, J.H., Ajo-Franklin, C.M., Cumbers, J., Czar, M.J., de Mora, K., Glieberman, A.L., Monie, D.D. and Endy, D., 2009. Measuring the activity of BioBrick promoters using an *in vivo* reference standard. *Journal of biological engineering*, 3(1), p.4.

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

Kercher, M.A., Lu, P. and Lewis, M., 1997. Lac repressor—operator complex. *Current opinion in structural biology,* 7(1), pp.76-85.

Kerppola, T.K. and Kane, C.M., 1991. RNA polymerase: regulation of transcript elongation and termination. *The FASEB Journal,* 5(13), pp.2833-2842.

Kingston, R.E. and Chamberlin, M.J., 1981. Pausing and attenuation of *in vitro* transcription in the *rrnB* operon of *E. coli*. *Cell,* 27(3), pp.523-531.

Knight, T., 2003. *Idempotent vector design for standard assembly of biobricks*. MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTELLIGENCE LAB.

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

Koslover, D.J., Fazal, F.M., Mooney, R.A., Landick, R. and Block, S.M., 2012. Binding and translocation of termination factor rho studied at the single-molecule level. *Journal of molecular biology,* 423(5), pp.664-676.

Kouzine, F. and Levens, D., 2007. Supercoil-driven DNA structures regulate genetic transactions. *Front Biosci,* 12(8-12), p.4409.

Kouzine, F., Gupta, A., Baranello, L., Wojtowicz, D., Ben-Aissa, K., Liu, J., Przytycka, T.M. and Levens, D., 2013. Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nature structural & molecular biology,* 20(3), pp.396-403.

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

Larson, M.H., Landick, R. and Block, S.M., 2011. Single-molecule studies of RNA polymerase: one singular sensation, every little step it takes. *Molecular cell*, 41(3), pp.249-262.

Larson, M.H., Mooney, R.A., Peters, J.M., Windgassen, T., Nayak, D., Gross, C.A., Block, S.M., Greenleaf, W.J., Landick, R. and Weissman, J.S., 2014. A pause sequence enriched at translation start sites drives transcription dynamics *in vivo*. *Science*, 344(6187), pp.1042-1047.

Lathe, W.C., Suyama, M. and Bork, P., 2002. Identification of attenuation and antitermination regulation in prokaryotes. *Genome biology,* 3(6), p.preprint0003.

Lee, D.J., Minchin, S.D. and Busby, S.J., 2012. Activating transcription in bacteria. *Annual review of microbiology,* 66, pp.125-152.

Leng, F., Chen, B. and Dunlap, D.D., 2011. Dividing a supercoiled DNA molecule into two independent topological domains. *Proceedings of the National Academy of Sciences*, 108(50), pp.19973-19978.

Lilley, D.M.J. and Higgins, C.F., 1991. Local DNA topology and gene expression: the case of the leu-500 promoter. *Molecular microbiology*, 5(4), pp.779-783.

Liu, L.F. and Wang, J.C., 1987. Supercoiling of the DNA template during transcription. *Proceedings of the National Academy of Sciences,* 84(20), pp.7024-7027.

Livak, K.J. and Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *methods*, *25*(4), pp.402-408.

Lodge, J., Fear, J., Busby, S., Gunasekaran, P. and Kamini, N.R., 1992. Broad host range plasmids carrying the Escherichia coli lactose and galactose operons. *FEMS microbiology letters*, *95*(2-3), pp.271-276.

Low, D.A., Weyand, N.J. and Mahan, M.J., 2001. Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. *Infection and Immunity*, 69(12), pp.7197-7204.

Luo, X., Hsiao, H.H., Bubunenko, M., Weber, G., Court, D.L., Gottesman, M.E., Urlaub, H. and Wahl, M.C., 2008. Structural and functional analysis of the *E. coli* NusB-S10 transcription antitermination complex. *Molecular cell,* 32(6), pp.791-802.

Mah, T.F., Kuznedelov, K., Mushegian, A., Severinov, K. and Greenblatt, J., 2000. The α subunit of *E. coli* RNA polymerase activates RNA binding by NusA. *Genes & development,* 14(20), pp.2664-2675.

Malan, T.P., Kolb, A., Buc, H. and McClure, W.R., 1984. Mechanism of CRP-cAMP activation of *lac* operon transcription initiation activation of the P1 promoter. *Journal of molecular biology*, *180*(4), pp.881-909.

Martin, F.H. and Tinoco, I., 1980. DNA-RNA hybrid duplexes containing oligo (dA: rU) sequences are exceptionally unstable and may facilitate termination of transcription. *Nucleic acids research*, 8(10), pp.2295-2300.

Martınez-Antonio, A. and Collado-Vides, J., 2003. Identifying global regulators in transcriptional regulatory networks in bacteria. *Current opinion in microbiology,* 6(5), pp.482-489.

Mattick, J.S. and Makunin, I.V., 2006. Non-coding RNA. *Human molecular genetics,* 15(suppl 1), pp.R17-R29.

McGary, K. and Nudler, E., 2013. RNA polymerase and the ribosome: the close relationship. *Current opinion in microbiology*, *16*(2), pp.112-117.

Mekler, V., Minakhin, L., Kuznedelov, K., Mukhamedyarov, D. and Severinov, K., 2012. RNA polymerase-promoter interactions determining different stability of the *Escherichia coli* and *Thermus aquaticus* transcription initiation complexes. *Nucleic acids research*, p.gks973.

Menzel, R. and Gellert, M., 1983. Regulation of the genes for E. coli DNA gyrase: homeostatic control of DNA supercoiling. *Cell*, *34*(1), pp.105-113.

Merino, E. and Yanofsky, C., 2005. Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends in Genetics*, 21(5), pp.260-264.

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

Mitchell, J.E., Zheng, D., Busby, S.J. and Minchin, S.D., 2003. Identification and analysis of 'extended–10'promoters in *Escherichia coli*. *Nucleic acids research*, 31(16), pp.4689-4695.

Monsalve, M., Calles, B., Mencía, M., Salas, M. and Rojo, F., 1997. Transcription activation or repression by phage φ29 protein p4 depends on the strength of the RNA polymerase–promoter interactions. *Molecular cell*, *1*(1), pp.99-107.

Mooney, R.A., Davis, S.E., Peters, J.M., Rowland, J.L., Ansari, A.Z. and Landick, R., 2009. Regulator trafficking on bacterial transcription units *in vivo*. *Molecular cell*, 33(1), pp.97-108.

Moxon, R., Bayliss, C. and Hood, D., 2006. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu. Rev. Genet.,* 40, pp.307-333.

Mukhopadhyay, J., Das, K., Ismail, S., Koppstein, D., Jang, M., Hudson, B., Sarafianos, S., Tuske, S., Patel, J., Jansen, R. and Irschik, H., 2008. The RNA polymerase "switch region" is a target for inhibitors. *Cell,* 135(2), pp.295-307.

Müller-Hill, B., Rickenberg, H.V. and Wallenfels, K., 1964. Specificity of the induction of the enzymes of the *lac* operon in *Escherichia coli*. *Journal of molecular biology*, 10(2), pp.303-318.

Murakawa, G.J., Kwan, C., Yamashita, J. and Nierlich, D.P., 1991. Transcription and decay of the *lac* messenger: role of an intergenic terminator. *Journal of bacteriology*, 173(1), pp.28-36.

Nielsen, A.A., Der, B.S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E.A., Ross, D., Densmore, D. and Voigt, C.A., 2016. Genetic circuit design automation. *Science*, 352(6281), p.aac7341-11.

Niu, W., Kim, Y., Tau, G., Heyduk, T. and Ebright, R.H., 1996. Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. *Cell*, 87(6), pp.1123-1134.

Nudler, E. and Gottesman, M.E., 2002. Transcription termination and anti-termination in E. coli. *Genes to Cells*, *7*(8), pp.755-768.

Nudler, E., 1999. Transcription elongation: structural basis and mechanisms. *Journal of molecular biology*, 288(1), pp.1-12.

Nudler, E., 2012. RNA polymerase backtracking in gene regulation and genome instability. *Cell,* 149(7), pp.1438-1445.

Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I. and Van Oudenaarden, A., 2004. Multistability in the lactose utilization network of *Escherichia coli*. *Nature*, 427(6976), pp.737-740.

Pagel, J.M., Winkelman, J.W., Adams, C.W. and Hatfield, G.W., 1992. DNA topology-mediated regulation of transcription initiation from the tandem promoters of the ilvGMEDA operon of *Escherichia coli*. *Journal of molecular biology*, 224(4), pp.919-935.

Paget, M.S. and Helmann, J.D., 2003. The sigma70 family of sigma factors. *Genome Biol*, 4(1), p.203.

Phan, T.T.P., Nguyen, H.D. and Schumann, W., 2012. Development of a strong intracellular expression system for Bacillus subtilis by optimizing promoter elements. *Journal of biotechnology*, 157(1), pp.167-172.

Philips, S.J., Canalizo-Hernandez, M., Yildirim, I., Schatz, G.C., Mondragón, A. and O'Halloran, T.V., 2015. Allosteric transcriptional regulation via changes in the overall topology of the core promoter. *Science*, 349(6250), pp.877-881.

Postollec, F., Falentin, H., Pavan, S., Combrisson, J. and Sohier, D., 2011. Recent advances in quantitative PCR (qPCR) applications in food microbiology. *Food microbiology*, *28*(5), pp.848-861.

Postow, L., Hardy, C.D., Arsuaga, J. and Cozzarelli, N.R., 2004. Topological domain structure of the Escherichia coli chromosome. *Genes & development*, *18*(14), pp.1766-1779.

Poteete, A.R., 2011. Recombination phenotypes of *Escherichia coli* greA mutants. *BMC molecular biology,* 12(1), p.1.

Prasch, S., Jurk, M., Washburn, R.S., Gottesman, M.E., Wöhrl, B.M. and Rösch, P., 2009. RNA-binding specificity of *E. coli* NusA. *Nucleic acids research*, 37(14), pp.4736-4742.

Qayyum, M.Z., Dey, D. and Sen, R., 2016. Transcription Elongation Factor NusA Is a General Antagonist of Rho-dependent Termination in *Escherichia coli*. *Journal of Biological Chemistry*, 291(15), pp.8090-8108.

Rabussay, D. and Zillig, W., 1969. A rifampicin resistent RNA-polymerase from *E. coli* altered in the β-subunit. *FEBS letters*, 5(2), pp.104-106.

Ray, A., Bellecourt, M.J. and Landick, R., 2016. Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annual Review of Biochemistry*, 85(1).

Reitzer, L. and Schneider, B.L., 2001. Metabolic Context and Possible Physiological Themes of ς54-Dependent Genes in *Escherichia coli*. *Microbiology and Molecular Biology Reviews*, 65(3), pp.422-444.

Revyakin, A., Liu, C., Ebright, R.H. and Strick, T.R., 2006. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science*, 314(5802), pp.1139-1143.

Reznikoff, W.S., 1992. The lactose operon-controlling elements: a complex paradigm. *Molecular microbiology*, *6*(17), pp.2419-2422.

Reznikoff, W.S., Siegele, D.A., Cowing, D.W. and Gross, C.A., 1985. The regulation of transcription initiation in bacteria. *Annual review of genetics*, 19(1), pp.355-387.

Richardson, J.P., 2002. Rho-dependent termination and ATPases in transcript termination. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression,* 1577(2), pp.251-260.

Riley, M., Abe, T., Arnaud, M.B., Berlyn, M.K., Blattner, F.R., Chaudhuri, R.R., Glasner, J.D., Horiuchi, T., Keseler, I.M., Kosuge, T. and Mori, H., 2006. *Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005. *Nucleic acids research,* 34(1), pp.1-9.

Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K. and Gourse, R.L., 1993. A Third Recognition Element in Bacterial Promoters: DNA Binding by the Subunit of RNA Polymerase. *Science*-New York Then Washington, 262, pp.1407-1407.

Ruff, E.F., Drennan, A.C., Capp, M.W., Poulos, M.A., Artsimovitch, I. and Record, M.T., 2015. *E. coli* RNA Polymerase Determinants of Open Complex Lifetime and Structure. *Journal of molecular biology*, 427(15), pp.2435-2450.

Saecker, R.M. and Record, M.T., 2011. Mechanism of bacterial transcription initiation: RNA polymerase-promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *Journal of molecular biology*, 412(5), pp.754-771.

Saha, A., Mukhopadhyay, J., Datta, A.B. and Parrack, P., 2015. Revisiting the mechanism of activation of cyclic AMP receptor protein (CRP) by cAMP in *Escherichia coli*: Lessons from a subunit-crosslinked form of CRP. *FEBS letters,* 589(3), pp.358-363.

Sahoo, M. and Klumpp, S., 2013. Backtracking dynamics of RNA polymerase: pausing and error correction. Journal of Physics: *Condensed Matter,* 25(37), p.374104.

Santangelo, T.J. and Artsimovitch, I., 2011. Termination and antitermination: RNA polymerase runs a stop sign. *Nature Reviews Microbiology*, *9*(5), pp.319-329.

Saunders, A., Core, L.J. and Lis, J.T., 2006. Breaking barriers to transcription elongation. *Nature Reviews Molecular Cell Biology,* 7(8), pp.557-567.

Savage, D.C., 1977. Microbial ecology of the gastrointestinal tract. *Annual Reviews in Microbiology*, 31(1), pp.107-133.

Schmid, M.B. and Roth, J.R., 1987. Gene location affects expression level in Salmonella typhimurium. *Journal of bacteriology*, 169(6), pp.2872-2875.

Schweimer, K., Prasch, S., Sujatha, P.S., Bubunenko, M., Gottesman, M.E. and Rösch, P., 2011. NusA interaction with the α subunit of *E. coli* RNA polymerase is via the UP element site and releases autoinhibition. *Structure*, 19(7), pp.945-954.
Science, (In press).

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

Shapiro, J., Machattie, L., Eron, L., Ihler, G., Ippen, K. and Beckwith, J., 1969. Isolation of pure lac operon DNA. *Nature*, *224*, pp.768-774.

Shultzaberger, R.K., Chen, Z., Lewis, K.A. and Schneider, T.D., 2007. Anatomy of *Escherichia coli* σ70 promoters. *Nucleic acids research*, 35(3), pp.771-788.

Silhavy, T.J. and Beckwith, J.R., 1985. Uses of *lac* fusions for the study of biological problems. *Microbiological reviews*, 49(4), p.398.

Sousa, C., de Lorenzo, V. and Cebolla, A., 1997. Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiology,* 143(6), pp.2071-2078.

Stefano, J.E. and Gralla, J.D., 1982. Spacer mutations in the *lac* ps promoter. *Proceedings of the National Academy of Sciences,* 79(4), pp.1069-1072.

Storz, G., Vogel, J. and Wassarman, K.M., 2011. Regulation by small RNAs in bacteria: expanding frontiers. *Molecular cell*, 43(6), pp.880-891.

Strobel, E.J. and Roberts, J.W., 2015. Two transcription pause elements underlie a σ70-dependent pause cycle. *Proceedings of the National Academy of Sciences*, 112(32), pp.E4374-E4380.

Suzuki, Y., Endo, M. and Sugiyama, H., 2015. Studying RNAP–promoter interactions using atomic force microscopy. *Methods,* 86, pp.4-9.

Sykes, R.B., Cimarusti, C.M., Bonner, D.P., Bush, K., Floyd, D.M., Georgopapadakou, N.H., Koster, W.H., Liu, W.C., Parker, W.L., Principe, P.A. and Rathnum, M.L., 1981. Monocyclic β-lactam antibiotics produced by bacteria. *Nature,* 291, pp.489-491.

Taliaferro, L.P., Keen, E.F., Sanchez-Alberola, N. and Wolf, R.E., 2012. Transcription Activation by *Escherichia coli* Rob at Class II Promoters: Protein–Protein Interactions between Rob's N-Terminal Domain and the σ 70 Subunit of RNA Polymerase. *Journal of molecular biology,* 419(3), pp.139-157.

Tlapák, H., Rydzewski, K., Schulz, T., Weschka, D., Schunder, E. and Heuner, K., 2017. Functional Analysis of the Alternative Sigma-28 Factor FliA and Its Anti-Sigma Factor FlgM of the Nonflagellated Legionella Species L. oakridgensis. *Journal of Bacteriology*, *199*(11), pp.e00018-17.

Ueguchi, C. and Mizuno, T., 1993. The *Escherichia coli* nucleoid protein H-NS functions directly as a transcriptional repressor. *The EMBO journal*, 12(3), p.1039.

Uptain, S.M., Kane, C.M. and Chamberlin, M.J., 1997. Basic mechanisms of transcript elongation and its regulation. *Annual review of biochemistry,* 66(1), pp.117-172.

VanGuilder, H.D., Vrana, K.E. and Freeman, W.M., 2008. Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques*, *44*(5), p.619.

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

Vidovic, S., Mangalappalli-Illathu, A.K. and Korber, D.R., 2011. Prolonged cold stress response of *Escherichia coli* O157 and the role of rpoS. *International journal of food microbiology,* 146(2), pp.163-169.

Vo, N.V., Hsu, L.M., Kane, C.M. and Chamberlin, M.J., 2003. *In vitro* studies of transcript initiation by *Escherichia coli* RNA polymerase. 3. Influences of individual DNA elements within the promoter recognition region on abortive initiation and promoter escape. *Biochemistry*, *42*(13), pp.3798-3811.

Vogel, U. and Jensen, K.F., 1994. Effects of guanosine 3', 5'-bisdiphosphate (ppGpp) on rate of transcription elongation in isoleucine-starved *Escherichia coli*. *Journal of Biological Chemistry,* 269(23), pp.16236-16241.

Vogel, U. and Jensen, K.F., 1997. NusA is required for ribosomal antitermination and for modulation of the transcription elongation rate of both antiterminated RNA and mRNA. *Journal of Biological Chemistry*, 272(19), pp.12265-12271.

Voliotis, M., Cohen, N., Molina-París, C. and Liverpool, T.B., 2009. Backtracking and proofreading in DNA transcription. *Physical review letters*, 102(25), p.258101.

Voliotis, M., Cohen, N., Molina-París, C. and Liverpool, T.B., 2008. Fluctuations, pauses, and backtracking in DNA transcription. *Biophysical journal,* 94(2), pp.334-348.

Voulgaris, J., French, S., Gourse, R.L., Squires, C. and Squires, C.L., 1999. Increased rrn gene dosage causes intermittent transcription of rRNA in *Escherichia coli*. *Journal of bacteriology,* 181(14), pp.4170-4175.

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

Vvedenskaya, I.O., Vahedian-Movahed, H., Bird, J.G., Knoblauch, J.G., Goldman, S.R., Zhang, Y., Ebright, R.H. and Nickels, B.E., 2014. Interactions between RNA polymerase and the "core recognition element" counteract pausing. *Science*, 344(6189), pp.1285-1289.

Wade, J.T., Roa, D.C., Grainger, D.C., Hurd, D., Busby, S.J., Struhl, K. and Nudler, E., 2006. Extensive functional overlap between σ factors in *Escherichia coli*. *Nature structural & molecular biology*, 13(9), pp.806-814.

Wang, J.C., 1985. DNA topoisomerases. *Annual review of biochemistry,* 54(1), pp.665-697.

Wang, J.C., 1996. DNA topoisomerases. *Annual review of biochemistry,* 65(1), pp.635-692.

Wang, M.D., Schnitzer, M.J., Yin, H., Landick, R., Gelles, J. and Block, S.M., 1998. Force and velocity measured for single molecules of RNA polymerase. *Science,* 282(5390), pp.902-907.

Wang, X., Llopis, P.M. and Rudner, D.Z., 2013. Organization and segregation of bacterial chromosomes. *Nature Reviews Genetics*, *14*(3), pp.191-203.

Waters, L.S. and Storz, G., 2009. Regulatory RNAs in bacteria. *Cell,* 136(4), pp.615-628.

Watson, J.D. and Crick, F.H., 1953. Molecular structure of nucleic acids. *Nature*, 171(4356), pp.737-738.

Welch, R.A., 2006. The genus *Escherichia*. In The prokaryotes (pp. 60-71). Springer New York.

Wilson, C.J., Zhan, H., Swint-Kruse, L. and Matthews, K.S., 2007. The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cellular and molecular life sciences*, 64(1), pp.3-16.

Wong, M.L. and Medrano, J.F., 2005. Real-time PCR for mRNA quantitation. *Biotechniques*, *39*(1), p.75.

Wösten, M.M.S.M., 1998. Eubacterial sigma-factors. *FEMS microbiology reviews*, 22(3), pp.127-150.

Wu, H.C., Tyson, K.L., Cole, J.A. and Busby, S.J., 1998. Regulation of transcription initiation at the *Escherichia coli* nir operon promoter: a new mechanism to account for co-dependence on two transcription factors. *Molecular microbiology,* 27(2), pp.493-505.

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E. and Ebright, R.H., 2012. Structural basis of transcription initiation. *Science,* 338(6110), pp.1076-1080.

Appendix I