Regulation of Transcription of the
*Escherichia coli* K5 Capsule Gene Cluster
Region One Promoter

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in Molecular Microbiology in the Faculty of Life Sciences

2014

JIA JIA
CHAPTER 1. INTRODUCTION ............................................................... 15

1.1 The *Escherichia coli* Capsule .................................................................. 15

1.2 Functions of Bacterial Capsule ........................................................................... 17
   1.2.1 Prevention of Desiccation ............................................................................... 17
   1.2.2 Adherence and Biofilm Formation ................................................................... 18
   1.2.3 Against the Host Immune System ..................................................................... 19

1.3 The Genetics and Biosynthesis of *Escherichia coli* Capsular Polysaccharides ................................................................. 21
   1.3.1 *E. coli* Group 1 Capsules ............................................................................... 22
   1.3.2 *E. coli* Group 4 Capsules ................................................................................. 25
   1.3.3 *E. coli* Group 3 Capsules ............................................................................... 29
   1.3.4 *E. coli* Group 2 Capsules ............................................................................... 29
     1.3.4.1 The Genetic Organization and expression of Group 2 Capsule Gene Cluster ........................................................................................................... 30
     1.3.4.2 The Biosynthesis of Group 2 Capsule Expression ...................................................... 32
     1.3.4.3 Export and Translocation of Group 2 Capsular Polysaccharide .................................................... 34

1.4 Mechanism of Bacterial Promoter Transcription ............................................................. 37
   1.4.1 Initiation of Transcription ................................................................................. 37
     1.4.1.1 RNA Polymerase Binding Site: Promoter .......................................................... 37
     1.4.1.2 RNA Polymerase .............................................................................................. 38
     1.4.1.3 Sigma-Factors ................................................................................................. 40
     1.4.1.4 THE SIGMA™ FACTOR .................................................................................. 41
     1.4.1.5 Steps of Transcription Initiation ....................................................................... 44
     1.4.2 Transcription Elongation .................................................................................. 47
     1.4.3 Regulation of Transcription Termination ............................................................. 47

1.5 Transcription Regulation in Bacteria ...................................................................... 49
   1.5.1 Regulation of Transcription Initiation in Bacteria ............................................... 49
     1.5.1.1 Sigma Factors Competition ............................................................................ 49
     1.5.1.2 Transcription Regulation by Transcription Factors .............................................. 50
   1.5.2 Transcription Regulation by Nucleoid Associated Proteins .................................. 54
     1.5.2.1 FIS ..................................................................................................................... 55
     1.5.2.2 H-NS ................................................................................................................. 56
1.5.2.3 IHF ................................................................. 59
1.5.2.4 DNA Superhelicity and Transcription Regulation .................. 61
1.5.3 Complex Promoters .................................................. 63
1.5.4 Transcriptional Regulation in E. coli Group 2 Capsule Gene Cluster .......................................................... 64
1.5.4.1 SlyA-H-NS Functional Interaction in the Regulation of Transcription .......................................................... 65
1.5.4.2 BipA .................................................................. 66
1.5.4.3 IHF Regulates K5 Capsule Gene Cluster ............................ 68
1.5.4.4 Other Factors Involved in Thermoregulation of Promoter Transcription .......................................................... 68

1.6 Examples of Regulation of Capsule Gene in Other Gram-negative Bacteria ......................................................... 69
1.6.1 Salmonella enterica serovar Typhi ....................................... 69
1.6.2 Neisseria meningitidis ..................................................... 70

CHAPTER 2. MATERIALS AND METHODS ................................................. 74

2.1 Bacterial Strains, Media and Growth Conditions .............................................. 74
2.2 Chemicals and solutions .............................................................................. 74
2.3 Strain Maintenance and Growth Conditions ................................................. 74

2.4 DNA manipulation ....................................................................................... 78
2.4.1 Plasmid preparation and purification .................................................. 78
2.4.2 Quantification of DNA or RNA samples ............................................. 78
2.4.3 Restriction Endonuclease Digestion of DNA ....................................... 78
2.4.4 Ligation Reaction .................................................................................. 78

2.5 Competent cells preparation ....................................................................... 79
2.5.1 Electropotent E. coli Cells ................................................................. 79
2.5.2 Chemically Competent DH5α Cells ...................................................... 79

2.6 Transformation of Chemically Competent E. coli Cells ................................ 80

2.7 Transformation of Electrocompetent E. coli Cells ........................................ 80

2.8 Agarose gel electrophoresis ......................................................................... 80

2.9 Polymerase Chain Reaction (PCR) .............................................................. 81

2.10 Colony PCR ............................................................................................. 85

2.11 DNA Sequencing ....................................................................................... 85

2.12 Beta-Galactosidase Assay ......................................................................... 86

2.13 RNA Extraction and Clean up ..................................................................... 87

2.14 Rapid Amplification of 5’ cDNA ends (5’ RACE) ....................................... 88

2.15 Modified 5’RACE- Capillary electrophoresis using a DNA sequencer .......... 89

2.16 pGEM-T Vector System ........................................................................... 89
2.16.1 Ligation Protocol ............................................................................... 89
2.16.2 Transformation Protocol ..................................................................... 90
2.16.3 Blue/White Selection of Recombinants ............................................... 90

2.17 Site Direct Mutagenesis by PCR ................................................................. 90

2.18 Generalized Transduction of E. coli ............................................................ 91
2.18.1 Preparation of P1 lysate ......................................................... 91
2.18.2 Determination of P1 Titre ...................................................... 92
2.18.3 Transduction ........................................................................ 92

2.19 Protein Purification and Analysis ............................................. 93
2.19.1 IHF over-expression and collection ........................................ 93
2.19.2 Sonication cell lysis ............................................................. 93
2.19.3 Ammonium sulfate precipitation and dialysis ....................... 93
2.19.4 Affinity chromatography purification .................................. 94
2.19.5 SDS-polyacrylamide gel electrophoresis (SDS-PAGE) .......... 94
2.19.6 Protein Quantification .......................................................... 95

2.20 Electrophoretic Mobility Shift Assay (EMSA) .......................... 96
2.21 Quantitative Real-time Polymerase Chain Reaction (qRT PCR) .................................................................. 96
2.22 Immunofluorescent staining and imaging ................................ 98

CHAPTER 3. IDENTIFICATION AND CHARACTERIZATION OF MULTIPLE PROMOTERS AT PROMOTER 1 REGION (PR1) OF THE E. COLI K5 ANTIGEN GENE CLUSTER .................. 99

3.1 Three additional promoters were determined in the 5’ Untranslated Region (5’ UTR) at PR1 .......... 99
  3.1.1 Introduction ........................................................................... 99
  3.1.2 Generating E. coli K5 PR1 promoter region, 5’ UTR region and upstream promoter region lacZ fusions ......................................................... 99
  3.1.3 The transcription from 5’ UTR showed modest transcription activity at 37 °C .................................................. 102
  3.1.4 Determination of the potential transcription start sites in the 5’ UTR of Region 1 Promoter by 5’ RACE assay ........................................................................... 104

3.2 Evaluating the relative transcriptional activities of multiple promoters ............................................. 111
  3.2.1 Construction of E. coli K5 region 1 minimal promoter lacZ fusions ......................................................... 111
  3.2.2 Transcriptional fusions confirm putative promoter functionality ............................................................... 113

3.3 Identification of functional elements of PR1 region tandem promoters .................................................. 115
  3.3.1 Generating single or multiple sites mutation in the predicted motif of putative promoters ......... 115
  3.3.2 Mutation of the second T of PR1-1 promoter -10 Pribnow box could abolished 95% of transcription activity ............................................................................ 117
  3.3.3 The transcription activity with 165A>C and 138G>A mutation showed significant reduction at PR1-4 respectively ................................................................. 118

3.4 Analysis of transcriptional level at PR1 from two major promoter PR1-1 and PR1-4 ...................... 121
  3.4.1 Quantification of the relative contribution of each transcriptional start site to the total transcript at 37 °C ............................................................................ 121
  3.4.2 Destruction of promoter PR1-1 caused attenuation of PR1-4 transcriptional level .................. 125

3.5 Discussion .................................................................................. 128
  3.5.1 Observation the multiple tandem promoters at PR1 5’ UTR region ................................................................. 128
  3.5.2 Identification of the promoter elements for multiple promoters .............................................................. 131

CHAPTER 4. INVESTIGATING HOW THE MULTIPLE PROMOTERS AT PR1 OF THE E. COLI K5 GENE CLUSTER ARE REGULATED BY A TRANSCRIPTIONAL REGULATOR(S) .......... 134

4.1 Role of H-NS in transcription of PR1 region of K5 capsule gene ....................................................... 134
  4.1.1 H-NS as a negative regulator of the capsule gene cluster .......................................................................... 134
  4.1.1.1 H-NS regulates transcription at PR1 negatively and is important for temperature regulation at PR1 ................................................................................... 134
  4.1.1.2 In vivo transcription of PR1 tandem promoters were influenced negatively by H-NS .......... 140
  4.1.2 EMSA analysis of H-NS binding region at PR1 ................................................................................. 144
4.1.3 The integrity of PR1 region is important for promoter PR1-1 temperature regulation which may related to H-NS binding.......................................................................................................................... 147

4.2 The role of IHF in regulating transcription at PR1 ................................................................. 149
  4.2.1 IHF play a positive role in regulating the transcription at PR1-1 but repressing the transcription from PR1-4.......................................................................................................................... 149
  4.2.2 Purification of IHF .............................................................................................................. 151
  4.2.3 IHF is also directly involved in the transcriptional regulation at PR1 promoters .......... 153
    4.2.3.1 EMSA analysis of IHF binding at PR1 ........................................................................ 153
    4.2.3.2 Destruction of IHF binding site at +130 of PR1 region by Site-direct Mutagenesis .... 155
    4.2.3.3 The binding of IHF may affect the temperature regulation at PR1-4 when the promoter PR1-4 was separated out from the PR1 region ......................................................................................... 161
  4.2.4 Analysis the activation of promoter PR1-4 ........................................................................ 162
    4.2.4.1 Design and characterization of artificial Rho-independent transcriptional terminators... 163
    4.2.4.2 Transcription at promoter PR1-1 can simulate downstream promoter PR1-4 transcription activation .......................................................................................................................... 167

4.3 Role of SlyA in transcription of PR1 region of K5 capsule gene ........................................... 170

4.4 Discussion ............................................................................................................................... 173
  4.4.1 H-NS negatively regulating transcription from PR1 .......................................................... 173
  4.4.2 SlyA may antagonizes H-NS-mediated silencing at PR1 ..................................................... 176
  4.4.3 The role of IHF in regulating PR1 transcription .................................................................... 177
  4.4.4 Transcription from promoter PR1-4 is dependent on promoter PR1-1 activation .......... 179

CHAPTER 5. STUDYING THE TRANSCRIPTION FROM MULTIPLE PROMOTERS AT PR1 DURING GROWTH PHASE ....................................................................................................... 183

5.1 Analysis the time courses of multiple promoters transcriptional activity at PR1 in E. coli K5 strain... 183

5.2 Examining transcriptional level of PR1-1 and PR1-4 at PR1 in strain P90C(pJPR1) during growth phase at 37 °C by qRTPCR .............................................................................................................................................. 185

5.3 Examining the transcriptional activity of PR1-1 and PR1-4 at PR1 during the growth phase following a temperature shift from 20 °C to 37 °C ....................................................................................... 187

5.4 Discussion ............................................................................................................................... 195

CHAPTER 6. GENERAL DISCUSSION ......................................................................................... 199

APPENDIX ................................................................................................................................. 232

Final Word Count: 55,277
INDEX OF FIGURES

Figure 1. 1 Diagrammatic drawing of Gram-negative bacterial cell wall. ................................................. 17
Figure 1. 2 A model for biosynthesis and assembly of group 1 and 4 capsules (Whitfield, 2006). ........................................... 24
Figure 1. 3 Genetic organization of E. coli capsule gene cluster in groups 1(K30), 2(K5), 3(K10) and 4(K40). .............................................................................................................................................. 27
Figure 1. 4 Genetic organization of the E. coli K5 capsule gene cluster......................................................... 28
Figure 1. 5 A model for biosynthesis and assembly of group 2 capsules (Whitfield, 2006). .... 36
Figure 1. 6 Promoter recognition by σ70 and a subunit ................................................................................. 43
Figure 1. 7 Structure of complex between E. coli σ70 RNAP holoenzyme and interaction s between promoter and DNA. ........................................................... 43
Figure 1. 8 RNA polymerase interactions at a promoter and the pathway of transcription initiation. ........................................................................................................................................ 46
Figure 1. 9 Transcription activation at simple promoter. ................................................................................. 52
Figure 1. 10 Transcription repression. ........................................................................................................... 54
Figure 1. 11 Model of H-NS silencing mechanism. ....................................................................................... 58
Figure 1. 12 Structure of the IHF-DNA complex ......................................................................................... 60
Figure 2. 1 Schematic picture of 5’RACE work flow ..................................................................................... 89
Figure 2. 2 Schematic picture of site-directed mutagenesis (Loening, 2005) .................................................. 91
Figure 3. 1 Generation of PR1 transcription fusion JJPR1, JJPR2, 5′ UTR region lacZ fusions pJJ1 and upstream promoter region lacZ transcriptional fusion pJJ2 .................................................................. 101
Figure 3. 2 Restriction enzyme digestions of the purified plasmids pJJ1 & pJJ2 from strain P90C showing from the agarose gel .............................................................................. 102
Figure 3. 3 β-galactosidase activities of PR1 transcriptional fusions JJPR1, JJPR2, pJJ1 and pJJ2 grown at 37° C and 20° C ............................................................................................ 103
Figure 3. 4 Schematic representation of the strategy used for mapping transcription starts sites by 5′RACE assay ........................................................................................................ 105
Figure 3. 5 The 5’ RACE reaction using mRNA extracted from strain P90C (pJJ1) grown at 37° C ................................................................. 106
Figure 3. 6 Colony PCR for DH5a bearing plasmid pGEM-T Easy ligated with P90C (pJJ1) second round PCR products in 5′ RACE assay. .................................................................................. 106
Figure 3. 7 The detail of putative transcription start sites analyzed by 5′ RACE. ........................................ 108
Figure 3. 8 Analysis the putative kpsF promoter within the E. coli K1 capsule gene showing on the polyacrylamide sequencing gel, Cieslewicz and Vmir (1996). ................................. 109
Figure 3. 9 Portions of different sizes of amplified DNA fragments with 6-FAM modified SP2 primer in modified 5’ RACE assay. .......................................................................................................................... 110
Figure 3. 10 Construction map of transcriptional lacZ fusion inserts. ......................................................... 112
Figure 3. 11 β-galactosidase activities of different minimal putative promoter-lacZ fusions grown at 37° C and 20° C. ........................................................................................................... 114
Figure 3. 12 Schematic representation of constructing Site-Direct Mutagenesis plasmids in PR1 region. .............................................................................................................................................. 116
Figure 3. 13 Site direct mutation analysis of the promoter PR1-1 with second T to C substitution at -10 hexamer................................................................................................................. 118
Figure 3. 14 Nucleotide sequence and map of the predicted promoter functional elements. ........................... 120
Figure 3. 15 Illustration of qRT-PCR primers and corresponding amplicons performed in qRT-PCR assay of strain P90C (pJJPR1). .................................................................................................................. 123
Figure 3. 16 Standard curves for 1-1A, 1-1, 1-4 and rpoD amplicons performed in qRT-PCR assay. ................................................................. 123

Figure 3. 17 The transcript copy number of promoters PR1-1 and PR1-4 in promoter destruction mutants grown at 37 °C. ........................................ 127

Figure 4. 1 Colony PCR screening of P90C and P90C hns::kan transductants. ................... 135

Figure 4. 2 The effect of the hns:kan mutants on region 1 promoter activity at 37 °C and 20 °C in the vector pRS415 background. ........................................ 137

Figure 4. 3 The effect of the hns:kan mutants on region 1 promoter activity at 37 °C and 20 °C in the vector pCB192 background. ........................................ 137

Figure 4. 4 Standard curves for MS101-kpsf amplicon performed in qRT-PCR assay. .......... 139

Figure 4. 5 H-NS represses the transcription of PR1 promoter region of E. coli K5 capsule gene cluster. .............................................................................. 139

Figure 4. 6 Different minimal promoter- lacZ fusion inserts .............................................. 141

Figure 4. 7 H-NS is repressed the transcription at PR1 in vivo at 37 °C. .............................. 142

Figure 4. 8 H-NS is repressed the transcription at PR1 in vivo at 20 °C. ............................. 143

Figure 4. 9 DNA fragments used in analysis of H-NS binding at PR1 region ..................... 146

Figure 4. 10 H-NS binds to PR1 upstream region but not UTR region in vitro. .................. 146

Figure 4. 11 Comparison the different PR1-1 transcriptional lacZ fusion inserts. ............... 148

Figure 4. 12 Temperature regulation of PR1-1 require both upstream and downstream region of PR1-1 .............................................................. 148

Figure 4. 13 IHF is essential for the transcription at PR1 in vivo at 37 °C. .......................... 150

Figure 4. 14 IHF elution profiles of HiTrap Heparin HP (1 ML) fractions ............................... 152

Figure 4. 15 Coomassie blue stained SDS-PAGE analysis of IHF protein induction .......... 152

Figure 4. 16 Coomassie blue stained SDS-PAGE indicated purified IHF protein ................. 152

Figure 4. 17 IHF sequence for EMSA analysis ............................................................... 154

Figure 4. 18 EMSA analysis of purified IHF ...................................................................... 154

Figure 4. 19 Inactivation of IHF binding sites by site-direct mutagenesis at PR1 +130 region. 156

Figure 4. 20 Inactivation of IHF binding sites by site-direct mutagenesis at PR1 +130 region. 156

Figure 4. 21 IHF binding of IHF was directly repressed the transcriptional activity of promoter PR1-1 and PR1-4. ..................................................... 161

Figure 4. 22 IHF binding of IHF was directly repressed the transcriptional activity of promoter PR1-1 and PR1-4. ..................................................... 161

Figure 4. 23 PR1-4 lost temperature regulation in the absence of IHF binding. ................. 162

Figure 4. 24 Schematic representation of the strategy used for constructing PR1 region inserted with Rho-independent terminator .................................................. 165

Figure 4. 25 Nucleotide sequence of the 5’ end upstream region of the kpsf gene with Rho-independent terminator at NcoI site. ................................. 166

Figure 4. 26 Illustration of qRT-PCR primers and corresponding amplicons performed in qRT-PCR assay of strain P90C (JP1T) and P90C (JP1T-IHF-BSM). .............. 168

Figure 4. 27 Standard curves for 1_1’ amplicon in JP1T and JP1T-IHF-BSM performed in qRT-PCR assay ................................................................. 168

Figure 4. 28 Transcription initiation from PR1-4 was dependent on the activation of promoter P1-1. ................................................................. 169

Figure 4. 29 SlyA is required for the transcription from PR1-1 in vivo at 37 °C. .................... 171

Figure 4. 30 SlyA has no effect on the transcription at PR1 region in vivo at 20 °C. .......... 172

Figure 4. 31 A hypothesized model of activation of PR1-4 is dependent on the transcription activation of PR1-1 ..................................................... 182
Figure 5.1 Analysis of multiple promoters at PR1 transcriptional activity during growth at 37 °C. .......................................................... 184
Figure 5.2 Transcripts level of PR1-1 and PR1-4 measured by qRT-PCR in strain pJPR1 (P90C) during growth phase at 37 °C. .......................................................... 186
Figure 5.3 Illustration of qRT-PCR primers and corresponding amplicons performed in qRT-PCR assay of strain UTI89. .................................................................................. 189
Figure 5.4 Standard curves for 1-1A, 1-1, 1-4 and 16s amplicons performed in qRT-PCR assay. ........................................................................................................ 189
Figure 5.5 Time course of PR1 promoter transcription when the cells grown at 20 °C continuously. ............................................................................... 191
Figure 5.6 Time course of PR1 promoter transcription when the cells grown at 37 °C continuously. ............................................................................... 191
Figure 5.7 Time course of PR1 promoter transcription following a temperature upshift from 20 °C to 37 °C. .................................................................................. 192
Figure 5.8 Immunofluorescence microscopy of strain UTI89 using K1-specific monoclonal antibody following a temperature upshift from 20 °C to 37 °C. .................................................................................. 194
Figure 6.1 Regulation of the Region 1 Promoter of Escherichia coli K5 at 37 °C. .......... 207
# INDEX OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>Classification of <em>Escherichia coli</em> capsules (Whitfield, 2006)</td>
<td>22</td>
</tr>
<tr>
<td>Table 1.2</td>
<td>Sigma factors of <em>Escherichia coli</em></td>
<td>41</td>
</tr>
<tr>
<td>Table 2.1</td>
<td>Strains Used During the Course of this Study</td>
<td>75</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>Plasmids Used During the Course of this Study</td>
<td>76</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>Primers Used During the Course of this Study</td>
<td>85</td>
</tr>
<tr>
<td>Table 2.4</td>
<td>Constituents of polyacrylamide gels</td>
<td>95</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Summarization of detected transcription start sites from different constructs used in 5′RACE assay</td>
<td>108</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Identification of the functional elements of the PR1-4 promoter</td>
<td>120</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>Transcripts copy number of the promoters PR1-1 and PR1-4 in pJPR1 (P90C) mid-log cultures grown at 37°C</td>
<td>125</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Terminator Sequence used for this study</td>
<td>164</td>
</tr>
</tbody>
</table>
Abstract

Submitted by JIA JIA for the Degree of Doctor of Philosophy in The University of Manchester and entitled ‘Regulation of Transcription of the *Escherichia coli* K5 Capsule Gene Cluster Region One Promoter’ in September, 2014

Encapsulated *Escherichia coli* are responsible for a number of life threatening infections of man. These range from urinary tract infections to septicemia and neonatal meningitis. A common property of these *E. coli* strains is the expression of a polysaccharide capsule or K antigen. The expression of a capsule is an essential virulence factor protecting the bacterium from host defenses. Like many virulence factors capsule gene expression is regulated by temperature, such that at 37 °C inside the host the capsule is expressed whereas at 20 °C it is not.

The project used the K5 capsule gene cluster as a model system to study in detail the regulation of capsule gene expression. Expression of *E. coli* K5 gene cluster is regulated at the transcriptional level by two convergent promoters PR1 and PR3. The temperature regulation-dependent expression is in part controlled at the level of transcription by complex regulatory network involving the regulators SlyA, H-NS and IHF acting at PR1 and PR3. A large 5’ untranslated region (5’ UTR) is involved in transcriptional regulation by interacting with global regulator proteins.

In this study, a combination of lacZ reporter gene fusions, 5’ RACE analysis and site-direct mutagenesis at promoter functional elements were used to investigate the promoter. These studies identified that the PR1 promoter was more complex than initially thought and contains, in addition to previously characterized PR1-1 promoter at +1, three additional tandem promoters PR1-2, PR1-3 and PR1-4 transcribing in the same direction from the site +133, +142 and +182, respectively. In order to analyse the contribution for the transcription from PR1 among these multiple promoters, these multiple tandem promoters’ activities were measured by β-galactosidase assay and Real-time quantitative reverse PCR assay. We determined that PR1-2 and PR1-3 are two cryptic promoters with very low transcription activity while PR1-1 and PR1-4 are the major promoters that contributed evenly to the total transcripts into *kps* operon in the mid-exponential phase. Furthermore, we demonstrated that the promoter PR1-1 and PR1-4 are tightly coupled and the activity of PR1-4 can be co-ordinately reduced by disrupted PR1-1.

Different minimal PR1-lacZ promoter fusions were also transformed into strains with mutations in the genes that encode these regulatory proteins (IHF, SlyA and H-NS) and the transcription activity was examined by β-galactosidase assay at both 37 °C and 20 °C. IHF is required indirectly for maximum transcription at PR1-1 promoter but directly represses transcription from PR1-4 due to binding at +160 region at 37 °C. Global regulator H-NS represses the transcription at both 37 °C and 20 °C at PR1 and plays an important role for transcriptional temperature regulation at PR1 region. The anti-repressor SlyA activates transcription at PR1-1 at 37 °C.

This study identified for the first time growth phase dependent expression from the PR1 promoter. Also, this study discovered different temporal patterns of promoter PR1-1 and PR1-4 transcription was coordinated with bacterial growth cycle. Overall this study will be helpful to decipher the complex regulation of capsule gene expression in *E. coli*.
Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses.
Acknowledgement

First of all, I would like to express my deep and sincere gratitude to my supervisor Professor Ian S. Roberts, who provided me with the precious opportunity to study for the Ph.D degree in Molecular Microbiology at the University of Manchester. Without his supervision, encouragement and valuable input during the course of my Ph.D, this doctoral thesis would not have been achievable.

I would also like to thank Dr. Nicky High for her instructions and encouragement during all stage of my works. I wish to thank Dr. David Corbett and Dr. Jane King for their helpful advice and assistance to my project. I would like to thank Temitope Adeyemi, Emily White, Hasan Aal Owaif, Eva Haas, Tomas Pointon, Warren Flood and everyone else in lab C. 1202 for their assistance and cooperation. Special thanks to Dr. Ashley Houlden, Dr. Adrian Jervis, Dr. James Thompson and Marie Goldrick, for their assistance and advice on experimental materials and protocols.

I am also deeply grateful to my best friends Dr. Jiahui Wang, Dr. Yonghui Ma and Siyao Wang. Without the help and support of the kind people around me, my Ph.D life would not be fun and productive. My deepest gratitude should give to my parents for their constant great financial and emotional help and support during my study abroad.

Finally, great gratitude is given to China Scholarship Council (CSC) for funding my Ph.D study.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC-2</td>
<td>ABC-2 ATP-binding cassette type 2</td>
</tr>
<tr>
<td>Amp</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>cAMP</td>
<td>cyclic adenosine monophosphate</td>
</tr>
<tr>
<td>C4bp</td>
<td>C4 binding protein</td>
</tr>
<tr>
<td>CPS</td>
<td>capsular polysaccharide</td>
</tr>
<tr>
<td>CPR</td>
<td>Catabolite regulatory protein</td>
</tr>
<tr>
<td>DEPC</td>
<td>Diethyl pyrocarbonate</td>
</tr>
<tr>
<td>EC</td>
<td>Elongation complex</td>
</tr>
<tr>
<td>ECA</td>
<td>Enterobacterial common antigen</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylendiaminetetraacetic acid</td>
</tr>
<tr>
<td>EPS</td>
<td>Extracellular polysaccharide</td>
</tr>
<tr>
<td>GlcA</td>
<td>Glucuronic acid</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetyl glucosamine</td>
</tr>
<tr>
<td>GT</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>HhH</td>
<td>Helix-Hairpin-helix</td>
</tr>
<tr>
<td>IL8</td>
<td>Interleukin-8</td>
</tr>
<tr>
<td>JUMPStart</td>
<td>Just Upstream of Many Polysaccharide Gene Starts</td>
</tr>
<tr>
<td>Kdo</td>
<td>2-keto-3-deoxymanno-octonic acid</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>MAC</td>
<td>Membrane attack complex</td>
</tr>
<tr>
<td>MAP</td>
<td>mitogen-activated protein</td>
</tr>
<tr>
<td>ManNAc</td>
<td>N-acetyl mannosamine</td>
</tr>
<tr>
<td>MES</td>
<td>2-(N-morpholino) ethanesulfonic acid</td>
</tr>
<tr>
<td>MPA2</td>
<td>Cytoplasmic-membrane-periplasmic auxiliary</td>
</tr>
<tr>
<td>MS</td>
<td>Membrane-spanning</td>
</tr>
<tr>
<td>NeuNac</td>
<td>N-acetyl neuraminic acid</td>
</tr>
<tr>
<td>ONPG</td>
<td>O-nitrophenyl-β-D-galactosidase</td>
</tr>
<tr>
<td>ops</td>
<td>Operon polarity supressor</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PG</td>
<td>phosphatidyglycerol</td>
</tr>
<tr>
<td>ppGpp</td>
<td>guanosine tetraphosphate</td>
</tr>
<tr>
<td>PST</td>
<td>polysialyltransferase</td>
</tr>
<tr>
<td>RPo</td>
<td>RNA polymerase open complex</td>
</tr>
<tr>
<td>Rpc</td>
<td>RNA polymerase close complex</td>
</tr>
<tr>
<td>RNAP</td>
<td>RNA polymerase</td>
</tr>
<tr>
<td>SD</td>
<td>Shine-Dalgarno Sequence</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SDM</td>
<td>Site-direct mutagenesis</td>
</tr>
<tr>
<td>SPI-7</td>
<td>Salmonella pathogenicity island 7</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate-EDTA</td>
</tr>
<tr>
<td>TE</td>
<td>Tris-EDTA</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription start site</td>
</tr>
<tr>
<td>TNFα</td>
<td>Tumor necrosis factor-alpha</td>
</tr>
<tr>
<td>Und-PP</td>
<td>Undecaprenyl phosphate</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>UPEC</td>
<td>Uropathogenic Escherichia coli</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-bromo-4-chloro-3-indolyl- β-galactoside</td>
</tr>
<tr>
<td>5’ RACE</td>
<td>Rapid amplifying the 5’ end of the cDNA</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction

1.1 The *Escherichia coli* Capsule

Both Gram-positive and Gram-negative bacteria may produce cell-surface polysaccharides. These polysaccharides can form an amorphous layer of extracellular polysaccharide coating the outer surface which may further be formed into a distinct structure named the capsule (Roberts, 1996). *Escherichia coli* are capable of causing a range of diseases in humans and animals, including gastro-intestinal and urinary tract infections, meningitis, and septicaemia. A common feature of *E. coli* isolates from extraintestinal infections is the expression of a polysaccharide capsule (Corbett & Roberts, 2008). The term ‘capsule’ commonly indicates an extracellular polysaccharide that is physically linked to the cell surface via covalent attachment (Corbett & Roberts, 2008). The major components of bacterial capsules are highly hydrated, high molecular weight acidic polysaccharides that confer upon bacteria an overall negative charge and hydrophilic properties (Orskov *et al*., 1977).

There is a great structural diversity in capsular polysaccharides both among different bacterial species but also among the strains of the same bacterial species. This diversity is an outcome of not only differences in the repeat monosaccharide components but also differences in linkage between the monosaccharides (Roberts, 1996). *E. coli* is capable of expressing a diverse range of cell surface polysaccharides. The cell surface polysaccharides of *E. coli* are serotype-specific and are in two general forms: lipopolysaccharide (LPS) O antigen and capsular polysaccharide (K antigen) (Figure 1.1). The O antigen is a thermostable surface antigen found in all smooth *Enterobacteriaceae* (Whitfield, 2006). It is distinguished from LPS by the absence of terminal lipid A-core. LPS is a unique component of the outer membrane of Gram-negative bacteria (Whitfield & Valvano, 1993) and is composed of three regions: proximal lipid-A region, distal O-specific polysaccharide chain (O antigen) and the core oligosaccharide connecting lipid-A and O-antigen (Jann and Jann, 1997). Lipid A and core oligosaccharide are
two conserved constituent domains of LPS, the third being the highly variable O antigen. All three components are synthesized separately and ligated together (Jann and Jann, 1997).

The expression of K antigens is strongly associated with certain infections. For example, expression of the K1 capsule is common among isolates of *E. coli* that cause neonatal meningitis and urinary tract infection (Robino *et al*., 2013; Barichello *et al*., 2014). The K5 polysaccharide is associated with urinary tract infection and sepsis (Devine *et al*., 1989; Ullerød *et al*., 1994). Different O and K serotypes contain a variety of different sugar components as well as differences in the glycosidic bonds between repeating sugar residues. Further some polysaccharides may be substituted with non-carbohydrate residues (Orskov *et al*., 1977). Variations in structures of these polysaccharides contribute to form 167 different O antigens and more than 80 K antigens in *E. coli* and the primary structures of many of these antigens have been determined (Roberts, 1996).

In addition, there are two non-serotype specific surface polysaccharides produced by most *E. coli* strains, namely colanic acid (M antigen) and the enterobacterial common antigen (ECA). Colanic acid (M antigen) is a polymer of glucose, galactose, fucose and glucuronic acid modified with acetyl and pyruvyl groups, which has been widely found within *E. coli* (Gottesman & Stout, 1991). It has been suggested that the M antigen is important for survival of *E. coli* outside the host and perhaps plays a role in its resistant to desiccation (Ophir & Gutnick, 1994). Enterobacterial common antigen (ECA) is a surface antigen of all enteric bacteria. The biological significance of ECA is not fully understood but it had been shown that ECA together with O-specific polysaccharide of endotoxin cause resistance to acetic acid and bile salts (Barua *et al*., 2002).
Figure 1. Diagrammatic representation of the Gram-negative bacterial cell wall.

The cell wall of Gram-negative bacteria is characterized by the presence of two lipid bilayers, the outer membrane and inner membrane. The polysaccharide coat on the outer membrane surface formed capsule including the lipopolysaccharide and the capsular polysaccharide. Many different proteins are localized in the outer and inner membrane.

1.2 Functions of Bacterial Capsule

It is clear that the expression of a capsule is an important virulence factor in encapsulated strains and provides a selective advantage to the bacteria in a variety of environments. Possible functions for a capsule include adhesion, transmission, resistance to innate host defenses, resistance to the host’s adaptive immune response, and intracellular survival.

1.2.1 Prevention of Desiccation

The highly hydrated bacterial capsule (water content > 95%) that coats the cell surface may protect the bacteria from desiccation (Roberson & Firestone, 1992). The prevention of desiccation may be associated with the expression of colanic acid. It was shown that a mucoid strain, such as *E. coli* K12, was significantly more resistant to desiccation than corresponding isogenic non-mucoid mutant (Ophir & Gutnick, 1994). It was speculated that the colanic acid may play a role in maintaining an appropriately humid environment surrounding the cell surface.
Exposure to desiccation can induce the expression of genes encoding enzymes for colanic acid synthesis in *E. coli* (Ophir & Gutnick, 1994). In addition, it was reported that the NeuO-mediated O-acetylation of capsular polysaccharide could enhance desiccation resistance in *E. coli* K1 strain (Mordhorst *et al*., 2009). The expression of O-antigen capsule in *Salmonella* is also important for survival during desiccation stress (Gibson *et al*., 2006). However, the mechanism of the capsule expression regulation in response to desiccation is unclear. It was hypothesized that desiccation could change the external osmolality, which triggers increased capsule expression (Roberts, 1996).

1.2.2 Adherence and Biofilm Formation

Capsular polysaccharides play a key role in adherence of bacteria to each other and to surfaces, thereby forming biofilms. By comparing various clinical isolates of *Haemophilus influenzae*, it was found that a capsulated strain could promote biofilm formation (Qin *et al*., 2014). Biofilm formation can confer nutritional advantages upon bacteria and consists of four sequential steps: initial attachment, micro-colony formation, extracellular polysaccharide production and maturation of the biofilm (Davey & O’toole, 2000). Initially, the specific colonizing bacteria may provide bridges for the subsequent attachment of other bacterial cells (Aparna & Yadav, 2008). Biofilm formation can significantly increase bacterial resistance to antibiotics and innate host defenses (Costerton *et al*., 1999; Aparna & Yadav, 2008). Bacteria in a biofilm can reach levels of resistance approximately 10 - 1000 times higher than during planktonic growth (Costerton *et al*., 1999). This may be due to the predominantly polysaccharide-rich biofilm matrix that can decrease the penetration of antimicrobial agents to the constituent cells (Sutherland, 2001). Resistance may also result from biofilm heterogeneity. It was hypothesized that a small subpopulation of cells in a biofilm may survive increased concentrations of one antibacterial substance due to a specific physiological status. The surviving cells can prevent the colony from being erased entirely, even though most of the cells were killed (Keren *et al*., 2004).
1.2.3 Against the Host Immune System

In the absence of specific antibody, the presence of a capsule is thought to mediate resistance to non-specific host defense mechanisms, especially the action of complement-mediated killing and phagocytosis. Capsule can provide a steric barrier protecting the outer membrane from host defenses including deposition of complement factors (Buckles et al., 2009). Complement is a complex system that can be activated by antigen-antibody complex via the classical pathway, lectin pathway or alternative pathway (Sarma & Ward, 2011). Activation of any of these three pathways leads to the opsonization of the target with the complement component C3b and results in the formation of the membrane attack complex (MAC) (Miajlovic & Smith, 2014). Factor H is the major regulator of the alternative pathway of complement system and the second most abundant complement factor in plasma (Walport, 2001). Factor H can inhibit the alternative pathway C3 convertase and promote inactivation of C3b into iC3b (Sarma & Ward, 2011). Negatively charged sialic acid could enhance the binding of factor H to C3b on the cell surface and prevent amplification of the alternative pathway. For example, the capsular polysaccharides of Neisseria meningitidis that contains sialic acid could recruit factor H for avoidance of complement-mediated killing. However it has been shown that N. meningitidis affects immune responses using the surface lipoprotein (Factor H binding protein) instead of charged-carbohydrate chemistry to recruit the factor H (Schneider et al., 2009). What is more, it was hypothesized that inhibition of complement-mediated lysis maybe because the capsule masks the cell surface components critical for binding of C3b to the bacterium, thus preventing insertion of the Membrane Attack Complex (MAC) into the bacterial outer membrane (Lo et al., 2009). This complex forms transmembrane pores in the membranes of susceptible bacteria, and hence leading to bacterial cell death (Sarma & Ward, 2011). It has been shown that a long O-polysaccharide chain of the LPS of Salmonella typhimurium can prevent insertion of the forming MAC into the outer membrane (Joiner, 1985).
The classical complement pathway is initiated by the binding of C1 to the Fc portion of antibody–antigen complexes on the bacterial surface (Sarma & Ward, 2011). For some Gram negative bacteria, such as the *E. coli* and *Salmonella spp.*, the long O-polysaccharide chain of LPS could sterically hinder the binding of complement components to the bacteria membrane in terms of blocking the access of the C1q to the bacterial membrane and thus halt classical complement pathway activation at the C1 stage (Rautemaa & Meri, 1999). Some bacterial surface proteins may also act in concert to destroy the classical complement response. For example, Wooster *et al.* (2006) demonstrated that the outer membrane protein OmpA could bind to the C4b-binding protein (C4bp) and disrupt activation of the classical pathway in *E. coli* K1 during exponential growth. C4bp is an inhibitor of classical complement activation, which can bind to the C4b and stop it binding to C2a, thereby inhibiting the formation of C4b2a (C3 convertase) (Wooster *et al.*, 2006).

As a result of the structural similarities between certain bacterial capsular polysaccharides and the host own tissue molecules, some capsules are poorly immunogenic. Therefore, the expression of such capsules provides resistance to the host’s specific immune response. For instance, the *E. coli* K5 polysaccharide consists of a repeated disaccharide of D-glucuronic acid (GlcA) and N-acetyl glucosamine (GlcNAc) as the repeating unit -4)-β-GlcA-(1-4)-α-GlcNAc-(1- (Vann *et al.*, 1981). Its chemical structure is identical to the biosynthetic precursor of mammalian heparin sulphate (Navia *et al.*, 1983). Therefore, it is hypothesized that the host immune system struggles to detect this capsule because the polysaccharide coat had the features that are very similar to heparin sulphate, resulting in a significant increase in virulence of these strains. Eventually, this hypothesis was proven. It was shown that the K5 capsular polysaccharide confers virulence to *E. coli* K5 by acting as a 3D mimetic of host heparin sulphate, helping to evade detection by the mammalian immune system (Blundell *et al.*, 2009). It has been known for a long time that the expression of capsular polysaccharides
confer some resistance to complement-mediated killing system even though the mechanism is still not clear (Roberts, 1996).

Capsular polysaccharides may also have immunomodulatory activities and act as signalling molecules modulating the inflammatory response of epithelial cells to maximize colonization and promote bacteria survival in the host (Corbett & Roberts, 2008). The expression of K5 capsule of *E. coli* strain Nissle 1917 plays an important role in induction of interleukin-8 (IL-8) through the mitogen-activated protein (MAP) kinase pathway (Hafez *et al.*, 2010). In the case of *Staphylococcus aureus*, both capsular polysaccharide types 5 and 8 were capable of binding to epithelial cells and induce interleukin-8 (IL-8) production then inducing IL-8, IL-6, IL-1b, and tumor necrosis factor-alpha (TNFα) from monocytes (Soell *et al.*, 1995).

In *S. aureus*, it was proposed that these capsular polysaccharides were binding to a protein on the cell membrane and acting as adhesins to promote immunomodulatory effects on human cells (Soell *et al.*, 1995). Evrard *et al.* (2010) also demonstrated that *Klebsiella pneumonia* capsular polysaccharides could induce a defective immunological host response involved in Th1 cytokine production and maturation of dendritic cells (Evrard *et al.*, 2010).

### 1.3 The Genetics and Biosynthesis of *Escherichia coli* Capsular Polysaccharides

The 80 different capsular serotypes in *E. coli* were originally divided into groups based on serological properties, and later revisions incorporated genetic and biochemical criteria (Table 1.1) (Whitfield & Roberts, 1999). Now the classification has been expanded to four groups (Table 1.1). The capsule gene cluster of *E. coli* have been cloned and studied which indicates that the *E. coli* capsule gene are clustered at a single chromosomal locus (Whitfield, 2006). The capsule single gene cluster allows the coordinate regulation of a large number of genes that may be involved in the biosynthesis and export of capsular polysaccharides (Figure 1.3). *E. coli* group 1 and 4 capsules share a common assembly system, and this is
fundamentally different from the one used for group 2 and 3 capsules (Whitfield, 2006).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>1</th>
<th>2</th>
<th>3 or 3</th>
<th>4 and 1 (O-antigen capsules)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Former K-antigen group</td>
<td>IA</td>
<td>II</td>
<td>IA/II</td>
<td>IB</td>
</tr>
<tr>
<td>Thermostability of K antigen</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Coexpressed with O serogroups</td>
<td>Limited range (O8, O9, O20, O101)</td>
<td>Many</td>
<td>Many</td>
<td>Often O8, O9 but occasionally none</td>
</tr>
<tr>
<td>Coexpressed with colanic acid</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Genetic locus</td>
<td>gap near bsl</td>
<td>kps near ser A</td>
<td>kps near ser A</td>
<td>Ksl</td>
</tr>
<tr>
<td>Thermoregulated expression</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Elevated levels of CMP-Kdo synthetase</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Terminal lipid motif</td>
<td>LPS lipid A core in K_{LPS}; unknown for capsular K antigen</td>
<td>α-glycerophosphate</td>
<td>α-glycerophosphate?</td>
<td>LPS lipid A core in K_{LPS}; unknown for capsular K antigen</td>
</tr>
<tr>
<td>Polymer chain grows at</td>
<td>Reducing terminus</td>
<td>Nonreducing terminus</td>
<td>Nonreducing terminus?</td>
<td>Reducing terminus</td>
</tr>
<tr>
<td>Polymerization system</td>
<td>Wy dependent</td>
<td>Processive glycosyltransferase activity</td>
<td>Processive glycosyltransferase activity?</td>
<td>Wy dependent</td>
</tr>
<tr>
<td>PST-1 protein</td>
<td>Wzx</td>
<td>None</td>
<td>None</td>
<td>Wzx</td>
</tr>
<tr>
<td>ABC transporter</td>
<td>None</td>
<td>KpsMT</td>
<td>KpsMT?</td>
<td>None</td>
</tr>
<tr>
<td>MPA-1 protein</td>
<td>Wzx</td>
<td>None</td>
<td>None</td>
<td>Wzx</td>
</tr>
<tr>
<td>MPA-2 protein</td>
<td>None</td>
<td>KpsE</td>
<td>KpsE?</td>
<td>None</td>
</tr>
<tr>
<td>OMA protein</td>
<td>Wza</td>
<td>KpsD</td>
<td>KpsD?</td>
<td>Wza</td>
</tr>
<tr>
<td>Model systems</td>
<td>Serotypes K30</td>
<td>Serotypes K1, K5</td>
<td>Serotypes K10, K54</td>
<td>Serotypes K40, O111</td>
</tr>
<tr>
<td>Similar to capsules in</td>
<td>Klebsiella, Erwinia</td>
<td>Neisseria, Haemophilus</td>
<td>Neisseria, Haemophilus</td>
<td>None known</td>
</tr>
</tbody>
</table>

Table 1. 1 Classification of *Escherichia coli* capsules (Whitfield, 2006)

1.3.1 *E. coli* Group 1 Capsules

*E. coli* Group 1 capsules are acidic polysaccharides, typically containing hexuronic acid, and usually co-expressed with a limited range of O antigens (O8, O9, O20, O101) (Jann & Jann, 1992). The K antigen from Group 1 is expressed on the cell surface in two different forms. One form comprises low-molecular-weight K-antigenic polysaccharides and is linked to a LPS lipid A core which is termed K_{LPS} (Dodgson *et al.*, 1996) to distinguish it from LPS molecules carrying the serological O antigen in the same isolate (MacLachlan *et al.*, 1993). The second form is the high-molecular weight capsular K antigen that forms the noticeable capsule
structure in electron micrographs and masks O antigen in serotyping (MacLachlan et al., 1993; Whitfield, 2006). The genetic locus of group 1 is a 16 kb region of DNA encoding 12 ORFs mapping to a chromosomal region near his (histidine-biosynthesis) operon (Roberts, 1996). Group 1 capsule gene cluster have a highly conserved genetic organization and are characterized by the presence of four genes: wzi, wza, wzb and wzc (Rahn et al., 1999; Drummelsmith & Whitfield, 1999).

The group 1 capsule biosynthesis locus (cps) comprises two regions separated by a putative stem-loop transcriptional attenuator (Rahn & Whitfield, 2003). The 5′ part of the locus includes four conserved genes (wzi, wza, wzb, and wzc) present in all group 1 cps loci (Figure 1.3). The 3′ region of the locus is serotype specific and encodes enzymes for a Wzy-dependent biosynthesis system. This 3′ region contains the enzymes for producing any sugar nucleotide precursors used for capsule synthesis, glycosyltransferases (GTs), and two integral inner membrane proteins (Wzy and Wzx) (Whitfield, 2006).

The defining characteristic of the group 1 (Wzy-dependent) pathway is that the individual repeat units are assembled on a carrier lipid (undecaprenyl phosphate; und-PP) by the sequential activities of glycosyltransferases (GTs) enzymes, WbaP. WbaP initiates the reaction in the cytoplasm, transferring galactose from UDP-galactose to the carrier und-PP (Roberts, 1996; Drummelsmith & Whitfield, 1999). A further GTs, WbaZ, then completes the formation of the repeating unit backbone, -2)-α-Man-(1-3)-β-Gal-1(1-. A side-branch is also present, formed from repeating glucuronic acid and galactose residues by the GTs - WcaN, which is linked to the main polysaccharide chain by WcaO (Drummelsmith & Whitfield, 1999). The lipid-linked repeat units are then flipped across the plasma membrane by the Wzx protein, using a mechanism that has yet to be established (Willis & Whitfield, 2013). Polymerization occurs at the periplasmic face of the plasma membrane with the help of polymerase enzyme, Wzy, and also requires the activity of the tetrameric Wzc protein (Collins et al., 2006). Translocation of the finished polymer involves the products of the genes wza, wzb and wzc. Wza is an integral outer
membrane lipoprotein (Willis & Whitfield, 2013). The Wza octamer spans the outer membrane and is comprised of four novel domains forming a large central cavity for translocating capsular polysaccharide (Dong et al., 2006; Nickerson et al., 2014). Wzc is an inner membrane protein with tyrosine auto-kinase activity and the cytosolic Wzb is its cognate phosphatase (Wugeditsch et al., 2001). Wzb is a responsible for dephosphorylating Wzc. It is possible that Wzb serves to dephosphorylate the spent und-PP carrier, to allow its re-entry into the polymerization cycle (Whitfield & Roberts, 1999). The cycling phosphorylation of Wzc is crucial for export (Wugeditsch et al., 2001; Paiment et al., 2002). Wzi is required for the efficient assembly of the capsular layer (Rahn & Whitfield, 2003). The current model of biosynthesis and export of the group 1 capsule is shown in figure 1.2.

![Figure 1. 2 A model for biosynthesis and assembly of group 1 and 4 capsules (Whitfield, 2006).](image)

Firstly, und-PP-linked repeat units are synthesized at the interface between the cytoplasm and the inner membrane. Then newly synthesized und-PP-linked repeats are then exported across the membrane in a process requiring Wzx. This provides the substrates for Wzy-dependent polymerization. Then the repeat units are polymerized in a Wzy-dependent reaction. High-level polymerization requires transphosphorylation of C-terminal tyrosine residues in the Wzc oligomer and dephosphorylation by the Wzb phosphatase. Exporting of polymer to the surface requires Wza, which likely acts as a channel. Wzi is unique to group 1 capsule and appears to be required for efficient assembly of the capsules on the cell surface.
1.3.2 *E. coli* Group 4 Capsules

Similar to Group 1 K antigens, Group 4 antigens are found on the cell surface in two different forms. The first is a high molecular weight capsular form that masks the underlying shorter O polysaccharide molecules in agglutination reactions and the second form is the K\text{\textsubscript{LPS}}, consisting of K oligosaccharides covalently attached to the outer membrane via the lipid-A core of LPS (Roberts, 2000). Members of this group are alternatively known as O-antigen capsules as they are very similar to LPS, with up to 50% of the capsule polymer expressed by these strains linked to lipid A-core (Dodge\textit{son et al.}, 1996; Roberts, 1996). Some of the differences between Group 1 and 4 K antigens include the presence of amino sugars or amino acids as a component of the repeat structure in Group 4 capsules but not in the Group 1 counterparts (Jann & Jann, 1992). What is more, although both groups form K\text{\textsubscript{LPS}}, Group 1 K\text{\textsubscript{LPS}} consists primarily of a single repeat unit of oligosaccharides that attached to lipid-A core while that of Group 4 is synthesized as longer polysaccharides chains attached to lipid-A (Roberts, 2000). Additionally, unlike strains bearing Group 1 capsules, the colanic acid locus of Group 4 strains is intact, and colanic acid can be co-expressed with the capsule polysaccharide (Whitfield, 2006).

The Group 4 capsule serotypes, K40 (Amor & Whitfield, 1997) (Figure 1.3) and O111 (Wang \textit{et al.}, 1998), are also assembled by Wzy-dependent pathways. Biosynthesis of the Group 4 capsular antigen occurs in the same manner as described for Group 1 capsules except that the initiating glycosyltransferase enzyme is WecA, transferring GlcNAc or N-acetyl galactose (GalNAc) (Whitfield & Roberts, 1999). There is an enzyme called Wzz (chain length regulator), regulating the extent of individual O-antigen repeats (Whitfield, 1995). Despite organizational similarities to Group 1 capsules, the genes necessary for the assembly of the Group 4 capsule are not known. Peleg \textit{et al.} (2005) suggest that etp and etk are required for the assembly of Group 4 capsule (Peleg \textit{et al.}, 2005). Dedicated Wza-Wzb-Wzc homologues are not encoded by gene cluster involved in group 4 K antigen
expression, leaving the identity of their translocation components unclear (Whitfield & Roberts, 1999).
Figure 1. 3 Genetic organization of *E. coli* capsule gene cluster in groups 1(K30), 2(K5), 3(K10) and 4(K40).

The large boxes represent the defined functional regions. Horizontal arrows show the transcription direction.
Figure 1. Genetic Organization of the *E. coli* K5 Capsule Gene Cluster.

The central serotype-specific region 2 is flanked by the two conserved regions 1 and 3. The two convergent arrows represent region 1 and 3 promoters (PR1 and PR3). Horizontal arrows show the transcription direction from both promoters. The promoter 1 could initiate a single major transcript from region 1, and the major transcript originating from region 3 that proceeds through region 2. The lower half of the figure details the region 1 promoter with the regulatory proteins binding sites, which the IHF binding site located at +130 indicated by green, SlyA indicated by red and H-NS indicated by blue. SlyA binding site I and H-NS binding site I are overlapping range from -224 to -134. SlyA binding site II and H-NS binding site are also overlapping range from -121 to -79. H-NS binding site III located from +1 to +32. SlyA and H-NS binding sites were mapped by DNaseI footprinting by Corbett *et al.* (2007). Numbering indicates nucleotide position relative to the transcription start site.
1.3.3 *E. coli* Group 3 Capsules

The group 3 typified by the K10 and K54 antigens (Orskov & Nyman, 1974). Group 2 and Group 3 capsules have similar heat stability, composition and charge density (Whitfield, 2006). The group 3 *kps* genes show a segmental gene organization, with two conserved regions flanking a central serotype-specific region in a manner analogous to that of group 2 capsule gene clusters (Pearce & Roberts, 1995) (Figure 1.3). However, the group 3 capsule genes appear to have little detectable nucleotide sequence in common with the Group 2 capsule genes, which indicates that these capsule gene clusters may have originated from a different source than that of the Group 2 capsule genes, but inserted at the same *serA* site on the *E. coli* chromosome (Pearce & Roberts, 1995). The Group 3 capsule cluster region 1 contains four genes (*kpsD, kpsM, kpsT, kpsE*) which are responsible for encoding region 1 and region 3 proteins homologues of the Group 2 (Roberts, 1996). Region 3 is composed of two genes, *kpsC* and *kpsS*, which encode homologues of the Group 2 region 1 proteins (Russo et al., 1998).

1.3.4 *E. coli* Group 2 Capsules

Group 2 capsule are expressed by many extra-intestinal isolated of *E. coli* and closely resemble capsules from *N. meningitidis* and *H. influenzae* (Frosch et al., 1991). Capsular polysaccharides of this group are linked via their reducing termini to phosphatidyl-Kdo acceptor, which play a role in anchoring the capsular polysaccharide to the cell surface (Jann & Jann, 1992). The Kdo (2-keto-3deoxy-manno-octonic acid) works as a reducing sugar linking the reducing end of the polysaccharide chain and outer membrane lipid (Finke et al., 1991). Group 2 capsules are unique among other *E. coli* capsule groups because they are temperature-regulated at the level of transcription with capsule gene being expressed at 37 °C, but not below 20 °C (Simpson et al., 1996).
1.3.4.1 The Genetic Organization and expression of Group 2 Capsule Gene Cluster

Group 2 capsule gene clusters are located near the serA locus on the E. coli chromosome and organized into three regions (Bounois et al., 1987). Capsule gene clusters of E. coli K1 (Silver et al., 1981; Vimr et al., 1989), K4 (Drake et al., 1990), K5, K7, K12 and K92 (Roberts et al., 1986) have been cloned and analysed which reveal that the Group 2 capsular polysaccharides have a conserved genetic organization consisting of three functional regions (Bounois et al., 1987). The serotype-specific Region 2 is flanked by two conserved regions: region 1 and region 3. Region 1 and 3 contain the genes responsible for transport of newly synthesized capsular polysaccharides from the cytoplasm to the bacterial cell surface, and are conserved in all Group 2 capsule gene clusters. Region 2 encodes the genes responsible for the synthesis of the polysaccharide and its precursors (Roberts et al., 1988). The transcription organization of Group 2 capsule gene cluster is of two convergent transcripts, one of which originates from the region 1 promoter and covers region 1, and the other originates from the region 3 promoter and span region 1 and 3 (Rowe et al., 2000). Both promoters are temperature regulated at the level of transcription with genes being expressed at 37 °C but not at 20 °C (Cieslewicz & Vimr, 1996; Simpson et al., 1996).

Region 1 of the K5 locus contains 6 genes, kpsFEDUCS, organized in a single transcriptional unit. A single σ70 promoter binding site (promoter 1, PR1) is located 225 base pairs (bp) upstream from the initiation codon for the first gene kpsF (Simpson et al., 1996) (Figure 1.4). Analysis revealed that no alternative sigma factor was found within this promoter (Rowe et al., 2000). Transcription from the region 1 promoter generates an 8.0 kb polycistrionic transcript and processed to give a separately stable 1.3 kb kpsS transcript (Simpson et al., 1996). This processing has been proposed to allow the differential expression of the kpsS gene, but the mechanism is unclear (Roberts, 2000). An intragenic Rho-dependent transcriptional terminator was discovered within the kpsF gene (Simpson et al.,
This intragenic terminator is believed to play a role in reducing unnecessary transcription suffering physiological stress (Corbett & Roberts, 2008).

Region 3 contains two genes kpsM and kpsT that are organized in as a single transcriptional unit with a $\sigma^{70}$ promoter (Promoter 3, PR3) (Pavelka et al., 1991). PR3 can initiate transcription from a site 741 bp upstream of the start of kpsM (Petit et al., 1995). Promoter 3 has typical E. coli $\sigma^{70}$ -10 consensus sequence but no -35 sequence and there are no consensus binding sequences for other $\sigma$ factors nor IHF present within this promoter (Stevens et al., 1997). An ops (operon polarity suppressor) sequence, a cis-acting regulatory sequence conserved in RfaH regulated operons is located 30 bp upstream of the start codon of kpsM (Stevens et al., 1997). The role of RfaH in regulation of Group 2 capsule gene expression in E. coli was confirmed by observing that mutation of rfaH or deletion of the ops sequence resulted in a reduced capsule expression (Stevens et al., 1997). It was shown that RfaH increases the transcription of region 2 genes by read through transcription from the Region 3 promoter (Stevens et al., 1997). RfaH acts as an antiterminator and this read through transcription is essential to obtain expression levels of kfi genes necessary to synthesize K5 capsule. The ops sequence is also contained within a 39 bp regulatory element called JUMPstart (Just Upstream of Many Polysaccharide-associated gene starts) (Hobbs & Reeves, 1994). JUMPstart sequences are present in many polysaccharide biosynthesis genes in enteric bacteria including O antigen gene clusters and Group 2 capsule gene clusters in E. coli (Hobbs & Reeves, 1994). RfaH may interact with RNA polymerase complex to permit transcription elongation to proceed from region 3 to region 2. It is hypothesized that RfaH is recruited to the transcription elongation complex at the ops site and is able to affect transcription up to 20 kb downstream (Bailey et al., 1997; Bailey et al., 2000).

In the K5 gene cluster, the region 2 gene cluster is 8.0 kb and comprises 4 genes kfiABCD, all of which are transcribed in the same direction as region 3 (Petit et al., 1995). In the K1 gene cluster, region 2 consists of six genes, neuDBACES (Silver et
The mRNA generated from *E. coli* K5 region 2 promoter surprisingly has two large untranslated intergenic regions: one is 340 bp between *kfiA* and *kfiB* genes and the other is 1293 bp between *kfiB* and *kfiC* genes (Petit *et al.*, 1995). The role of the untranslated regions (UTR) from region 2 is still unclear. Three weak promoters in region 2 were identified 5’ to the *kfiA*, *kfiB* and *kfiC* genes, generating 8, 6 and 3.5 kb transcripts respectively (Petit *et al.*, 1995). The transcription from the *kfiA* promoter generates 8.0 kb transcripts spans the entire region 2, including the two large intergenic regions between the *kfiA* and *B* gene and *kfiB* and *C* genes (Petit *et al.*, 1995). The transcription from these three weak promoters is independent of temperature with equivalent transcription at both 37 °C and 18 °C (Roberts, 1996). It was believed that the weak promoters generate insufficient transcription to yield detectable levels of Kfi proteins and expression of a K5 capsule (Stevens *et al.*, 1997). Transcription of region 2 is in the same direction as that of region 3 and the vast majority of region 2 message is produced by read through transcription from the PR3, generating a polycistronic transcript containing both region 3 and 2 genes (Stevens *et al.*, 1997). It was suggested that the region 2 gene maybe acquired from other bacteria capsule gene sequence by the homologous recombination events between the region 1 and 3 of incoming and resident capsule DNA clusters (Roberts, 1996).

### 1.3.4.2 The Biosynthesis of Group 2 Capsule Expression

The biosynthesis of the *E. coli* K1 and K5 antigens have been extensively studied (Finke *et al.*, 1991; Jann & Jann, 1992; Steenbergen *et al.*, 1992). The biosynthesis of Group 2 capsules occurs on the inner surface of the cytoplasmic membrane (Roberts, 1996). The initiation of polymer synthesis occurs by the attachment of sugar monomers to an unknown endogenous acceptor in the membrane (Finke *et al.*, 1991). It was reported that the capsular polysaccharide chain could attach to phospholipid, lyso-phosphatidyglycerol (PG), via a novel β-link-poly-3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) linker (Willis & Whitfield, 2013). The conserved terminal lipid has been suggested to be recognized by the ABC transporter and
direct the transport anchor of K antigen at the cell surface (Jann & Jann, 1992). Polymerization then proceeds at the non-reducing terminus by addition of activated sugar molecules to the growing polysaccharide chain (Troy, 1992). The K1 capsule is a linear homopolymer of an (α-2,8)-linked sialic acid (NeuNAc) while the \textit{E. coli} K5 capsular polysaccharide consists of a repeat structure of glucuronic acid (GlcA) and N-acetylglucosamine (GlcNAc) as [-4]GlcA-β(1,4)-GlcNAc-α(1-) (Jann & Jann, 1992). In the \textit{E. coli} K1 serotype, the polysialic acid capsule (NeuNAc) is synthesized from activated UDP-NeuNAc acid produced from N-acetylmannosamine (ManNAc) and phosphoenolpyruvate (Bliss & Silver, 1996). The polymers in \textit{E. coli} K1 are synthesized by the enzyme polysialyltransferase (PST) while the \textit{E. coli} K5 polymer is synthesized by glycosyltransferases (Willis & Whitfield, 2013). The biosynthesis and export of Group 2 serotype K5 capsules involves a heterooligomeric membrane-bound complex on the plasma membrane (Rigg \textit{et al.}, 1998), consisting of the proteins KfiA-D, KpsC, M, S and T (Petit \textit{et al.}, 1995; Bliss & Silver, 1996; Roberts, 1996). The synthesis of the K5 capsule involves the addition of alternating GlcA and GlcNAc residues from their nucleoside-sugar precursors to the non-reducing end of the polysaccharide (Corbett & Roberts, 2008).

Mutations within \textit{E. coli} K5 region 2 \textit{kfi} genes abolished polysaccharide biosynthesis, which suggest that this region encodes enzymes for the synthesis of this particular K antigen (Roberts \textit{et al.}, 1988). The \textit{kfi} cluster encodes the proteins required for the synthesis, activation and polymerization of the sugar subunits into a large 200-residue polymer that terminates in a phosphodiester bound to 1, 2 - diacylglycerol (Silver \textit{et al.}, 2001). The functions of three of the four \textit{Kfi} gene products of region 2 have been determined for K5 antigen. The KfiA protein is an α-UDP-GlcNAc glycosyltransferase (GT) that can initiate K5 polysaccharide biosynthesis. KfiA can catalyze the addition of GlcNAc to the non-reducing terminus of a nascent K5 polysaccharide chain via a α - 1, 4 - glycosidic bond (Hodson \textit{et al.}, 2000; Roman \textit{et al.}, 2003). KfiC is the β-UDP-GlcA glycosyltransferase, which acts to add UDP-GlcA residues to the non-reducing end.
of the growing polysaccharide chain (Griffiths et al., 1998). KfiD is UDP-Glc dehydrogenase, which catalyzes the formation of essential substrate UDP-GlcA from UDP-glucose (Roman et al., 2003). No role has been established for the KfiB protein, but it may be a structural one for supporting the biosynthetic complex (Roberts, 2000). It was suggested that the interaction between the KfiA, KfiB and KfiC proteins is essential for the stable association of these proteins with the cytoplasmic membrane and the biosynthesis of K5 polysaccharide (Hodson et al., 2000).

1.3.4.3 Export and Translocation of Group 2 Capsular Polysaccharide

Mutations of individual kps genes revealed that the biosynthesis of Group 2 capsule is intimately linked to its export via a biosynthetic-export complex on the plasma membrane (Whitfield & Roberts, 1999). This complex consists of the proteins (KfiA-D) required for polymerization and KpsC, M, S and T necessary for polysaccharide transport across the cytoplasmic membrane (Rigg et al., 1998; Whitfield & Roberts, 1999). The conservation of Kps proteins among all E. coli Group 2 capsules would suggest that these proteins might provide the scaffold onto which the specific capsule biosynthetic proteins can function. For example, KpsM and KpsT help the polysaccharide transfer across the cytoplasmic membrane, whereas KpsD and KpsE are responsible for transport across the periplasm (Corbett & Roberts, 2008).

The nascent capsule is exported across the inner membrane by a member of the ABC-2 family of ATP-binding cassette transporters (Bliss & Silver, 1996). The ABC-2 transporter of Group 2 capsule comprises KpsM (the transmembrane component) and KpsT (the ATPase component) (Willis & Whitfield, 2013). KpsT is the ATP binding (ABC) module and contains the ABC ATPase domain that couples ATP hydrolysis to transport (Willis & Whitfield, 2013). KpsM provides the hydrophobic membrane-spanning (MS) module and contains six membrane-spanning regions, with the N- and C-terminal domains facing the cytoplasm (Whitfield & Roberts, 1999). KpsF is responsible for catalysing the conversion of ribulose-5-phosphate to
arabinose-5-phosphate, an intermediate in Kdo synthesis (Meredith & Woodard, 2006). It was suggested that KpsU is responsible for formation of CMP-Kdo, for attachment of Kdo to phospholipid (Rosenow et al., 1995). The KpsC and KpsS proteins maybe responsible for the attachment of phosphatidyl-Kdo to the reducing terminus of the capsular polysaccharide immediately before export (Arrecubieta et al., 2001). KpsE is a 43 kDa protein that is anchored in the cytoplasmic membrane with the bulk of the protein existing in the periplasmic space (Roberts, 1996; Whitfield & Roberts, 1999). It may play a role in an analogous manner to the membrane export proteins associated with ABC exporters and interact with the outer membrane via its periplasmic domain (Willis & Whitfield, 2013). Therefore, KpsE could be important for the formation of the membrane adhesion sites associated with the capsule assembly complex (Whitfield & Roberts, 1999). KpsD is a 60 kDa outer membrane protein that is also required for translocation of capsular polysaccharides in E. coli (Bliss & Silver, 1996). The function of KpsD in polymer export is still unclear. One possibility is that KpsD might be involved in recruiting porins to the capsule assembly complex to permit the transport of polysaccharide across the outer membrane (Roberts, 1996). The current model for the Group 2 biosynthesis and export are shown in figure 1.5.
Figure 1.5 A model for biosynthesis and assembly of group 2 capsules (Whitfield, 2006).

Firstly, polymer formation is initiated on an unknown endogenous acceptor (open hexagon) and is elongated by processive glycosyltransferases (GTs), adding residues to the non-reducing terminus of the chain. The growing polymer is shown on an undecaprenyl pyrophosphate carrier before being export to the phosphatidyl-Kdo. The polymer is exported via the ABC transporter (KpsM and KpsT). KpsS and C play an essential role in this process and are responsible for the attachment of phosphatidyl-Kdo to the reducing terminus of the capsule, and KpsF and KpsU also participate, but the details are unknown. The orientation of the polymeric substrate during export is still unknown, and biosynthesis and export may be temporally coupled. Translocation across the periplasm and outer membrane requires KpsE and KpsD.
1.4 Mechanism of Bacterial Promoter Transcription

1.4.1 Initiation of Transcription

Transcription initiation is an essential step for the regulation of gene expression and RNA polymerase (RNAP) plays an important role. Transcription initiation is a multi-step process by which RNA is synthesised from a DNA template catalyzed by RNAP. Binding of RNAP to DNA triggers a series of conformational changes in both RNAP and template DNA, formation from the closed double helix DNA of a highly stable open complex.

1.4.1.1 RNA Polymerase Binding Site: Promoter

Initiation of the transcription is dependent on different DNA sequence elements, which are called promotors, defined by a conserved DNA sequence. Four different sequence elements have been identified. For RNA polymerase containing $\sigma^{70}$, there are two core promoter elements called -10 hexamer ‘TATAAT’ (also known as Pribnow box) and the -35 hexamers ‘TTGACA’, which are located 10 bp and 35 bp upstream from the transcription start site, respectively (Harley & Reynolds, 1987; Lisser & Margalit, 1993). The spacing (optimally 17±1 bp) separating the -10 and -35 elements is ensures communication of these two promoter elements which is important in positioning the two consensus hexamers at the correct distance and angle for optimal contacts with the surface of the RNA polymerase (Harley & Reynolds, 1987). It has been shown that the spacer plays an active role in integrating the functional consequences of RNA polymerase contacts with -10 and -35 promoter element (Sztiller-Sikorska et al., 2011). Also, the spacer can interact with the $\beta'$ zipper of RNAP to facilitate the formation of stable closed promoter complex and can even substitute for $\sigma$4 interactions with the -35 element during the open complex formation (Yuzenkova et al., 2011). The other two important promoter elements are the extended -10 ‘TG’ element and the UP element. The extended -10 ‘TG’ element is a 3-4 bp motif located just upstream of the -10 hexamer (positioned at -15/-14 with respect to the transcription start site), which is recognized by the $\sigma$3.0 (or the domain 2.5 of $\sigma$ factor)(Barne et al., 1997)
The TG motif is present in ~20% of *E. coli* promoters and provides a supplementary recognition sequence which can substitute for the lack of the -35 region (Burr *et al.*, 2000; Murakami *et al.*, 2002; Saecker *et al.*, 2011). The UP element is a ~20 bp sequence located upstream of -35 promoter element that is recognized by the αCTD (the C-terminal domain of RNAP α-subunit) (Ross *et al.*, 2001). The UP element was found upstream of many strong promoters and used to derive a consensus sequence: -59 nnAAA(A/T)(A/T)T(A/T)TTTnnAAAAnnn -38 (Estrem *et al.*, 1998) (Figure 1.6). It functions by increasing the recruitment of RNAP such that promoter strength can be increased up to 30-fold (Ross *et al.*, 1993). The interactions between the αCTDs and the UP element (specifically or non-specifically) can stimulate and accelerate the process of promoter complex formation (Strainic *et al.*, 1998).

### 1.4.1.2 RNA Polymerase

There are five subunits in the bacterial RNAP core enzyme (E): two identical α subunits and one subunit of each of β, β’ and ω (Burgess, 1969; Vassylyev *et al.*, 2002) (Figure 1.7). The core enzyme is capable of nonspecific DNA recognition and subsequent transcription initiation. In bacteria, one specific σ factor combines with the core RNAP (E) to form the holoenzyme (Eσ) to initiate specific transcription (Murakami & Darst, 2003).

The α subunit, the product of the *rpoA* gene, is a 36.5 kDa protein, with three domains: two independently folded domains connected by a flexible interdomain linker (Blatter *et al.*, 1994; Iyer *et al.*, 2003). The α subunit does not have direct catalytic activity but is still required for enzyme assembly, transcription initiation and elongation steps (Gross *et al.*, 1996; Iyer & Aravind, 2012). The αCTD, specifically the C-terminal helix-Hairpin-helix (HhH) domain, can contact the minor groove of the UP element, the HhH domain of σ factor and also make contact with the transcription factors binding upstream of promoter, while the N-terminal domain is responsible for dimerization and interactions with β and β’ subunits (Kimura & Ishihama, 1996; Ross *et al.*, 2001; Hudson *et al.*, 2009; Iyer & Aravind,
At some promoters, the promoter-proximal αCTD binds to the promoter DNA at position -41 and interacts directly with domain 4 of σ factor (Gourse et al., 2000). The interaction of α subunits with transcriptional regulators seems to be important for effective transcription initiation by RNAP (Iyer & Aravind, 2012).

The β subunit, the product of the rpoB gene, co-transcribing with the rpoC gene encoding the RNAP β’ subunit were identified (1342 and 1407 amino acids respectively) (Burgess, 1969; Vassylyev et al., 2002). The active site for the nucleotidyltransferase activity of the RNA polymerase is formed from residues from both the β and β’ subunits that together are termed the catalytic subunits (Opalka et al., 2010). These two large subunits are involved in the DNA binding, elongation as well as termination activities of the enzyme (Gross et al., 1996).

The ω subunit (91 amino acids) is the smallest RNAP subunits and its function remains elusive. It is an entirely α-helical protein and is asymmetrically positioned in the RNAP complex (Vassylyev et al., 2002). The ω subunit associates with the final step of recruitment of β’ to the subunits α₂β to form α₂ββ’. It is also involved in the stringent response by providing the binding site for (p)ppGpp (Ghosh et al., 2003; Mathew & Chatterji, 2006).

The first step of RNAP assemble pathway is the formation of a dimer of two α subunits, which acts as a scaffold for the addition of first β and then β’ω to give core enzyme with the help of ω subunit. The holoenzyme is then formed by the addition of the σ subunit (Ishihama et al., 1987).

A high-resolution structural study revealed that the core enzyme adopts a CRAB-CLAW STRUCTURES (Darst, 2001; Vassylyev et al., 2002; Browning & Busby, 2004). This ‘crab claw’ shape of core RNAP was about 115 Å tall, 110 Å wide and 150 Å long, with an internal channel of 27 Å diameter (Darst, 2001). On the back of the channel, the essential ion ‘Mg$^{2+}$’ is located close to the active site (Darst, 2001).
1.4.1.3 Sigma-Factors

The σ factor is a key determinant for promoter recognition. All sequence-specific contacts with RNAP core elements are mediated by the σ subunit (Murakami et al., 2002). There are seven species of σ subunits that have been identified in *E. coli*, these are σ^D^ (σ^70^), σ^N^ (σ^54^), σ^S^ (σ^38^), σ^H^ (σ^32^), σ^F^ (σ^28^), σ^E^ (σ^24^) and σ^FecI^, each participating in recognizing and transcribing specific sets of genes (Ishihama, 2000) (Table 1.2). All bacteria have a primary sigma factor that suffices for growth under nutrient-rich conditions. In *E. coli*, the primary sigma factor is σ^70^, reflecting its molecular mass of approximately 70 kDa (Saecker et al., 2011). Most bacterial sigma factors have significant homology to the σ^70^ and are collectively referred to as the σ^70^-family. The other six interchangeable alternative sigma factors (encoded by the *rpoS*, *rpoH*, *rpoE*, *rpoF*, *rpoN* and *fecI* genes in the *E. coli* genome) are concerned with the management of different stresses (Ishihama, 2000).

Most of the housekeeping genes expressed during exponential-phase are regulated by the σ^70^ while the σ^S^ is responsible for transcription of some specific genes that are important for survival at stationary-phase (Hengge-Aronis, 1996; Loewen et al., 1998). σ^S^ and σ^70^ can recognize the same promoter elements, so that many promoters can be recognized by both σ^70^ and σ^S^ (Gaal et al., 2001). Other alternative sigma factors are required for transcription of some stress response genes. Heat shock σ factors, σ^H^ and σ^E^, are required to transcribe genes for heat shock proteins that are essential for survival (Yura, Nagai, & Mori, 1993; Arsène, Tomoyasu, & Bukau, 2000). The holoenzyme Es^E^ is also responsible for transcribing genes that can help to refold and degrade thermally denatured protein, such as mis-folded proteins in the periplasm (Erickson & Gross, 1989; Rouvière et al., 1995). The holoenzyme Es^N^ is important for transcription of genes essential for the nitrogen metabolism and is activated in other stress conditions (Merrick, 1993a; Shingler, 1996). The σ^N^ family are very distantly related to the σ^70^-family, and except for the small σ^N^ family, all other sigma factors share common features (Merrick, 1993b). Es^F^ is involved in the synthesis of the proteins that
necessary for flagella formation and chemotaxis (Helmann, 1991). $\sigma^{\text{FecI}}$ acts as a transcriptional activator in the ferric citrate transport system in \textit{E. coli} and belongs to a subfamily of $\sigma^{70}$ factors that respond to extracytoplasmic stimuli (Angerer et al., 1995).

<table>
<thead>
<tr>
<th>Sigma factors</th>
<th>Upstream recognition sequence (-35 Region)</th>
<th>Optimal number of spacer nucleotides</th>
<th>Downstream recognition sequence (-10 Region)</th>
<th>Genes under the control of each sigma</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^D$ ($\sigma^{70}$)</td>
<td>TTGACA</td>
<td>17±1</td>
<td>TATAAT</td>
<td>Housekeeping genes; Growth-related genes (~1000)</td>
</tr>
<tr>
<td>(613 aa)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^N$ ($\sigma^{54}$)</td>
<td>ttGGcaca</td>
<td>4</td>
<td>ttGCA</td>
<td>Nitrogen-regulated/stress response genes (~15)</td>
</tr>
<tr>
<td>(477 aa)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^S$ ($\sigma^{38}$)</td>
<td>CCGGCG</td>
<td>17±1</td>
<td>CTATACT</td>
<td>Stationary phases/stress response genes (~100)</td>
</tr>
<tr>
<td>(330 aa)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^H$ ($\sigma^{32}$)</td>
<td>TntCNCCCTTG AA</td>
<td>13-17</td>
<td>CCCCATtTA</td>
<td>Heat shock/stress response genes (~40)</td>
</tr>
<tr>
<td>(284 aa)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^F$ ($\sigma^{28}$)</td>
<td>TAAA</td>
<td>15</td>
<td>GCCGATAA</td>
<td>Flagella-chemotaxis genes (~40)</td>
</tr>
<tr>
<td>(239 aa)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^E$ ($\sigma^{24}$)</td>
<td>GAACCTT</td>
<td>16</td>
<td>TCTGAT</td>
<td>Extreme heat shock/extracytoplasmic genes (~5)</td>
</tr>
<tr>
<td>(202 aa)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^{\text{FecI}}$</td>
<td>AAGGAAAT</td>
<td>17</td>
<td>TCCTTT</td>
<td>Ferric citrate transport/extracytoplasmic genes (~5)</td>
</tr>
<tr>
<td>(173 aa)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.2 Sigma factors of \textit{Escherichia coli} (Wagner, 2000).

$\sigma^D$ and $\sigma^S$ will recognize the same promoters. The RNA polymerase holoenzyme containing sigma subunit recognizes and transcribes a specific set of genes.

1.4.1.4 THE SIGMA$^{70}$ FACTOR

The primary structures of the $\sigma^{70}$-family has been divided into four regions, numbered 1-4 (with subregions 1.1, 1.2, 2.1, 2.2, 2.3, 2.4, 3.0, 3.1, 3.2, 4.1 and 4.2), which were mapped on the basis of their functional properties and sequence conservation (Murakami & Darst, 2003; Paget & Helmann, 2003). These functions include binding to the core RNA polymerase, recognition of the consensus -10 and
-35 promoter elements, DNA melting and interaction with certain transcriptional activators.

The σ1 domain of the σ^{70} unit is required for the early steps in transcription initiation complex formation (Murakami, 2013). The σ2 domain contains the most highly conserved amino acids. It was found that a single α helix near the C-terminal end of the region 2 (region 2.4) recognizes the -10 hexamer (Gross et al., 1996). σ^{70} subregions 2.3 together with 2.1 play an essential role in DNA melting which stabilizes the initial transcription bubble (deHaseth and Helmann, 1995). The σ3 domain of E. coli σ^{70} is involved in recognition of the extended -10 element (Barne et al., 1997), thereby stabilizing the open complex (Voskuil & Chambliss, 2002). The linker separating region 3 and 4, comprised of the σ3.2 region (the σ3.2 loop), loops into the RNAP active site channel and go around the way out through the RNA exit channel (Murakami & Darst, 2003) (Figure 1.7). The σ4 domain at the C-terminal of the σ^{70} subunit is C-shaped, which has been shown to participate in the recognition of the -35 promoter element. The helix-turn-helix motif (HTH motif) of σ4.2 binds to the -35 element and bends the DNA backbone 36° around the recognition helix (Murakami et al., 2002). The bend of DNA at -35 site by σ4 region can alter the track of upstream DNA which can bring the DNA closer to the RNAP and facilitating interactions between the α-CTD domain and the upstream DNA and interactions with transcription activators that bind at the UP-elements (Ross et al., 1993, Murakami et al., 2002).
Figure 1.6 Promoter recognition by $\sigma^{70}$ and $\alpha$ subunit.

Orange and blue arrows indicate recognition of promoter regions as double-stranded DNA elements by the $\alpha$ and $\sigma^{70}$ subunits, respectively. The two red arrows point a region of the non-template strand DNA recognized by $\sigma^{70}$ subsequent to strand separation. Region 1-4 are divided into sub-regions. The consensus promoter elements -10 and -35 regions are recognized by regions 2.4 and 4.2, respectively. Region 3.0 is shown making direct interaction with the TG motif located one base pair upstream of the -10 element (indicated by purple). A typical *E. coli* promoter does not have all elements shown and exhibits deviations from the consensus sequences shown here for the -10, -35, and UP element, as well as the consensus spacer length (17 bp). (Modified from Murakami & Darst, 2003)

![Promoter recognition by $\sigma^{70}$ and $\alpha$ subunit.](image)

Figure 1.7 Structure of complex between *E. coli* $\sigma^{70}$ RNAP holoenzyme and interaction with promoter and DNA.

The components of holoenzyme is shown as a molecular surface, with the $\alpha l$ in grey, all in dark grey, $\beta$ subunit in purple, $\beta'$ subunit in green, $\omega$ subunit in yellow and $\sigma^{70}$ subunit in pink. The DNA phosphate backbone is shown with the template strand in light grey and the non-template strain in dark grey. This figure was generated using software UCSF Chimera.

![Structure of complex between *E. coli* $\sigma^{70}$ RNAP holoenzyme and interaction with promoter and DNA.](image)
1.4.1.5 Steps of Transcription Initiation

There are three main steps involved in the transcription initiation: promoter recognition at specific consensus sequence; isomerization to form an open complex; the initial step of RNA-chain synthesis (Figure 1.8). Multiple events during initial transcription include σ release from the core enzyme, DNA scrunching and promoter escape have been proposed (Kapanidis et al., 2006; Revyakin et al., 2006; Tchernaenko et al., 2008).

In the first step of transcription initiation, the RNAP recognizes fork junction DNA containing conserved -10 and -35 promoter elements in a sequence specific manner (Guo & Gralla, 1998; Matlock & Heyduk, 2000) (Figure 1.8). Differences between promoters, such as the TG region and UP element are also specifically and selectively recognized by RNAP and complete the initial binding of RNAP (Harley & Reynolds, 1987; Ross et al., 1993; Blatter et al., 1994; Gross et al., 1996). The promoter DNA will lie across one face of the holoenzyme, completely sitting outside of the RNAP active-site channel (Figure 1.8).

The RNAP binds to the DNA to form a closed complex (RP$_C$) and then triggers a series of conformational changes in both RNAP and DNA template to form an open complex (RP$_O$), which is often called ‘isomerization’ (Murakami & Darst, 2003). The mechanistic details of isomerization are still unsolved, but probably include the RNAP introducing of a series of kinetically distinct intermediate states, bending the DNA around the RNAP to increase the available binding interface (Murakami et al., 2002). During the transition from RP$_C$ to unstable intermediate stage, a short DNA segment is melted to make the template strand accessible to the catalytic core (Gross et al., 1996). The DNA melting is associated with unwinding of the DNA and induces flexibility in the DNA, allowing the downstream DNA to bend across. Once the two strands are separated, with the transcription start site base (+1) of the template strand in the active site while the non-template strand is being hold in region σ2.3 (Gross et al., 1996; Murakami et al., 2002; Vassylyev et al., 2002; Saecker et al., 2011). In the RP$_O$ model, the upstream edge of the open complex
bubble begins at -11 followed by melting of the two DNA strands that extends 13/14 bp downstream to site +2/+3 around the transcription start site at +1 (Browning & Busby, 2004; Gries et al., 2010) (Figure 1.8). During the open complex formation, the -10 downstream double stranded DNA plays an important role to stabilize formation of the open complex (Murakami et al., 2002), and the interaction between the upstream DNA and RNAP is also essential for directing the downstream DNA moving into the RNAP active channel (Davis et al., 2007; Saecker et al., 2011). Once the open complex is formed, the RNAP starts catalyzing the synthesis of RNA with adding NTP substrate to form a short RNA-DNA hybrid at the catalytic channel (Davis et al., 2007; Saecker et al., 2011). The universally conserved positive charged amino acid exposed at the entrance of the tunnel in the holoenzyme, such as Arg and Lys, contacts the negatively charged phosphate backbone of DNA in directing the strand of template DNA template to move into the active site of RNAP so that the RNA synthesis can begin (Murakami et al., 2002; Browning & Busby, 2004). The newly synthesized 11nt RNA chain will displace the σ3.2 loop out of its path and make the end of abortive initiation, which plays an important role in the σ ejection (Basu et al., 2014). Then the initially transcribing complex will end the initiation-to-elongation transition and enter into the elongation phase of transcription (Saecker et al., 2011).
**Figure 1.8 RNA polymerase interactions at a promoter and the pathway of transcription initiation.**

(a) A schematic representation of RNA polymerase bound to a bacterial promoter and forms a closed complex (RPc). The transcription start point is indicated by a bent arrow. (b) The duplex DNA around -10 to +2 is unwound to form an open complex (RPo), which is called isomerization. (c) The initiating complex (RPinit) is formed and synthesis of the transcript (shown as red dash line) begins with formation of a phosphodiester bond between the initiating and adjacent phosphodiester nucleotide triphosphate (NTPs). (d) Elongation is the final step and the RNA chain length increases along with the RNAP running through the DNA template. (Modified from Browning & Busby, 2004)
1.4.2 Transcription Elongation

RNA synthesis involves covalent bond formation between the 3’ OH end of the nascent newly synthesized RNA and the α-phosphate of the incoming nucleotide triphosphate. Phosphodiester bond synthesis results in the extension of the chain by one residue and release of pyrophosphate (Saecker et al., 2011).

During transcription elongation, RNAP releases from the promoter and transitions into the elongation complex (EC). The EC moves along the template strand, adding ribonucleotides to the 3’ hydroxyl of the growing RNA transcript. Typically, σ⁷₀ is likely to be triggered to release for elongation when the nascent RNA-DNA complex reaches 8-9 nt with the DNA strand downstream of the transcription start site (Mooney et al., 2005). However, during the transition from initiation to elongation, an alternative mechanism has been proposed in which the σ⁷₀ factor is retained on the elongation ternary complex and translocates with the RNAP during elongation (Bar-Nahum & Nudler, 2001; Mukhopadhyay, 2001). Based on studies on the σ cycle paradigm, two versions of the σ cycle were concluded. One is called the obligate-release model in which the RNAP cannot form a stable elongation complex until the σ factor is released from the complex (normally occurring after synthesis of 8-9 nt of RNA transcript (Mooney et al., 2005). The second model is called the stochastic release model in which the affinity of σ factor decreases during elongation but release happens stochastically after RNAP initiated transcription (Mooney et al., 2005). The σ factor also associated with the RNAP elongation that cause pausing at promoter proximal sites by rebinding promoter -10-like DNA sequence downstream of the transcription start site to regulate transcription elongation (Mooney et al., 2005).

1.4.3 Regulation of Transcription Termination

Bacteria have two alternative ways to terminate the transcription reaction. Transcription may stop as a consequence of a special sequence signal that causes transcription to adopt a termination proficient structure leading to the RNAP to leave the DNA template. The majority of intrinsic rho-independent terminator
signals consist of stable GC-rich palindromic structures and followed by a poly-T tail in the nontemplate DNA (Yarnell & Roberts, 1999). The transcribed palindromic region forms a step-loop structure in the nascent RNA and the hairpin structure can cause RNAP pause (Artsimovitch & Landick, 1998) and weaken the interaction between the elongated oligo RNA and template DNA (Wilson & von Hippel, 1995).

The transcription reaction can also be interrupted by the action of terminator factor-Rho (ρ), a homohexameric ring protein that binds to the nascent RNA and then threads RNA 5’ to 3’ though the centre of the ring as ATP-powered translocase/helicase (Peters et al., 2011). Rho binds to C-rich unstructured RNA binding sites called rut, located upstream from the points of termination, and once the nascent RNA passes through the ring, Rho dissociates the RNAP from RNA and template DNA (Peters et al., 2011). The proteins Rho and NusG mediate the Rho-dependent transcription termination, but the mechanism is still unclear (Nudler & Gottesman, 2002; Banerjee et al., 2006; Ciampi, 2006; Peters et al., 2011). NusG is a small protein (21 kDa in E. coli) that enhances Rho termination. Rho has been reported to directly bind RNAP (Epshtein et al., 2010) while NusG has two domains to bind Rho and RNAP, respectively (Chalissery et al., 2011). The binding of NusG to both RNAP and Rho is required for increasing the rate of RNAP release by an unknown mechanism (Chalissery et al., 2011). Recently, it was reported that RNA-binding protein CsrA can bind at the upstream portion of 5’ UTR region of mRNA to unfold secondary structure and promote Rho-dependent termination (Figueroa-Bossi et al., 2014).

Protein RfaH functions as an antiterminator, which is recruited to the transcription elongation complex through specific interaction with a 12 nt conserved DNA element called ops (operon polarity suppressor) (5’-RGCCGTTAGYnT-3’) (Santangelo & Roberts, 2002; Belogurov et al., 2009). The N-terminal region of RfaH binding at the ops element can support antitermination by reducing RNAP pausing and termination at the intrinsic terminators and Rho-dependent terminators in vitro (Svetlov et al., 2007; Belogurov et al., 2010).
1.5 Transcription Regulation in Bacteria

1.5.1 Regulation of Transcription Initiation in Bacteria

Regulation can occur in any step during the transcription pathways. Transcription initiation is the most fundamental step in controlling gene expression in prokaryotes. The regulation of promoter activity is in response to environmental cues and principally mediated by different sigma factors and transcription factors (Browning & Busby, 2004).

1.5.1.1 Sigma Factors Competition

RNAP in the cells are always in low supply. The amount of free RNAP that is available to copy most of the 4000-5000 genes in the cell is limited (Ishihama, 2000). The majority of the RNAP is channelled into transcribing the genes encoding stable RNAs that are needed for translation (Browning & Busby, 2004).

The sigma (σ) factor of RNAP holoenzyme is an important determinant of promoter recognition. In *E. coli*, seven different sigma factors have been described (see above) and each of them has its own unique promoter sequence that can switch the regulation of transcription (Lonetto *et al.*, 1992). Therefore, the transcriptional gene regulation can be also determined by σ factor competition among various alternative σ factors to bind to the same core RNAP (Ishihama, 2000). The interaction among the interchangeable σ factors is dependent on many aspects, such as during bacterial growth, the concentration of RNAP availability exhibited different binding affinity to the core RNAP (Maeda *et al.*, 2000), the levels of the alarmone guanosine tetraphosphate (ppGpp) (Jishage *et al.*, 2002) and the changes in external stimuli (nutritional status, thermal and osmotic stress) (Nyström, 2004; Ganguly & Chatterji, 2012). This allows the reversible σ-switching from housekeeping factors to alternate sigma factors when the cells sense a signal under changing physiological conditions (Ganguly & Chatterji, 2012).

The ppGpp is involved in the stringent control which can cause a rapid reduction in rRNA transcription by binding to the RNAP (Kolmsee *et al.*, 2011). The ppGpp plays
a positive role in expression of a large number of $\sigma^{70}$ dependent genes that are required for stationary phase and starvation (Nyström, 2004). When the cells are under stress or growing under limited nutrition for growth, the level of ppGpp is increased and allow the alternative sigma factors ($\sigma^5$, $\sigma^{54}$, $\sigma^{32}$) to work in concert with $\sigma^{70}$ by increasing the relative competitiveness of the sigma factors (Jishage et al., 2002). It was suggested that ppGpp could slow down nucleotide addition cycles and destabilize the initial transcription complexes (Zuo et al., 2013).

Since different $\sigma$ factors have different affinities for binding core RNAP, the concentration of alternative $\sigma$ factors is a key factor in regulating transcription. The concentration of the $\sigma$ factors is partially controlled by the interaction of regulatory elements: the anti-$\sigma$ factors (Treviño-Quintanilla et al., 2013). An anti-sigma factor is defined by the ability to form a complex with its cognate $\sigma$ subunit and thereby blocking the formation of the transcription complex in a certain environmental condition (Ishihama, 2000). Anti-$\sigma$ sigma factors can work in different ways. They can weaken RNAP function by binding to the RNAP though the $\sigma$ factor or inhibit the association of the cognate $\sigma$ factor (Hughes & Mathee, 1998). In this way, promoters will not be recognized properly and expression of the corresponding genes will be silenced. Anti-$\sigma$ factors regulate a wide range of cellular processes including stress response, sporulation, bacteriophage growth, flagella biosynthesis and virulence (Hughes & Mathee, 1998).

### 1.5.1.2 Transcription Regulation by Transcription Factors

Transcription is mostly regulated by transcription factors (activator and repressors) that either recruit or block RNAP binding to the promoter, thereby activating or inhibiting transcription (Browning & Busby, 2004; Ishihama et al., 2014). From the genome sequence of *E. coli*, the total number of DNA-binding proteins that may influence transcription is estimated to be around 300 (Ishihama et al., 2014). Most of these DNA-binding proteins interact directly with the RNA polymerase and modulate its specificity of transcription initiation (Ishihama et al., 2014). Many transcription factors have dual functional roles, a given transcription factors can
act as an activator for one transcription unit and a repressor for another (Pérez-Rueda & Collado-Vides, 2000).

For the transcription activators, most of them have a specific activation domain that can recognize a specific site within the RNAP and therefore directly bind to the RNAP to stabilize the RNAP-DNA closed complex formation or accelerate the transition to the open complex (Lee et al., 2012). There are three types of activation at a simple promoter: Class I activation, Class II activation and activation by conformation change (Figure 1.9) (Browning & Busby, 2004). The Class I activation binds to the upstream of -35 promoter element and recruits the RNAP to the promoter directly interacting with the RNAP αCTD domain (Browning & Busby, 2004). This Class I activation can be best demonstrated by the transcription activation at the lac promoter with the help of a unique protein called catabolite regulatory protein (CRP). CRP-regulation requires the presence of the obligatory co-regulator cyclic adenosine monophosphate (cAMP) (Shimada et al., 2011). The CRP dimer binds to a 22-bp of promoter DNA at position -61.5 and direct interaction with αCTD domain of the RNAP and thus stabilize the formation of the initial RNAP-DNA complex (Lawson et al., 2004). In the Class II activation, the activators bind to the region between the -35 element and the σ70 subregion4 (Dove et al., 2003), which also can recruit the RNAP to the promoter binding region but may affect other steps during transcription (Browning & Busby, 2004). The best characterized of this model is that of the CI protein of bacteriophage λ that binds to the OR2 operator site at P_{RM} promoter, which overlaps 2bp of the -35 hexamer at P_{RM} promoter and activating transcription by contacting with σ70 subregion4 residues (Meyer et al., 1980; Nickels et al., 2002). Transcription activation of rhaSR and rhaBAD promoters in E. coli also shows same regulatory mechanism. Homologous transcription activator RhaS and RhaR are members of a subset of the AraC/XylS family, which play a positive role in activating the transcription of rhaSR and rhaBAD operon by interaction with the σ70 subregion4 residues respectively (Wickstrum & Egan, 2004). Finally, in the activation by conformation change model, the activator will bind to the spacer between the -10
and -35 elements and alter the conformation of the target promoter to enable the RNAP preferable binding to the promoter (Browning & Busby, 2004). For instance, the MerR family of transcription factors those normally bind between the -10 and -35 promoter elements. The MerR transcription factors binds to the suboptimal $\sigma^{70}$ dependent promoters between -10 and -35 elements activating transcription by protein-dependent DNA distortion (Brown et al., 2003).

Figure 1.9 Transcription activation at simple promoter.

Three types of transcription activation are illustrated according to the summary of Browning and Busby (2004). Since many activators function as dimer, thus it was indicated as dimer here. **(a)** Class I activation. The activator directly bound to the $\alpha$-CTD of RNAP and thereby recruiting the RNAP to the promoter. **(b)** Class II activation. The activators bound adjacent to the -35 promoter element and had interaction with the $\sigma_4$ subdomain and hence activate transcription. **(c)** Activation by conformation change. The activator bound at the spacer between -35 and -10 elements and optimized the DNA template conformation for RNAP binding or transcription activation.
Transcription repression can be achieved at different points in the transcription initiation process. These include inhibiting RNAP binding to promoter, blocking the transition from open to closed complex and inhibiting promoter clearance (Rojo, 1999). There are three models of simple repression by transcription factors and are reviewed by Browning & Busby (2004) (Figure 1.10). Firstly, the repressor can bind close to the promoter elements and overlap the RNAP binding site, which can simply inhibit the RNAP-DNA closed complex formation or prevent promoter escape (Payankaulam et al., 2010). Several repressors have been studied for many years and shown to work in this way, such as the LacI repressor that overlaps the RNAP binding region when binding to the O1 operator of the lac promoter (Schlax et al., 1995), and the lexA repressor covering the promoter region when binding at the uvrA promoter (Bertrand-Burggraf et al., 1987). Secondly, some repressors can bind to the promoter-distal sites and form a DNA loop that may block the RNAP and inhibit the transcription initiation. One classic example is that the repression of gal promoters in E. coli, where the transcription repressor GalR is bound to two operators flanking two promoters and thereby induces DNA looping and thus blocks the transcription in the looped domain (Choy et al., 1995). Similar repression mechanism has been proposed in recent studies. The protein Ler (homologous to H-NS) plays an important role in expression of chromosomal pathogenicity island, locus of enterocyte effacement (LEE) in enteropathogenic Escherichia coli (EPEC). Transcription factor Ler binds at two AATT motifs located at position -111.5 and at +65.5 of LEE1 P1 promoter and induces DNA looping formation (Bhat et al., 2014). Thirdly, the transcription repressor may have interaction with the transcription activator at the promoter region that can repress the function of transcription activator as an anti-activator (Browning & Busby, 2004). The best understood example is that the E. coli cytR regulon that is regulated under the co-ordinate interaction between transcription activator CRP and the cytidine repressor (CytR). The repressor directly interacts with activator CRP and controls the CRP-mediated transcription activation (Shin et al., 2001; Chahla et al., 2003).
In nature, many genes are controlled by two or more transcription factors and the complex regulation depends on combinations of activators and repressors, or co-dependence on more than just one regulator (Martínez-Antonio & Collado-Vides, 2003).

**Figure 1.10 Transcription repression.**

Three types of transcription activation were illustrated according to the summary of Browning and Busby (2004). (a) Repression by steric hindrance. The binding of the repressors at the promoter region could occlude/block the RNAP binding at promoter. (b) Repression by looping. The loop/bridge formation at the promoter region could block the RNAP read through the DNA template or trap the RNAP at the loop. (c) Repression by the modulation of an activator. The repressor bound to an activator can prevent the activator functioning by blocking promoter elements.

**1.5.2 Transcription Regulation by Nucleoid Associated Proteins**

The mechanism of transcription regulation is affected by the overall structure of the bacterial chromosome, which is folded into a compact structure, known as a nucleoid. Undoubtedly, the distribution of RNAP between different regulatory regions will be influenced by this compaction. The nucleoid associated protein is
involved in maintaining the DNA compaction (McLeod & Johnson, 2001; Dame, 2005; Thanbichler et al., 2005). Twelve different nucleotide associated proteins have been identified; they are H-NS (histone-like nucleoid structuring protein), StpA (suppressor of td mutant phenotype A), HU (heat-unstable nucleoid protein), FIS (factor for inversion stimulation), IHF (integration host factor), CbpA (curved DNA-binding protein A), CbpB (curved DNA-binding protein B), Dps (DNA-binding protein from starved cells), Lrp (leucine-responsive regulatory protein), DnaA (DNA-binding protein A), Hfq (host factor for phage Qβ) and IciA (inhibitor of chromosome initiation A) (Azam & Ishihama, 1999). The levels of each different nucleoid associated protein varies throughout the growth and the nucleoid proteins FIS, HU, IHF, StpA and H-NS are the most abundant in the exponential phase of growth (Ali Azam et al., 1999).

1.5.2.1 FIS

FIS exhibits the most dramatic changes in abundance during growth with over 50,000 molecules in early-log phase and near zero in stationary phase (Ali Azam et al., 1999). It is a small protein with only 98 amino acids that preferentially binds to the A-tracts and AT-tracts and plays a myriad role in regulating up to 900 regulatory target regions of the E. coli chromosome (Grainger et al., 2006; Cho et al., 2008). Four different ways by which FIS can regulate complex promoters have been found in E. coli. At the nir promoter, FIS assists IHF to suppress global transcription activator FNR (an anaerobically triggered transcription activator) mediated activation (Browning et al., 2000; Browning et al., 2004). FIS act as a simple repressor to block the RNAP binding at the promoter -10 element at the nrf promoter (Browning et al., 2002; Browning et al., 2005). In addition, FIS can displace an important activator NarL (the nitrate/nitrite ion-controlled response regulators) at yeaR and ogt promoters (Squire et al., 2009). Finally, FIS also acts as a sigma factor-dependent repressor of σ70 RNAP holoenzyme at dps promoter in E. coli (Grainger et al., 2008).
1.5.2.2 H-NS

H-NS is a small (136 amino acid, 15.5 kDa) chromatin-associated protein found in enterobacteria (Falconi et al., 1988). It is very abundant (>20,000 copies per cell) and exists predominantly as dimer or tetramers (Falconi et al., 1988). H-NS has at least two independent structural domains, including a C-terminal domain involved in the DNA-protein interactions and a N-terminal domain containing determinants for protein-protein interactions. These regions are connected by a flexible linker region that is important for H-NS oligomerization (Rimsky, 2004). The formation of the H-NS dimer is caused by interaction between two leucine zipper-kind among the N-terminal regions of the two identical monomeric subunits of H-NS (Esposito et al., 2002).

H-NS is a global, almost exclusively negative, modulator of gene expression, able to repress the transcription of approximately 5% of E. coli genes (Hommais et al., 2001). Mutations in the hns gene lead to increased synthesis of approximately 30 genes that are negatively regulated by environmental factors such as osmolality, temperature, anaerobiosis, pH or growth phase (Atlung & Ingmer, 1997; Ono et al., 2005; Queiroz et al., 2011). The hns gene is a single copy and auto-regulated at the transcriptional level, which helps to maintain a constant H-NS to DNA ratio under normal growth conditions (Atlung & Ingmer, 1997). The expression of H-NS does not vary dramatically during cellular growth. However, a measureable increase in the H-NS concentration is observed at the end of the exponential phase or at reduced growth rates (Ali Azam et al., 1999).

H-NS is known to affect DNA supercoiling (Lim et al., 2014) and to condense DNA in vitro and in vivo, similar to eukaryotic histones (Dame & Wuite, 2003) hence the name histone-like nucleoid structuring protein, or H-NS. But it is better to describe it as a nucleoid-associated protein as it lacks any sequence homology with histones (Atlung & Ingmer, 1997). Previous studies have indicated that H-NS binding results in increasing the DNA coil length, but whether H-NS induces an increase in the

56
bending rigidity of DNA or increases compaction is still unclear (Spurio et al., 1992; Amit et al., 2003).

H-NS can bind to DNA in two different ways: first, in most cases of, specific binding, it binds to AT-rich intrinsically curved DNA (Singh & Grainger, 2013); second, in non-specific binding mode H-NS binds preferentially to DNA regions showing an intrinsic curvature (Owen-Hughes et al., 1992). There are two mechanisms to explain the role of H-NS in transcription regulation and they are not mutually exclusive (Figure 1.11). Firstly in the classical mechanism of steric hindrance, H-NS forms a rigid nucleoprotein filament and directly inhibits transcription initiation as a classic inhibitor by its ability to polymerize linearly along the DNA and then block RNAP access or trap the RNAP in a DNA loop formation (Williams & Rimsky, 1997; Rojo, 2001; Dame et al., 2002; Lim et al., 2012). These two binding modes of H-NS can coexist or are interchangeable under certain ionic (Mg$^{2+}$ and K$^+$) conditions (Liu et al., 2010). For example, in vitro, H-NS oligomerizes along a linear DNA in the absence of Mg$^{2+}$ but forms DNA-protein-DNA bridges on the same DNA molecules in the presence of 10 mM MgCl$_2$ (Liu et al., 2010). Secondly, H-NS might indirectly affect transcription initiation from supercoiling-sensitive promoters influenced at a distance by acting at a specific site through an affect DNA topology of promoter DNA that is incapable of interactions with RNA polymerase (Owen-Hughes et al., 1992; Tupper et al., 1994).

In most case, H-NS binding is not restricted to a single site and there are usually several adjacent binding sites exist and binding occurs in a highly cooperative manner (Rimsky, 2004).
Figure 1.11 Model of H-NS silencing mechanism.

(A) H-NS can block RNAP accessibility by forming a H-NS nucleoprotein along RNAP promoter site and hence preventing the RNAP close complex formation. (B) The H-NS nucleoprotein filament can also cause gene silencing by potentially obstructing the elongation process of RNAP. (C) The N-NS nucleoprotein can form a DNA loop/bridge, which can trap RNAP in the loop to cause gene silencing.

The precise mechanism as to how H-NS exerts its regulatory function is not yet known and it may actually be very different for different promoters. Several different transcription regulation models by H-NS have been proposed. In the case of the *E. coli* *rrnB* P1 promoter, H-NS facilitated loop formation and trapped the RNAP to stimulate open complex formation but inhibit the promoter clearance (Dame et al., 2002). In the thermoregulation of transcription of the *virF* promoter enteropathogen *Shigella flexneri*, H-NS can form a stable nucleoprotein bridge at the *virF* promoter region which occludes the access of RNAP (Prosseda et al., 2004). H-NS can also alter the DNA topology to repress the transcription of supercoiling sensitive *Salmonella typhimurium* proU promoter (a gene encoding a high-affinity transport system for the osmoprotective solute glycine betaine) when bound downstream of the promoter (Owen-Hughes et al., 1992).

Moreover, H-NS acts co-operatively or antagonistically with other regulatory proteins to influence transcription. For example, H-NS associates with another nucleoid-associated protein Hha to form a H-NS/Hha complex to repress acquired
foreign genes in *Salmonella* (Ali *et al.*, 2013). In *S. flexneri*, IHF overcomes the repression mediated by H-NS and plays a positive role in transcription regulation of *virB* and *virF* (Porter & Dorman, 1997). H-NS has also been shown to act as a transcriptional activator, for instance, H-NS and its homologous transcription factor StpA both play a positive role in the expression of genes in the *E. coli* maltose operon (Johansson *et al.*, 1998).

### 1.5.2.3 IHF

IHF is a 22 kDa heterodimeric protein comprising two subunits IHFα and IHFβ (Rice *et al.*, 1996) (Figure 1.12). The IHFα (99 amino acids, 11.2 kDa) and IHFβ (94 amino acids, 10.7 kDa) subunits are ∼30% identical to each other and are encoded by the *E. coli* himA and himD (*hip*) genes, respectively (Nash *et al.*, 1987). IHF is able to function as both an activator and an inhibitor of transcription initiation (Swinger & Rice, 2004). It binds to the minor groove of DNA at the specific consensus sequence WATCAANNNNTTR (W=A/T; R=A/G) (Rice *et al.*, 1996). IHF protects at least 25 bp of DNA and the consensus binding sequence was located on the right side of the binding site (Goodrich *et al.*, 1990). In many systems IHF is found as an architectural element, organizing the structure of nucleoprotein complexes (Dame, 2005).

IHF plays an important role in the activation of σ70–directed transcription reactions. At the phage λ P_L promoter, IHF induces a bend believed to appropriately position an overlapping UP element to enable the binding of one or both α-CTDs (Giladi *et al.*, 1998). In this case, IHF binds just upstream of the promoter and stabilizes the closed RNAP-DNA complex. Moreover, IHF is as a direct activator that activates the *E. coli* ilvP_σ promoter through transmission of destabilized DNA by increasing the rate of open complex formation (Sheridan *et al.*, 1998). It was proposed that transcriptional activation by IHF does not involve direct interactions between IHF and RNA polymerase and is more likely to occur through IHF bending DNA leading to stronger activation (Engelhorn & Geiselmann, 1998). IHF binding sites are normally located between the upstream activator binding sites and the promoter,
and Engelhorn and Geiselmann (1988) proposed that wild type IHF induces too strong a DNA bend ($180^\circ$) for optimal interactions between DNA upstream of the IHF binding site and the back of RNA polymerase. IHF may provide an architectural role in bending the DNA to allow the assembly of the DNA-protein complex, which is necessary for transcription activation. However, IHF was also believed to physically interact with the αCTD domain of RNAP and increased the affinity of RNAP for the promoter by protein-protein interaction (Giladi et al., 1992).

In some other cases, IHF can inhibit the transcription in a similar way to classical repressors, for example IHF binding sites overlap with the consensus promoter elements at the *E. coli* *ompB* promoter (Tsui et al., 1991). IHF may also act as a repressor indirectly by obstructing the binding of the activator OmpR for *ompF* expression in *E. coli* (Ramani et al., 1992).

**Figure 1.12 Structure of the IHF-DNA complex.**

The IHF heterodimer induced a sharp bend in its binding site, causing the DNA to wrap around the body of IHF. (A) α and β subunits of IHF are indicated in white and pink, respectively. The body of the protein makes extensive contacts with phosphates on the minor groove surface of the binding site, stabilizing the bend. The conserved sequence of IHF binding site was highlighted in green and contacted with the arm of α and β subunits. (B) The IHF-DNA complex viewed from the top of the bend. Figure adapted from (Rice et al., 1996).
1.5.2.4 DNA Superhelicity and Transcription Regulation

When transcription is initiated by RNA polymerases a series of DNA conformation changes occurs when the enzyme binds to the promoter and melts the DNA template to form an open complex (Saecker et al., 2011). These intermediate states exhibit varying degrees of helical unwinding and their stabilities, thus the promoter function would be expected to be affected by the topology of the DNA template.

One of the extrinsic determinants of the transcription efficiency is the superhelical density of the DNA. It has been demonstrated that changing the negative supercoiling of DNA mainly influenced many aspects of the steps involved in transcription, such as transcription initiation, promoter clearance, elongation and termination in bacteria (Lim et al., 2003). Negative supercoiling of the DNA template is usually more efficient for transcription initiation (Richardson, 1975; Botchan, 1976). It was believed that the effect of negative DNA supercoiling on the DNA helical twist of the promoter could alter the alignment of RNAP recognition promoter elements on the face of the DNA helix. Depending on their spacing, this can either promote or inhibit RNA polymerase binding (Wang & Syvanen, 1992).

In addition, negative supercoiling density in the chromosome also affects transcription elongation rates throughout the genome in bacteria (Rovinskiy et al., 2012). The degree of supercoiling in bacteria is regulated in response to the cell growth phases and the environmental conditions through the activities of DNA topoisomerases (Balke & Gralla, 1987; Schneider et al., 1997).

DNA supercoiling is under tight homeostatic control by different DNA topoisomerases in vivo and the global negative DNA supercoiling levels are maintained by the opposing actions of topoisomerase I and IV and DNA gyrase, in which DNA gyrase introduces and topoisomerases I and IV removes negative supercoils in the DNA helix (Zechiedrich et al., 2000; Champoux, 2001). Perturbation of supercoiling, expressing inhibitors of DNA gyrase or by introducing
topA or gyr mutations can alter the level of expression of many genes (Stern glanz et al., 1981; Menzel & Gellert, 1987; Lim et al., 2003). The leu-500 promoter of S. typhimurium is one of the best-characterized supercoiling-dependent promoters. Promoter leu-500 is a point mutation in -10 hexamer at the leuABCD operon promoter and was originally isolated in S. typhimurium as a mutation inactivating the leu operon promoter and resulting in leucine auxotrophy (Gemmill et al., 1984). Promoter leu-500 activity could be reactivated in topoisomerase I gene topA mutants, therefore, the accumulation of negative DNA supercoils in topA mutant compensate for the initiation defect of the leu-500 promoter and facilitate the open complex formation (Richardson et al., 1988).

Specific binding of certain nucleotide associate protein adjacent to a promoter can induce local changes in topology and can modulate promoter activity. Some supercoiling sensitive promoters regulated by FIS have been studied, such as the E. coli fis promoter (Schneider et al., 2000), hisR promoter (a tRNA operon) in S. typhimurium (Figueroa-Bossi et al., 1998) and the E. coli tyrT promoter (a tRNA operon) (Bowater et al., 1994; Free & Dorman, 1994). There is a positive correlation between the extent of negative supercoiling and promoter activity, with increased supercoiling being associated with increased promoter activity (Schneider et al., 1997). It was proposed that the FIS-induced loops were sufficiently flexible to make a redistribution of twist and writhe leading to different topological conformations (Muskhelishvili et al., 1997). In addition, FIS-mediated supercoiling-dependent mechanism, similar to the IHF-mediated mechanism mentioned above of the ilvP6 promoter (Sheridan et al., 1998), was also discovered at the leuV operon of E. coli. The binding of FIS could translocate the superhelical energy from the promoter-distal portion of a supercoiling-induced DNA complex destabilized region to the promoter-proximal portion of the leuV promoter (Opel et al., 2004).

Transcription can also affect DNA supercoiling at RNAP complex movement. Liu and Wang (1987) proposed the twin supercoiled-domain model that transcription
can lead to local variation in supercoiling (Liu & Wang, 1987). The movement of RNAP elongation complex induces a region of positive supercoiling ahead and one of negative supercoiling behind the direction of motion (Mielke et al., 2004). Therefore, local DNA supercoiling changes are expected to be dependent on the activity of nearby actively transcribed genes. It thus implies that if the supercoiling generated during the transcription is not removed by DNA topoisomerases, the supercoiling will travel or diffuse and affect the adjacent gene transcription. For example, the leu-500 promoter has been suggested to be operated in this topological coupling mechanism (Tan et al., 1994; Mojica & Higgins, 1996; Spirito & Bossi, 1996; Wu & Fang, 2003). When the leu-500 promoter is cloned into plasmid located next to a divergent promoter, the transcription activation of leu-500 promoter was found to be completely dependent on adjacent transcription. These studies indicated that the leu500 promoter could sense the negative supercoils resulting from the transcription of adjacent sequences (Chen & Wu, 2003). Another example is the E. coli ilvY promoter of the ilvYC operon divergently coupled to the ilvC promoter. Transcription activities of the ilvY and ilvC promoters are dependent on the localized superhelical density around the promoter region (Opel & Hatfield, 2001).

Overall, transcription and supercoiling are intricately linked processes. Supercoiling could affect both initial binding of polymerase and the rate of open complex formation, and in return, transcription by RNAP can affect the superhelicity of the template DNA.

1.5.3 Complex Promoters

In many cases genes are not controlled by a single promoter but by a complex arrangement of several promoters. Depending on the direction of transcription such promoter arrangements are defined as tandem promoters, convergent promoters and divergent promoters. The transcription initiation regions of bacterial genes and operons often contain multiple promoters (Opel & Hatfield, 2001).
Approximately 40% of all transcription units in *E. coli* are transcribed from divergent promoters. Depending on the multiple promoter arrangement, the multiple promoters can interact in different ways. The overlapping divergent transcribed promoters $P_{ilvY}$ and $P_{ilvC}$ are identified as a transcriptional-coupled promoter of the *ilvYC* operon of *E. coli* (Opel & Hatfield, 2001). However, if the divergent promoters overlap each other, they can interfere, causing repression (Hanamura & Aiba, 1991). For example, the two overlapping divergent promoters (P2 and div) of the *fis* gene in *E. coli* were found to counterbalance each other (Nasser *et al*., 2002). In addition, promoter interference can also occur when the upstream tandem promoters displaying a high initiation frequency and ‘occludes’ the downstream tandem promoter (Adhya & Gottesman, 1982; Zhang & Bremer, 1996).

The coupling of multiple promoters to a single operon increases the flexibility in the adjustment of gene expression under some certain growth conditions (Rhee *et al*., 1999; Opel & Hatfield, 2001). For example, expression of the *topA* gene involves tandem promoters that utilize multiple sigma factors for transcription initiation of *topA*, allowing expression of TopA to be varied according to the adaption of *E. coli* to different growth conditions (Qi *et al*., 1997). Also, four tandem *E.coli cydAB* promoters are regulated differentially depending on the changes in the oxygen availability (Govantes *et al*., 2000). These complex promoter arrangements are normally found in regulatory region of genes with pleiotropic functions (Nasser *et al*., 2002).

**1.5.4 Transcriptional Regulation in *E. coli* Group 2 Capsule Gene Cluster**

Expression of the K5 polysaccharide gene cluster is driven by two convergent temperature-regulated promoters, Promoter 1 (PR1) and Promoter 3 (PR3), located upstream of region 1 and region 3 (Simpson *et al*., 1996; Stevens *et al*., 1997). This temperature regulation appears to be controlled at the level of transcription, as no transcription is detectable at 20°C (Rowe *et al*., 2000).
1.5.4.1 SlyA-H-NS Functional Interaction in the Regulation of Transcription

H-NS plays an unusual dual role in the temperature regulation of transcription at PR1 in *E. coli* K5. It is required for maximum transcription at 37 °C, yet also contributes to repression of transcription at 20 °C (Rowe *et al.*, 2000). In the absence of H-NS at 37 °C, transcription from PR1 was reduced approximately 50% relative to the wild type (Corbett *et al.*, 2007).

SlyA was originally identified in *S. typhimurium* and recognized as a gene that confers a haemolytic phenotype on *E. coli* K12 (Libby *et al.*, 1994). SlyA was shown to be required for the virulence of *S. typhimurium* in mice and for the survival of Salmonellae in murine peritoneal macrophages (Libby *et al.*, 1994; Stapleton *et al.*, 2002). The *slyA* gene was assumed to encode a hemolytic protein until further studies revealed the SlyA does not exhibit hemolytic activity rather it plays a novel role as a regulator that induces hemolysin expression in *E. coli* K12 (Ludwig *et al.*, 1995). Sequence alignments revealed that SlyA is a member of a large family of regulatory proteins that control different physiological process in bacterial pathogens (Wu *et al.*, 2003). SlyA is homologue of RovA from *Yersinia enterocolitica* (Nagel *et al.*, 2001). RovA was found to be essential for the virulence of *Y. enterocolitica*, and was required to regulate the expression of invasion in *Y. enterocolitica* and *Y. pseudotuberculosis* (Revell & Miller, 2000; Nagel *et al.*, 2001) through H-NS anti-repression (Heroven *et al.*, 2004; Ellison & Miller, 2006). RovA is a thermosensor, which is more susceptible to proteolysis, but SlyA still remains active and stable at 37 °C (Quade *et al.*, 2012). Studies revealed that SlyA differentially affects the expression of various proteins in *E. coli* and *S. typhimurium*, acting as both a repressor and activator of gene express (Spory *et al.*, 2002). A study analyzing the interaction of the SlyA with target DNA reported that SlyA recognizes a palindromic DNA sequence with a consensus binding site (5′-TTAGCAAGCTAA) (Stapleton *et al.*, 2002). Structural data indicate that SlyA forms homodimers in which each subunit possesses a winged- helix DNA-binding domain (Dolan *et al.*, 2011).
A more detailed understanding of the control of transcription from PR1/PR3 has been elucidated which demonstrates that transcription involves a novel interaction between H-NS and the transcriptional activator SlyA (Corbett et al., 2007; Xue et al., 2009). Expression of the *E. coli* *slyA* gene was temperature-regulated, being expressed at higher levels at 37 °C than at 20 °C and that this regulation was independent of H-NS (Corbett et al., 2007).

In the case of transcription from PR1 in *E. coli* K5, SlyA can activate and enhance capsule gene transcription at PR1. For instance, inactivation of the *slyA* gene resulted in a dramatic reduction in transcription from the PR1 while multicopy *slyA* caused significant activation of transcription at both 37 °C and 20 °C (Corbett et al., 2007). SlyA was also shown to bind preferentially to sequences upstream of the transcription start point in PR1 (Corbett et al., 2007). It was shown that H-NS is required for maximal SlyA-mediated activation of transcription from PR1, and this would explain the observation that *hns* mutants have reduced transcription from PR1 at 37 °C. The mechanism of the regulation of transcription from PR1 by the interaction of H-NS and SlyA is still unclear. With respect to the temperature regulation, it seems likely that the reduced expression of SlyA at 20 °C will alter the ratio of H-NS and SlyA in the cell and reduce transcription (Corbett et al., 2007).

In the case of transcription from PR3 in *E. coli* K5, SlyA plays a role in the activation of transcription from PR3 at 37 °C. In the presence of SlyA, H-NS is required for maximal transcription from PR3 at 37 °C. But unlike the PR1 regulation, H-NS is not essential for SlyA activity because the significant SlyA-mediated activation of transcription can be achieved in an *hns* mutant (Xue et al., 2009).

### 1.5.4.2 BipA

BipA (67 kDa), also known as tyrosine-phosphorylated ribosome-binding GTPase, is a highly conserved prokaryotic GTPase that is able to regulate gene expression at the level of translation (Farris et al., 1998). Bacterial GTPases can directly act on the ribosome to affect cellular events such as viability, protein synthesis and
pathogenesis (Caldon & March, 2003; Karbstein, 2007). Studies also suggested that the bacterial GTPase could also act to protect and stabilize the ribosome in response to the stress response, thereby serving as a regulation protein (Brown, 2005).

BipA is necessary for bacterial survival and successful invasion of the host. For example, in *E. coli*, BipA has been implicated in controlling bacterial motility (Farris *et al.*, 1998), upregulating the expression of virulence genes (Grant *et al.*, 2003) and is also responsible for avoidance of host defense mechanisms (Farris *et al.*, 1998). BipA is also thought to be involved in bacterial stress responses, such as those associated with temperature. For example, BipA was required for the growth of *E. coli* K12 at low temperature (Pfennig & Flower, 2001), while BipA null strain of *Sinorhizobium meliloti* failed to grow at low temperatures (Kiss *et al.*, 2004). An investigation of the ribosome binding properties of the protein revealed that BipA has two distinct ribosome binding modes (deLivron & Robinson, 2008). Under normal growth conditions, BipA interacts with 70S ribosomes in a GTP-dependent manner. However, under conditions of stress, the ribosome binding properties of the protein are altered such that it associates with 30S ribosomal subunits (deLivron & Robinson, 2008).

In the case of transcription from PR1 in *E. coli* K5, RNA slot blot hybridization showed that the *bipA* mutation reduced *kpsE* expression (five-fold), *kpsT* expression (four fold) and *kfiA* (seven fold) at 37 °C compared to the wild type strain (Rowe *et al.*, 2000). At 20 °C there was a two-fold increase in expression of the three genes compared to wild type (Rowe *et al.*, 2000). BipA plays a similar role at PR3 (Stevens *et al.*, 1997; Rowe *et al.*, 2000). Both H-NS and BipA are essential for maximal transcription from the region 3 at 37 °C and for repression of transcription at 20 °C. These findings suggest that the requirement of BipA for maximal capsule gene expression at 37 °C with a possible regulatory role at 20 °C. The effect of the BipA mutation on transcription from the region 1 and 3 promoters is unlikely to be via a direct interaction between the BipA protein and
the respective promoter. The most likely interpretation is that BipA regulates the activity of proteins that are required for the regulation of transcription of the region 1 and 3 promoter (Rowe et al., 2000).

1.5.4.3 IHF Regulates K5 Capsule Gene Cluster

The IHF protein is required for maximum transcription from PR1 at 37 °C and bind to a single site located 130 bp downstream of the transcription start point at PR1 (Rowe et al., 2000), but there is no evidence of the role of IHF at PR3 (Stevens et al., 1997; Rowe et al., 2000). In the case of E. coli K5 PR1, IHF serves as a transcriptional activator at 37 °C (Rowe et al., 2000). Therefore, it was thought that the IHF binding site located 130 bp downstream of the transcription start point of PR1 might bring the upstream region of the kpsF gene into close proximity with RNA polymerase and allow transcriptional activation at PR1 (Rowe et al., 2000). It is worth mentioning that a himA mutation had a more significant effect on region 1 transcription than deletion of the IHF upstream 130 bp site, which suggests that IHF may also play an indirect role in the activation of region 1 transcription, perhaps by controlling the expression of an additional trans-acting regulator (Rowe et al., 2000).

1.5.4.4 Other Factors Involved in Thermoregulation of Promoter Transcription

Rowe et al. (2000) also identified two cis-acting elements in the region 1 promoter, which are termed conA and conB. The two conserved sequence conA and conB are separated by 17bp, and located upstream of the -35 region at both promoter 1 and 3. ConA and conB are defined as cis-acting region, which is necessary for maximum transcription from the region 1 promoter. In addition, the large 5’ UTR seems to be important in moderating the level of transcription that reaches kpsM (Xue et al., 2009). The combination of a powerful promoter and a long intervening UTR to moderate transcription reaching the operon is a novel mechanism to regulate gene expression in E. coli.
1.6 Examples of Regulation of Capsule Gene in Other Gram-negative Bacteria

1.6.1 Salmonella enterica serovar Typhi

The enteric Gram-negative bacterium *Salmonella enterica* serovar Typhi (*S. typhi*) is a facultative intracellular pathogen and the cause of typhoid fever in humans (Parry et al., 2002). *S. typhi* contains a large 134-kb pathogenicity islands, known as Salmonella pathogenicity island 7 (SPI-7), that encodes the *viaB* gene cluster that directs the biosynthesis and export of capsular polysaccharide Vi antigen of *S. typhi* (Pickard et al., 2003). The Vi antigen capsular polysaccharide is a linear homopolymer of α-1,4-linked-2-deoxy-2-N-acetylgalactosamine uronic acid, which is variably O-acetylated at the C-3 position (Szu et al., 1991). Synthesis of Vi antigen is controlled by the *viaA* and *viaB*, which are located at two widely separate chromosomal loci. The *viaA* locus is commonly found in enteric bacteria while the *viaB* locus appears to be specific to Vi antigen expressing strains of *Salmonella* (Virlogeux et al., 1996). The *viaB* cluster is transcribed within human macrophages and expression of the Vi antigen is important for bacterial survival in macrophages (Hirose et al., 1997; Daigle et al., 2001). Proteins TviB, TviC, TviD and TviE are responsible for catalysing the biosynthesis of Vi capsular polysaccharide and protein VexA, VexB, VexC, VexD and VexE are responsible for the Vi antigen transport and localization at the cell surface (Wetter et al., 2012). The ABC transporter genes *vexABC* are involved in transporting the Vi polysaccharide and thereby it is thought that biosynthesis of Vi is similar to *E. coli* group 2 capsular polysaccharide (Whitfield, 2006).

The *viaA* locus contains *rcsB* and *rcsC* genes, and the RcsB protein works in concert with protein TviA to activate *viaB* transcription at the *tviA* promoter (Virlogeux et al., 1996). The *S. typhi* proteins RcsB and RcsC are homologous to their *E. coli* K12 counterparts (Virlogeux et al., 1996). Vi antigen expression in *S. typhi* involves the RcsB-RcsC two component regulatory network at both transcriptional and post-translational levels, but the mechanism still unknown (Arricau et al., 1998). In *E. coli*, RcsA is an auxiliary factor that may interact with RcsB to form a heterodimer.
required for increased transcription of colanic acid synthesis \textit{cps} genes, however, it is not involved in Vi antigen expression in \textit{S. typhi} (Virlogeux \textit{et al.}, 1996). Expression of the Vi antigen is also dependent on a functional OmpR-EnvZ two component regulatory system (Pickard \textit{et al.}, 1994). Both RcsB-RcsC and OmpR-EnvZ regulatory system seems to interact with TviA and regions upstream of the \textit{tviA} promoter and thereby linking Vi expression to environmental cues such as osmolality (Arricau \textit{et al.}, 1998; Winter \textit{et al.}, 2009). Under high osmolality conditions, the co-transcription of the \textit{tviAB} genes required for Vi antigen synthesis was significantly inhibited (Arricau \textit{et al.}, 1998). Several regulator genes have been identified that are required for efficient Vi expression by performing mutagenesis screen, such as \textit{yrfF} gene that can interact with the \textit{rcsBCD} regulatory system and \textit{barA-sirA} that related to two component regulatory genes (Pickard \textit{et al.}, 2013).

1.6.2 \textit{Neisseria meningitidis}

\textit{N. meningitidis} is an encapsulated, aerobic Gram-negative diplococcus and the cause of septicaemia and infant meningitis. A major \textit{N. meningitidis} virulence factor is the capsular polysaccharide (CPS) that prevents bacterial phagocytosis or complement-mediated lysis (Coureuil \textit{et al.}, 2012). There are six highly virulent \textit{N. meningitidis} serogroups (-A, -B, -C, -W, -Y, and -X) producing negatively charged CPS. The genes involved in the CPS synthesis and translocation are clustered at a single chromosomal locus termed \textit{cps} which is composed of three regions: Region A is serospecific and responsible for CPS synthesis; Region B (\textit{ctrE} and \textit{ctrF}) and C (\textit{ctrA-D}) are highly conserved and encode the proteins that are necessary for export and assembly of polysaccharide on the cell surface (Harrison \textit{et al.}, 2013). The CPS A serotype is composed of repeating units of \(O\)-acetylated-(\(\alpha 1\rightarrow 6\))-linked-\(N\)-acetyl-D-mannosamine-1-phosphate (Liu \textit{et al.}, 1971).

The capsules expression of \textit{N. meningitidis} serogroup B, C, W-135 and Y are composed of sialic acid derivatives (Harrison \textit{et al.}, 2013) and the expressed polysialic acid capsular polysaccharide enhances meningococci survival during invasion of the bloodstream and cerebrospinal fluid (Coureuil \textit{et al.}, 2012). The \textit{N.}}
*meningitidis* group B gene cluster has similarities to the *E. coli* K1 capsule gene cluster and contains the conserved *siaA-C* (or *cssA-C*) genes for cytidine-5’-monophosphate-N-acetylneuraminic acid synthesis (Claus *et al.*, 1997) and a *csb* gene specific to serogroup B (Harrison *et al.*, 2013). An intergenic region (134 bp) between the capsule biosynthesis operon (*sia*) and the capsule transport operon (*ctrA-D*) controls transcription of these two divergently transcribed operons (Von Loewenich *et al.*, 2001). CtrA and CtrB together form an ABC transporter and play a role in export the CPS (Harrison *et al.*, 2013). There is an UP-like element located at this intergenic region which positively influences the transcription from *siaA* gene. Deletion of the UP-like element results in a half fold reduction of the transcription activity from *siaA* (Von Loewenich *et al.*, 2001). The 5’ UTR region proceeding to *siaA* has a negative effect on the transcription of the capsule synthesis gene expression (Von Loewenich *et al.*, 2001). The genes *ctrE* and *ctrF* (formerly *lipA* and *lipB*) show homology to *kpsC* and *kpcS* from the *E. coli* Group 2 gene cluster and are required for adding the phospholipid-anchoring group onto the reducing ends of the CPS chain (Willis & Whitfield, 2013a).

The expression of *siaA* is down regulated during the intimate adhesion directly regulated by protein CrgA, which binds to the promoter region of *siaA* and negatively affects transcription (Deghmane *et al.*, 2002). The *N. meningitidis* protein CrgA is a LysR-type transcriptional regulator that is proposed to be a repressor controlling expression of a set of genes during bacterial adhesion to epithelial cells (Deghmane *et al.*, 2000). The two-component regulatory system PhoQ/PhoP (MisR/MisS) was shown to negatively mediate the expression of CPS in *N. meningitidis* as increased capsule expression was detected in *misR* and *misS* mutants (Tzeng *et al.*, 2008). However the full details of this regulation is still to be established. Temperature can trigger immune evasion by *N. meningitidis* via its RNA thermosensor located at the 5’ UTR region of *css* gene (Loh *et al.*, 2013). This RNA thermosensor can form a stable hairpin structure at lower temperature that occludes the ribosome binding sites and then halts translation; when the
temperature is increased, the secondary structure is destabilized and allowing translation (Loh et al., 2013).

In conclusion, capsular polysaccharides are high molecular mass cell-surface polysaccharides and are synthesized by enzymes located in the inner membrane and are then transported to the cell surface. They are anchored to the outer membrane of many Gram-negative bacteria, including pathogens such as E. coli, N. meningitidis, H. influenzae and S. typhi, serving as important virulence factors. Biosynthesis of Vi antigen in S. typhi is thought to be similar to E. coli Group 2 capsular polysaccharide biosynthesis (Whitfield, 2006). The ABC transporter system (kpsMT/vexABC/ctrA-D) present in E. coli, S. typhi and N. meningitidis, respectively, suggests a common pathway for capsule polysaccharide export in Gram-negative bacteria. The conservation of capsule gene organization in these species also suggests they may have a common ancestral origin.
1.7 Aims and Objectives

The aim of this project was to elucidate the regulation of expression of the region 1 operon (PR1) of the *E. coli* K5 capsule gene cluster. Particular focus was on transcriptional regulation of this operon and the role of the PR1 promoter. Specific aims were to investigate the role of 5’ UTR in regulating transcription from PR1, establish fully the number of transcriptional starts within PR1 and their contribution to expression and how regulatory proteins H-NS, IHF and SlyA regulate transcription. This study will help us to understand how this promoter controlling expression of an important virulence factor in *E. coli* is regulated.
Chapter 2. Materials and Methods

2.1 Bacterial Strains, Media and Growth Conditions

A list of *E. coli* strains used in this study is provided in Table 2.1. Unless otherwise stated, all media and solutions were sterilized by autoclaving at 121 °C for 20 min at 15 psi.

2.2 Chemicals and solutions

Unless otherwise stated, analytical grade chemicals were from Sigma. Oligonucleotide primers were from Sigma-Genosys. All media were from Oxoid. Restriction endonucleases and other DNA modifying enzymes were obtained from Roche.

2.3 Strain Maintenance and Growth Conditions

Strains were routinely grown in 5 or 10 ml Luria-Bertani broth (LB: 1% w/v tryptone, 0.5% w/v yeast extract, 1% w/v NaCl) at 37 °C, with shaking at 200 rpm for 18 h to generate an ‘overnight’ culture. For experimental assays, overnight cultures were diluted 1:100 into fresh, pre-warmed media supplemented with antibiotics as appropriate at the following concentrations: chloramphenicol, 34 μg/ml; ampicillin, 100 μg/ml; kanamycin, 50 μg/ml. Antibiotics were filter-sterilized by passing solutions through 0.22 μm Millipore filters, with the exception of chloramphenicol, which was dissolved in 100% ethanol. Growth of cultures was monitored spectrophotometrically by measuring optical density at a wavelength of OD600 nm. Over short periods, strains were maintained on LB agar plates (LB broth supplemented with 1.5% w/v agar) and stored at 4 °C. Long-term storage of strains was achieved by mixing 500 μl of an overnight culture of the required strain with 500 μl of sterile 50% (v/v) glycerol, followed by storage at -80 °C. For blue-white screening, 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) was added to final concentration of 40 μg/ml.
### Table 2.1 Strains Used During the Course of this Study

<table>
<thead>
<tr>
<th>strains</th>
<th>genotype</th>
<th>reference/source</th>
</tr>
</thead>
<tbody>
<tr>
<td>P90C</td>
<td>$F_-$, ara, $\Delta$(lacZ-pro), thi-1</td>
<td>Miller et al., 1977</td>
</tr>
<tr>
<td>DH5α</td>
<td>$F_-$, $\Phi80$lacZ$\Delta$M15, $\Delta$(lacZYA-argF), U169, recA1, endA1, hsdR17 (rK−, mK+), phoA, supE44 $\lambda$− thi-1 gyrA96, relA1</td>
<td>Hanahan, 1983</td>
</tr>
<tr>
<td>MS101</td>
<td>serA+, rpsL, K5+</td>
<td>Stevens et al., 1994</td>
</tr>
<tr>
<td>HA1</td>
<td>P90C with single copy of PR1-lacZ transduced by $\lambda$RS45 from pRS415 carrying 869 bp of PR1 fragment</td>
<td>Askar, 2004</td>
</tr>
<tr>
<td>P90C hns::kan</td>
<td>P90C × P1 (MS101 hns::kan)</td>
<td>David Corbett</td>
</tr>
<tr>
<td>P90C slyA::kan</td>
<td>P90C × P1 (MG1655 slyA::kan)</td>
<td>David Corbett</td>
</tr>
<tr>
<td>MS105 ihfB::cm</td>
<td>MS101 × P1 (MS101 $\Delta$himD::Cm)</td>
<td>Steven et al., 1994</td>
</tr>
<tr>
<td>P90C ihfB::cm</td>
<td>P90C × P1 (MS105 ihfB::cm)</td>
<td>This study</td>
</tr>
<tr>
<td>P90C bipA::kan</td>
<td>Generated by PCR mutagenesis and insertion of a kanR cassette from pKD4</td>
<td>David Corbett</td>
</tr>
<tr>
<td>MS101 hns::kan</td>
<td>MS101 × P1(P90C hns::kan)</td>
<td>This study</td>
</tr>
<tr>
<td>UTI89</td>
<td>K1+, cystitis-derived isolate of serotype O18:K1:H7</td>
<td>Cusumano et al., 2010</td>
</tr>
<tr>
<td>K5746</td>
<td>Strain N5271 bearing plasmid pPLhiphimA-5 for overproducing $\alpha$ and $\beta$ subunits of IHF</td>
<td>Nash et al., 1987</td>
</tr>
</tbody>
</table>
Table 2.2 Plasmids Used During the Course of this Study

<table>
<thead>
<tr>
<th>Plasmid Code</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBHA1</td>
<td>pBluescript SK+ carrying 872bp PR1 region from -645 to +228 in the BamH1 site</td>
<td>This study</td>
</tr>
<tr>
<td>pBluescript SK+</td>
<td>pMB1 replication origin, Amp⁴</td>
<td>Fermentas</td>
</tr>
<tr>
<td>pBTNc13</td>
<td>pBluescript containing PR1 region from -645 to +228 inserted with designed Rho-independent terminator at NcoI site, Amp⁴</td>
<td>This study</td>
</tr>
<tr>
<td>pBTNc13-IHF-BSM</td>
<td>pBluescript containing PR1 region from -645 to +228 with mutated IHF binding region and inserted with designed Rho-independent terminator at NcoI site, Amp⁴</td>
<td>This study</td>
</tr>
<tr>
<td>pCB192</td>
<td>lacZ promoter probe vector, Amp⁴</td>
<td>(Schneider and Beck, 1986)</td>
</tr>
<tr>
<td>pCBIHF-1</td>
<td>Region 1 promoter with IHF consensus binding region 13 bp deletion at +130</td>
<td>(Rowe et al., 2000)</td>
</tr>
<tr>
<td>pDSHcH</td>
<td>1.1 kb HincII-HindIII fragment containing the region 1 promoter cloned into pCB192, Amp⁴</td>
<td>(Simpson et al., 1996)</td>
</tr>
<tr>
<td>pGEM-T-easy</td>
<td>Cloning Vector</td>
<td>Promega</td>
</tr>
<tr>
<td>pHA2</td>
<td>Translational fusion vector pRS414 carrying 873bp PR1 region from -645 to +228, includes the first ATG of kpsF, Amp⁴</td>
<td>(Askar, 2004)</td>
</tr>
<tr>
<td>pJJ1</td>
<td>pRS415 containing the PR1 region from +1 to +218, Amp⁴</td>
<td>This study</td>
</tr>
<tr>
<td>pJJ133</td>
<td>pRS415 containing the PR1 region from +1 to +140 with respect to TSS +133, Amp⁴</td>
<td>This study</td>
</tr>
<tr>
<td>pJJ133-142</td>
<td>pRS415 containing the PR1 region from +1 to +146, Amp⁴</td>
<td>This study</td>
</tr>
<tr>
<td>pJJ182</td>
<td>pRS415 containing the PR1 region from +146 to +218 with respect to TSS +182, Amp⁴</td>
<td>This study</td>
</tr>
<tr>
<td>pJJ182-SDM2</td>
<td>Single site mutation at +165 with A to C substitution in PR1 region of pJJ182, Amp⁴</td>
<td>This study</td>
</tr>
<tr>
<td>pJJ182A</td>
<td>pRS415 containing the PR1 region from +93 to +218, Amp⁴</td>
<td>This study</td>
</tr>
<tr>
<td>pJJ182A-IHF-BSM</td>
<td>pRS415 containing the PR1 region from +93 to +218 with mutated IHF binding region, Amp⁴</td>
<td>This study</td>
</tr>
<tr>
<td>pJJ2</td>
<td>pRS415 containing the PR1 region from -645 to +15 with respect to TSS +1, Amp⁴</td>
<td>This study</td>
</tr>
<tr>
<td>pJJ2-Δ+1</td>
<td>Single site mutation at +12 with T to C substitution in PR1 region of pJJ2, Amp⁴</td>
<td>This study</td>
</tr>
<tr>
<td>pJJ2A</td>
<td>pRS415 containing the PR1 region from -645 to +125 with respect to TSS +1, Amp⁴</td>
<td>This study</td>
</tr>
<tr>
<td>pJJ2B</td>
<td>pRS415 containing the PR1 region from -36 to +218, Amp⁴</td>
<td>This study</td>
</tr>
<tr>
<td>pJJTNc13</td>
<td>pGEMT easy vector containing PR1 region from -645 to +228 inserted with designed Rho-independent terminator at NcoI site, Amp⁴</td>
<td>This study</td>
</tr>
<tr>
<td>pJJTNc13-IHF-BSM</td>
<td>pGEMT easy vector containing PR1 region from -645 to +228 with mutated IHF binding region and inserted with designed Rho-independent terminator at NcoI site, Amp⁴</td>
<td>This study</td>
</tr>
<tr>
<td>pJP1T</td>
<td>Transcriptional fusion vector pRS415 carrying 873bp</td>
<td>This study</td>
</tr>
<tr>
<td>Description</td>
<td>Action</td>
<td>Reference</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>PR1 region from -645 to +228 inserted with designed Rho-independent terminator at NcoI site, Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>pJP1T-IHF-BSM</td>
<td>Transcriptional fusion vector pRS415 carrying 873bp PR1 region from -645 to +228 with mutated IHF binding region and inserted with designed Rho-independent terminator at NcoI site, Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pJPR1</td>
<td>Transcriptional fusion vector pRS415 carrying 873bp PR1 region from -645 to +228, includes the first ATG of kpsF, Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pJPR1-182SDM1</td>
<td>Single site mutation at +160 with T to C substitution in PR1 region of pJPR1, Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pJPR1-182SDM2</td>
<td>Single site mutation at +165 with A to C substitution in PR1 region of pJPR1, Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pJPR1-182SDM3</td>
<td>PR1 region of pJPR1 with bases +164 to +169 changed from TAAATT to CCGATC, Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pJPR1-182SDM4</td>
<td>Single site mutation at +138 with G to A substitution in PR1 region of pJPR1, Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pJPR1-IHF-BSM</td>
<td>PR1 region of pJPR1 with bases +139 to +151 changed from TTACAACCCATTG to GCATGTGACGGAC with respect to IHF consensus binding region, Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pJPR1-SDM+1-IHF-BSM</td>
<td>PR1 region of pJPR1 with single site mutation at +12 with T to C substitution and bases +139 to +151 changed from TTACAACCCATTG to GCATGTGACGGAC, respectively, Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pJPR1-Δ+1</td>
<td>Single site mutation at +12 with T to C substitution in PR1 region of pJPR1, Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pJPR2</td>
<td>Transcriptional fusion vector pRS415 carrying 869bp PR1 region from -645 to +224, Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pLhiphimA-5</td>
<td>Expression plasmid for α and β subunits of protein IHF, Amp&lt;sup&gt;R&lt;/sup&gt; and cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Nash et al., 1987)</td>
</tr>
<tr>
<td>pRS414</td>
<td>Translational fusion reporter plasmid, Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Simons et al., 1987)</td>
</tr>
<tr>
<td>pRS415</td>
<td>Transcriptional fusion reporter plasmid, Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Simons et al., 1987)</td>
</tr>
</tbody>
</table>
2.4 DNA manipulation

2.4.1 Plasmid preparation and purification

To extract small-scale high purity plasmid DNA, the QIAquick Mini-prep Kit (Qiagen) was used in accordance with the manufacturer’s instructions. Briefly, 5 ml of a bacteria overnight culture was centrifuged for 10 min at 3600 g and the pellet was resuspended in 200 μl Buffer P1 provided. Cell lysis was performed by the addition of 250 μl Buffer P2 provided. Neutralization was achieved with the addition of 350 μl Buffer N3 provided. The precipitate was removed from the lysate by centrifugation at 13,000 rpm for 10 min in a microcentrifuge. After lysate centrifugation, the supernatant containing the soluble plasmid DNA was decanted to a spin column. Trace nuclease activity was removed by washing the spin column with 500 μl Buffer PB followed by 750 μl of Buffer PE. Plasmid DNA was recovered by eluting with 50 μl distilled Milli-Q water.

2.4.2 Quantification of DNA or RNA samples

DNA or RNA concentration was measured using the ND-1000 apparatus (NanoDrop, Rockland, DE, USA).

2.4.3 Restriction Endonuclease Digestion of DNA

All restriction endonucleases and buffers were purchased from Roche Molecular Biochemicals and used according to the manufacturer’s instruction. Restriction digests were incubated at 37 °C for at least 1 h. Endonucleases were inactivated by heating at 65 °C or 80 °C (according to different enzymes) for 20 minutes or removed by the PCR purification procedure.

2.4.4 Ligation Reaction

Unless otherwise stated, ligation reactions were typically performed in a 20 μl volume consisting of 1 μl of 10× ligation buffer (Roche), digested and purified insert fragments and vector DNA (the with a molar ratio of between them is distinct from 3:1 and to 10:1), 1 μl T4 DNA ligase (5 Units/μl) and appropriate
volume of sterile distilled water. The mixtures were incubated for 3 to 4 h at room temperature or overnight at 4 °C. Ligations were analyzed by agarose gel electrophoresis and transformed into competent DH5α cells as described below.

2.5 Competent cells preparation

2.5.1 Electro-competent E. coli Cells

Overnight cultures of the strain to be transformed were diluted 1:100 into 10 ml fresh LB Broth, supplemented with the appropriate antibiotics, and grown to the mid-exponential phase (OD600nm = 0.4 - 0.6). Cells then were transferred into a 15 ml Falcon tube and incubated on the ice for 10 min. Cells were collected by centrifugation at 3,600 g, 4 °C for 10 min and the cell pellet resuspended in 1 ml ice-cold sterile distilled water in a 1.5 ml sterile Eppendorf tube. Cells were collected again by centrifugation at 10,000 rpm for 1 min in the microcentrifuge and the supernatant carefully discarded and cell pellet resuspended in 1 ml ice-cold sterile distilled water. This wash step was repeated a further three times. After the final ice-cold water wash, the cells were washed again with 1 ml ice-cold sterile 10% glycerol and centrifuged for 1 min. Finally, cells are were resuspended in 200 µl ice-cold sterile 10% (v/v) glycerol split into 100 µl aliquots to be used immediately or stored at -80 °C for later use.

2.5.2 Chemically Competent DH5α Cells

Overnight cultures were diluted 1:100 into 200 ml LB broth, supplemented with 3 ml 1 M MgCl₂. Cells are were grown to the mid-exponential phase (OD600nm = 0.4 - 0.6), transferred into a 1.5 ml sterile Eppendorf tube and collected by centrifugation at 3,600 g for 10 min at 4 °C. The supernatant was removed and cells resuspended in 60 ml ice-cold sterile solution A (600 µl of 1 M MgCl₂, 3 ml of 1 M CaCl₂, 12 ml of 50 mM MES, add sterile distilled water to the final volume 60 ml). Following incubation on the ice for 20 min, cells were pelleted by centrifuging at 6,000 rpm for 10 min at 4 °C. Resuspended the pellet in 12 ml sterile solution B (120 µl of 1 M MgCl₂, 600 µl of 1 M CaCl₂, 2.4 ml of 50 mM MES, 3.6 ml of 50% glycerol, add sterile distilled water to the final volume 60 ml). The mixture was
incubated on ice for 2 h and used immediately or split into 100 µl aliquots sterile eppendorf tubes and quick-frozen in liquid nitrogen for storage at -80 °C.

2.6 Transformation of Chemically Competent E. coli Cells

Thawed 100 µl aliquots of competent cells were mixed with 10 µl ligation mix or plasmid DNA purified form from cells and incubated on ice for 30 min. Cells were heat-shocked for 60 sec at 42 °C water bath and place on ice for 5 min. Cells were recovered by addition of 800 µl sterile LB broth and following incubat at 37 °C for 1 h, they were plated on the selective agar plates appropriate for the transformed DNA. Plates were incubated at 37 °C overnight. Analysis of transformants was carried out by DNA plasmid Mini-Preparation, followed by restriction digest analysis and agarose gel electrophoresis, or colony polymerase chain reaction (Colony PCR) as described below (see section 2.10).

2.7 Transformation of Electrocompetent E. coli Cells

Thawed 100 µl competent cells were mixed with 5 – 10 µl purified plasmid DNA as appropriate and transferred to a pre-chilled 0.2 mm electroporation cuvette (Bio-Rad). Cells were electroporated with the following parameters: 2.5 kV, 200 Ω, 25 µF. Electroporated cells were mixed with 800 µl LB broth and incubated for 1 to 3 h at 37 °C. Aliquots of 200 µl of cells were spread onto selective agar plates and incubated overnight at 37 °C. Analysis of transformants was carried out by DNA plasmid Mini-Preparation, followed by restriction digest analysis and agarose gel electrophoresis, or colony polymerase chain reaction (Colony PCR) as described below (see section 2.10).

2.8 Agarose gel electrophoresis

Agarose gels were prepared by melting of 0.75% - 2.0% (w/v) agarose (Longza) in Tris-Acetate-EDTA buffer (TAE Buffer: 0.5 M Tris, 5.7% acetic acid, 10 mM EDTA, pH 8.0). 6× DNA loading buffer (0.1% (w/v) bromophenol blue; 20 mM Tris, pH 8.0; 30% glycerol was added to the sample and the mixture loaded in the wells of the gel. DNA samples were electrophoresed at 6 V/cm in TAE buffer with 5 µg/ml of
ethidium bromide. DNA fragments sizes were calibrated with 1 kb or 100 bp markers (Hyperladder I and IV respectively, Bioline).

**2.9 Polymerase Chain Reaction (PCR)**

Standard PCR was performed using a Touchdown Thermal Cycler (Hybaid). Each 50 μl reaction contained 5 μl 10 μM of appropriate forward and reverse primers, 5 μl 10× PCR Buffer (Roche Molecular Biochemicals), 5 μl 2.5 mM dNTPs (Bioline), 0.5 - 1 μg template DNA, 2 - 3 μl Taq polymerase (Roche Molecular Biochemicals) with sterile distilled water added to a final volume 50 μl. PCR was carried out under the following PCR conditions: 95 °C Denaturation for 30 sec, Annealing (Tₐ is estimated as 5 °C lower than the primer melting temperature Tₘ) for 30 sec and 72 °C Extension for n min (“n” is the estimated length in kb of the fragment being amplified) for 30 cycles, followed by 1 cycle of 72 °C for 10 min. All primers used in this study were presented in Table 2.3 and were purchased from Sigma-Aldrich. PCR products were assessed by agarose gel electrophoresis and purified by using QIAquick PCR purification kits (Qiagen) or High Pure PCR products kits (Roche) according to the manufacturer’s instructions. The sequences and characterization of the primers used in this study are listed below in Table 2.3.
<table>
<thead>
<tr>
<th>Primer</th>
<th>sequence (5’-3’)</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>pJJ1-F-EcoR1</td>
<td>AGCCCTgaatttcCAAAAATTTGGTCCCTTCTCG</td>
<td>Forward primer for construction of pJPR1, pJJ1, pJJ3, pJJ133-142, pJJ133</td>
</tr>
<tr>
<td>pJJ1-R-Bam H1</td>
<td>AATCACgagatccCTTTTGACGGAAAATAATGC</td>
<td>Reverse primer for construction of pJPR1, pJJ1, pJJ182A, pJJ2B, pJJ182; Specific primer 2 for pJPR1/pHA2/pJJ1/MS101 5’RACE second round PCR</td>
</tr>
<tr>
<td>pJJ2-F-EcoR1</td>
<td>CGACCACgagatccGACTGTATGCGTAAATTGCTG</td>
<td>Forward primer for construction of pJJ2, pJJ2A</td>
</tr>
<tr>
<td>pJJ2-R-Bam H1</td>
<td>GATAAggatccGGGAACAAATTTTGCCACC</td>
<td>Reverse primer for construction of pJJ2</td>
</tr>
<tr>
<td>pRS415-LacZ-SEQ</td>
<td>GCAAGGCGATTAAGTGGGAAC</td>
<td>Specific primer 1 for pJPR1/pHA2/pJJ1 5’RACE first round PCR; Screen for insertion in the pRS415 plasmid from the lacZ gene</td>
</tr>
<tr>
<td>P1-Internal-SEQ</td>
<td>TGAATGCGGAATTAATCTTG</td>
<td>Screen for P1 promoter inserts in the pRS415 plasmid from the promoter region</td>
</tr>
<tr>
<td>pJJ182A-F-EcoR1</td>
<td>AGCCTgagatccATATAAGCATGGAATTAAATCTGG</td>
<td>Forward primer for construction of pJJ182</td>
</tr>
<tr>
<td>pJJ133-142-R-Bam H1</td>
<td>GTACAggatccGGTTGTAACAAAAATTCTTGGAAC</td>
<td>Reverse primer for construction of pJJ133-142</td>
</tr>
<tr>
<td>pJJ2A-R-Bam H1</td>
<td>GTACAggatccGAATAAACACCTGCTAGTCC</td>
<td>Reverse primer for construction of pJJ2A</td>
</tr>
<tr>
<td>pJJ2B-F-EcoR1</td>
<td>AGCTtgagatccCTTTGCTATGCTTGGGAT</td>
<td>Forward primer for construction of pJJ2B</td>
</tr>
<tr>
<td>pJPR2-F</td>
<td>CTGCCAGGAAATGGGATCGGA</td>
<td>Forward primer for construction of pJPR2</td>
</tr>
<tr>
<td>pJPR2-R-Bam H1</td>
<td>GACGCCATTTGGGATCGGA</td>
<td>Reverse primer for construction of pJPR2</td>
</tr>
<tr>
<td>pJJ182-F-EcoR1</td>
<td>AGACTtgagatccCTTTGCTATGCTTGGGAT</td>
<td>Forward primer for construction of pJJ182, pJJ182-SDM2</td>
</tr>
<tr>
<td>pJJ133-R-BamH1</td>
<td>GGTCAgagatccCAGAAAAATTTTAACTGATTAATAAC</td>
<td>Reverse primer for construction of pJJ133</td>
</tr>
<tr>
<td>SP1-MS101</td>
<td>AAGTAggatccGAAGTTTGTGCTTGGGATCGCG</td>
<td>Specific primer 1 for MS101 5’ RACE first round PCR</td>
</tr>
<tr>
<td>SP2-MS101-6FAM</td>
<td>[6FAM]GGCGAACAGAGGTAATTAGATATGCG</td>
<td>Specific primer 2 for MS101 modified 5’ RACE second round PCR</td>
</tr>
<tr>
<td>SP6 Promoter Primer</td>
<td>GATTTAGGTGACACTATAG</td>
<td>Forward primer for screening pGEMT-easy vector inserts</td>
</tr>
<tr>
<td>T7 Promoter Primer</td>
<td>GTAATACGACTCAGTATAGGG</td>
<td>Reverse primer for screening pGEMT-easy vector inserts</td>
</tr>
<tr>
<td>HNS-F</td>
<td>CCACCCCCAATATAAGTGGAG</td>
<td>Screen for the hns gene mutant in P90C</td>
</tr>
<tr>
<td>HNS-R</td>
<td>GCTGCGGCGAATTTAAGACAG</td>
<td>Screen for the hns gene mutant in P90C</td>
</tr>
<tr>
<td>Smut-F</td>
<td>GGTGTTCTGATCGGACGGTTGG</td>
<td>Screen for the slyA gene mutant in P90C</td>
</tr>
<tr>
<td>Smut-R</td>
<td>CGCCCCCTTCATTCCAAC</td>
<td>Screen for the slyA gene mutant in P90C</td>
</tr>
<tr>
<td>IHF-F</td>
<td>GAAGCTAAATTTCCCGGACTTGG</td>
<td>Screen for the IHF gene mutant in P90C</td>
</tr>
<tr>
<td>IHF-R</td>
<td>ATACACCAGCTCGTTATTTTCACTG</td>
<td>Screen for the IHF gene mutant in P90C</td>
</tr>
<tr>
<td>Primer ID</td>
<td>Primer Sequence</td>
<td>Description</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>EMSA_IHF_F</td>
<td>GCACCTCCATGAGACATT</td>
<td>Forward primer for amplification of IHF consensus region for IHF-EMSA</td>
</tr>
<tr>
<td>EMSA_IHF_R</td>
<td>CAGCTCCTTTGCACGG</td>
<td>Reverse primer for amplification of IHF consensus region for IHF-EMSA</td>
</tr>
<tr>
<td>M13_F</td>
<td>TGTA AACGACGGCCAGT</td>
<td>Forward primer for amplification of 225 bp length negative control free DNA for EMSA</td>
</tr>
<tr>
<td>M13_R</td>
<td>GGAACAGCTATGACCAG</td>
<td>Reverse primer for amplification of 225 bp length negative control free DNA for EMSA; primer for sequencing of pBHA1</td>
</tr>
<tr>
<td>HNS01_F</td>
<td>GGCAGATTTAATTCCGCATTCA</td>
<td>Forward primer for amplification of 372 bp length DNA from 5' UTR of kpsf gene for HNS-EMSA</td>
</tr>
<tr>
<td>HNS01_R</td>
<td>CGATATTACAGGCGAGAAAGGG AA</td>
<td>Reverse primer for amplification of 372 bp length DNA from 5' UTR of kpsf gene for HNS-EMSA</td>
</tr>
<tr>
<td>HNS02_F</td>
<td>GTCTTCGCACCTCCATGAG</td>
<td>Forward primer for amplification of 170 bp length DNA from 5' UTR of kpsf gene for HNS-EMSA</td>
</tr>
<tr>
<td>HNS02_R</td>
<td>GAAATAATGCACAGTCACAC</td>
<td>Reverse primer for amplification of 170 bp length DNA from 5' UTR of kpsf gene for HNS-EMSA</td>
</tr>
<tr>
<td>SDMPR1-1_F</td>
<td>GTTCATAATGTAGGAGTGTGTA AATAGGTTGGGCAAAATTGG</td>
<td>Forward primer for Site-direct mutagenesis in the second T to C of -10 TATA box of PR1-1</td>
</tr>
<tr>
<td>SDMPR1-1_R</td>
<td>CCAAAATTTCGCCACCTATTTG GAC ACACTCTACATTAGAAC</td>
<td>Reverse primer for Site-direct mutagenesis in the second T to C of -10 TATA box of PR1-1</td>
</tr>
<tr>
<td>SDMPR1-4_F1</td>
<td>GTTACAACCCATTGATTAGCACA AAATAAATTATAGGTTGGTTCGGG</td>
<td>Forward primer for Site-direct mutagenesis in the second T to C of predicted -10 TATA box of PR1-4</td>
</tr>
<tr>
<td>SDMPR1-4_R1</td>
<td>CCCGAACCCACTATAATTATTTG TGCTAAATCAATGGTGTTGTAAC</td>
<td>Reverse primer for Site-direct mutagenesis in the second T to C of predicted -10 TATA box of PR1-4</td>
</tr>
<tr>
<td>SDMPR1-4_F2</td>
<td>CAACCCATTGATTAGCATAAAATC AAATTATAGGTTGGTGTTTTG</td>
<td>Forward primer for Site-direct mutagenesis in the second A to C of predicted -10 TATA box of PR1-4</td>
</tr>
<tr>
<td>SDMPR1-4_R2</td>
<td>CAACCCCGAACCCACTATAATTTG ATTATGCTAAATCAATGGTGTTGG</td>
<td>Reverse primer for Site-direct mutagenesis in the second A to C of predicted -10 TATA box of PR1-4</td>
</tr>
<tr>
<td>SDMPR1-4_F3</td>
<td>CAACCCATTGATTAGCATAAAACCGATCAATGGGTTGGTGTTTTGTG</td>
<td>Forward primer for Site-direct mutagenesis of predicted -10 TATA box of PR1-4</td>
</tr>
<tr>
<td>SDMPR1-4_R3</td>
<td>CAACAAACCCGAACCCACTATAGA TCGGTTTATGCTAAATCAATGGGGTTTGG</td>
<td>Reverse primer for Site-direct mutagenesis of predicted -10 TATA box of PR1-4</td>
</tr>
<tr>
<td>SDMPR1-4_F4</td>
<td>GCATGGACTGACCATGGTTTTTAT ATTCAATTAAAAATTTATTACAAC CATTGATTAGTC</td>
<td>Forward primer for Site-direct mutagenesis in the third G to A of predicted -35 box of PR1-4</td>
</tr>
<tr>
<td>SDMPR1-4_R4</td>
<td>GCTAAATCAATGGGTTGTAATAA AATTTTTATGAAATATTAAACCCAT GTCAGTCCATGC</td>
<td>Reverse primer for Site-direct mutagenesis in the third G to A of predicted -35 box of PR1-4</td>
</tr>
<tr>
<td>Primer Type</td>
<td>Primer Sequence</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>pJJ182_SDM_F</td>
<td>AGAGT[gaattc]CCATTGATTTAGCA TAAATCAATTATAG</td>
<td>Forward primer for construction of pJJ182-SDM2</td>
</tr>
<tr>
<td>IHF_BSM_F</td>
<td>CCACTGGTTATATTTATCTAAAAATTTGGCATGTCAGCACTTTAATTAGTCAG</td>
<td>Forward primer for Site-direct mutagenesis of IHF consensus region at +130</td>
</tr>
<tr>
<td>IHF_BSM_R</td>
<td>CGAACCACACTATAATTATTTATATGCTCAATAGTCCTCAATGGCAAAAAATTTTTATGAATATAAAACCACATGG</td>
<td>Reverse primer for Site-direct mutagenesis of IHF consensus region at +130</td>
</tr>
<tr>
<td>MS101-qPCR-F</td>
<td>TGGCATTATTTCCGTGCAAAGGACG</td>
<td>Forward primer for amplification of 126 bp 5' region of kpsf gene in MS101 for qPCR</td>
</tr>
<tr>
<td>MS101-qPCR-R</td>
<td>TGCTGCGCCTTTGTTCTGCCA</td>
<td>Reverse primer for amplification of 126 bp 5' region of kpsf gene in MS101 for qPCR</td>
</tr>
<tr>
<td>rpoD-qPCR-F</td>
<td>CGGACCGTACGCTCGGTTCTTCT</td>
<td>Forward primer for amplification of 121 bp of rpoD gene for qPCR</td>
</tr>
<tr>
<td>rpoD-qPCR-R</td>
<td>GCCAGTTCCGGGTCGATGCT</td>
<td>Reverse primer for amplification of 121 bp of rpoD gene for qPCR</td>
</tr>
<tr>
<td>1_1a_F</td>
<td>GCCTGTAATATCCGACCGTC</td>
<td>Forward primer for amplification of 80 bp 3' region of +1 TSS in pRS415 promoter fusion for qPCR</td>
</tr>
<tr>
<td>1_1a_R</td>
<td>TGCTTATATGCAGGACTTATCACC</td>
<td>Reverse primer for amplification of 80 bp 3' region of +1 TSS in pRS415 promoter fusion for qPCR</td>
</tr>
<tr>
<td>1_1_F</td>
<td>CGTCTTCGACCTCCATGAGACATG</td>
<td>Forward primer for amplification of 177 bp 3' region of +1 TSS in pRS415 promoter fusion for qPCR</td>
</tr>
<tr>
<td>1_1_R</td>
<td>TGCACGGAATAATGCCACAGTC</td>
<td>Reverse primer for amplification of 177 bp 3' region of +1 TSS in pRS415 promoter fusion for qPCR</td>
</tr>
<tr>
<td>1_4_F</td>
<td>GCAGTTAATCCACAGCGAGCCAG</td>
<td>Forward primer for amplification of 129 bp 5' upstream region of lacZ gene in pRS415 promoter fusion for qPCR</td>
</tr>
<tr>
<td>1_4_R</td>
<td>AGGGTTTTCCACAGCGACGGTTG</td>
<td>Reverse primer for amplification of 129 bp 5' upstream region of lacZ gene in pRS415 promoter fusion for qPCR</td>
</tr>
<tr>
<td>UTI89_1_1a_F</td>
<td>GTAATATTGCAGCATCTTCGACC</td>
<td>Forward primer for amplification of 76 bp 3' region of +1 TSS at PR1 region of kpsf gene in UTI89 for qPCR</td>
</tr>
<tr>
<td>UTI89_1_1a_R</td>
<td>TGCTTATATGCAGGACTTATCACC</td>
<td>Reverse primer for amplification of 76 bp 3' region of +1 TSS at PR1 region of kpsf gene in UTI89 for qPCR</td>
</tr>
<tr>
<td>UTI89_1_1_F</td>
<td>GTAATATTGCAGCATCTTCGACC</td>
<td>Forward primer for amplification of 188 bp 3' region of +1 TSS at PR1 region of kpsf gene in UTI89 for qPCR</td>
</tr>
<tr>
<td>UTI89_1_1_R</td>
<td>GCACGGAAAAATGCGACAGTC</td>
<td>Reverse primer for amplification of 188 bp 3' region of +1 TSS at PR1 region of kpsf gene in UTI89 for qPCR</td>
</tr>
<tr>
<td>UTI89_kpsf_F</td>
<td>TTTCCGTGCAAAAGGAGCTGA</td>
<td>Forward primer for amplification of 168 bp 5' coding region of kpsf gene in UTI89</td>
</tr>
<tr>
<td>UTI89_kpsf_R</td>
<td>AAGGACACGGCTGGTACTGCC</td>
<td>Reverse primer for amplification of 168 bp 5' coding region of kpsf gene in UTI89</td>
</tr>
</tbody>
</table>
bp 5' coding region of kpsf gene in UTI89

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>16s-qPCR-F</td>
<td>ATTCCAGGTGTAGCGTGAAATGC</td>
<td>Forward primer for amplification of 169 bp of 16s gene for qPCR</td>
</tr>
<tr>
<td>16s-qPCR-R</td>
<td>CAAGGGCACAACCTCCAAGTCG</td>
<td>Forward primer for amplification of 169 bp of 16s gene for qPCR</td>
</tr>
<tr>
<td>PR1-1 terminator</td>
<td>CACAGGCCTccatggAAATAGACTGTCCCGAACCCGCGGTTCTGCTGTTTCTGTTTTTGGCAGAACCCGACCTCAACGTAAGAGCTCATAATGCGTCCc catggAAGTAGCT</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.3 Primers Used During the Course of this Study.

Lower case indicates the enzyme cleavage site, underlining indicates the consensus functional elements, bold characters indicates mutated nucleotides.

2.10 Colony PCR

Colony PCR can be used as a convenient screen following a transformation to identify successful ligations. The method was performed as standard PCR apart from using a single bacterial colony instead of a DNA template. Adding a single colony to a reaction mix by picking the colony with either a sterile pipette tip and resuspending in the reaction tube containing sterile distilled water, then streaked it onto a fresh replicate agar plate using a numbered plate. Therefore, streaking to this numbered plate will retrieve these colonies. For colony PCR, the reaction needs an additional step of 5 min at 95 °C for one cycle at the beginning to lyse cells.

2.11 DNA Sequencing

DNA sequencing was carried out using the BigDye version V1.1 Terminator Cycle Ready Reaction kit (Life Technologies) for cycle sequencing according to the manufacturer’s instructions. For each 20 μl sequencing reaction, 100 - 300 ng of plasmid DNA was mixed with 3.2 μl 1 pmol/μl custom forward or reverse primer, 3 μl 5 × BigDye Sequencing Buffer and 2 μl BigDye Terminator Mix. The reactions were made up to 20 μl with sterile dH₂O in a thin-walled 200 μl Eppendorf tubes. The reaction was then incubated in a Hybaid thermal cycler with the following parameters for 25 cycles: 96 °C × 30 s; 50 °C × 15 s; 60 °C × 4 min. To purify DNA after the sequencing reaction, to each PCR tube, 2 μl of 125 mM EDTA, 2 μl 3 M NaOAc (pH=4.5) and 60 μl 100% ethanol were added and mixed, and following 15
min incubation at room temperature. After 20 min of centrifugation at 13,000 rpm, the Ethanol-EDTA supernatant was carefully removed and the DNA pellet washed with 200 µl 70% ethanol and air-dried. The DNA pellets were resuspended and rearrayed into a plate format appropriate for ABI Prism 3100 Genetic Analyzer. Nucleotides were compared using the nucleotide BLAST website (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.12 Beta-Galactosidase Assay

Assays were carried out essentially as described by Miller (1972). Triplicate overnight cultures were diluted 1:100 into fresh, pre-warmed LB broth supplemented with appropriate antibiotics and grown to mid-exponential phase. Aliquots of each 100 in mid-log culture were mixed in a 2 ml eppendorf tube with 900 µl Z-Buffer (400 ml of 0.85 % w/v Na₂HPO₄, 0.55 % w/v NaH₂PO₄, 0.07 % w/v KCl, 0.025 % w/v MgSO₄, filter-sterilized into 50 ml aliquots in 50 ml falcon tubes, and 135 µl 2-Mercaptoethanol into each 50 ml portion just before use), and lysed with 40 µl chloroform, 20 µl 0.01 % SDS. After thorough mixing for 1 min, the samples were allowed to settle for 1 h. Two 176 µl samples from each tube were placed in adjacent wells on a 96 well plate and 35 µl o-nitrophenyl-β-D-galactoside (ONPG: 4 mg/ml in Z-Buffer, store in dark) added to the even-numbered wells using a multi-channel pipette. Allow a yellow coloration to develop and keep the plate in dark. Once sufficient colour developed, the reaction was terminated by addition of 88 µl of 1 M Na₂CO₃ (normally between 20 min to 60 min). The extent of the yellow developed coloration was assessed by measuring the OD₄₂₀ nm in a plate reader (Synergy™ HT Multi-Mode Microplate Reader) for each reaction. The Beta-Galactosidase activity in was calculated as below.

\[
\text{β-Galactosidase activity} = \frac{\text{OD}_{420 \text{nm (Measure)}} - \text{OD}_{420 \text{nm (Blank)}}}{\text{OD}_{600 \text{nm}} \times T \times V} \times 1000
\]

(Miller Units)
where $\text{OD}_{420\text{nm(measure)}}$ = the value of the even-numbered wells, $\text{OD}_{420\text{nm(Blank)}}$ = the value of the odd-numbered well corresponding to the even-numbered well, $T$ = time (min), $V$ = volume (ml), and 1 Miller Unit is equivalent to the amount of enzyme which produced 1 $\mu$mol o-nitrophenol/min.

2.13 RNA Extraction and Clean up

Overnight *E. coli* cultures were inoculated into fresh LB medium (1:100) and grown at 37 °C to mid-exponential phase ($\text{OD}_{600\text{nm}}$=0.4 - 0.6) and 1 ml of the culture removed into a 15 ml falcon tube and mixed with 2 ml RNA Protect Bacteria Reagent (Qiagen), following with suspension mixing was incubated at room temperature for 5 min. Cells were collected by centrifugation at 3600 $g$ for 10 min at room temperature. The supernatant was carefully discarded and the cell pellet stored at -20 °C or used immediately. Cells pellets were resuspended in 100 µl lysosome-TE buffer (20 µl 20 mg/ml lysosome stock in 1 ml TE buffer) and incubated at room temperature for 5 min. 350 µl Buffer RLT was added (add 10 µl β-ME in 1 ml Buffer RLT aliquots just before use), vortexed and centrifuged for 2 min at 13,000 rpm. To this 250 µl of 100% ethanol was added and mixed by pipetting. Discard the sample to the RNeasy mini column provided and centrifuge for 15 s at 13,000 rpm to bind the RNA into the silica-gel membrane. Wash with 700 µl Buffer RW1 before centrifuge 15 s at 13,000 rpm. Then wash with 500 µl Buffer RPE provided and centrifuge for 2 min at 13,000 rpm. Centrifuge for another 1 min to eliminate the residual Buffer RPE. Finally, elute the RNA from the membrane by adding 50 µl RNase-free water provided. All equipment was treated with 75% ethanol and RNase ZAP (Ambion) before use. The disposable equipments were used sterile un-opened package of eppendorf tubes and un-opened RNase-free, DNase-free tips. RNase free water was achieved by adding 0.1 ml DEPC into 100 ml of the Milli-Q water (0.1% DEPC) and left it overnight with stirring, then autoclaved twice. All the solutions were made with RNase free water with the exception of those containing Tris-buffer. After RNA was extracted, it can be stored at – 80 °C. Analysis whether the RNA sample get DNA contamination or not, the RNA sample should be done PCR with appropriate primers and agarose gel
electrophoresis. If the RNA sample got DNA contamination, the RNA samples should be treated with DNase digestion kit (Qiagen). Briefly, 50 µl RNA solution mixed with 10 µl provided Buffer RDD, 2.5 µl provided Dnase I solution, added RNase free water to the final volume 100 µl, followed with incubation at 37 °C water bath for 40 min. The DNase treated RNA sample have to re-clean up by RNeasy Protect Bacteria Mini Kit (Qiagen) according to the manufacturer’s instructions. Finally, RNA was quantified using a NanoDrop spectrophotometer.

2.14 Rapid Amplification of 5’ cDNA ends (5’ RACE)

5’ RACE was carried out using 5’/3’ RACE kit (2nd Generation, Roche) according to the manufacturer’s instructions. The protocol can be briefly summarized as follows:
1. First strand cDNA synthesis. The extracted mRNA was used as starting sample material. The first strand cDNA is synthesized from total RNA using a gene-specific primer SP1, Transcription Reverse Transcriptase and the deoxynucleotide mixture.
2. Purification. The first-strand cDNA was purified from unincorporated nucleotides and primers using the High Pure PCR Product Purification Kit (Roche).
3. Addition of homopolymeric A-tail. By using the terminal transferase to add a homopolymeric A-tail to the 3’ ends of the cDNA.
4. First Round PCR amplification. Tailed cDNA was amplified by PCR using a gene-specific primer SP2 and the Oligo dT-anchor primer.
5. Second Round PCR amplification. The obtained cDNA was further amplified by a second PCR using a nested PCR using gene-specific primer SP3 and the PCR anchor primer (Figure 2.1). The schematic picture of 5’RACE workflow can be seen in Figure 2.1. As the result, the obtained 5’ RACE products were cloned into an appropriate vector for subsequent characterization procedures, which included pGEM-T Easy Vector System ligation, Blue-white Screening, Colony PCR and Sequencing.
Figure 2. 1 Schematic picture of 5’RACE work flow

2.15 Modified 5’RACE- Capillary electrophoresis using a DNA sequencer

In the modified 5’ RACE experiments, all the steps were completely same as described above except the second round PCR reverse primer (SP3) was modified by fluorescent dye 6-FAM at the 5’ end of the specific reverse primer. The obtained PCR products labeled by 6-FAM can be directly sequenced for the length of all amplified products. Each sequencing PCR plate contains 9.4 µl Hi-Di formamide (ABI), 0.1 µl GS500LIZ (ABI) size standards and 0.5 µl of 150 ng/µl DNA products. Each sample was run in triplicate.

2.16 pGEM-T Vector System

The pGEM-T Vector System (Promega) was used in accordance with the manufacturer’s instructions described as follows.

2.16.1 Ligation Protocol

Set up ligation reactions named Standard Reaction and Positive Control. The standard reaction only contained 5 µl 2 × Rapid Ligation Buffer provided, 1 µl pGEM-T easy Vector, 3 µl PCR products wished to be ligated into pGEM-T vector, 1 µl T4 DNA ligase provided and 1 µl nuclease-free water to final volume 10 µl. Positive control was set up as same as the standard reaction except adding 1 µl
nuclease-free water instead of PCR products. The reaction tubes were incubated at room temperature for 1 h.

2.16.2 Transformation Protocol

Transformation for the ligated pGEM-T Vector used high-efficiency DH5α competent cells (≥1 × 108 cfu/μg DNA). Thawed 100 μl competent cells were mixed with 10 μl ligated pGEM-T vector followed by incubation on ice for 30 min. Then cells were heat-shocked for 90 sec in the 42 °C water bath before being returned to the ice for 5 min. After the cells were recovered by addition of 900 μl sterile LB broth and incubated at 37 °C for 1 h, they were plated on the Xgal-Amp plates. Plates then were incubated at 37 °C for overnight. Analysis of transformants was carried out by Blue/White Selection of Recombinant (Methods showed below), colony polymerase chain reaction and checked by the agarose gel electrophorasis in order to select the potential colony for sequencing.

2.16.3 Blue/White Selection of Recombinants

The pGEM-T Easy Vectors are high-copy-number vectors containing T7 and SP6 RNA polymerase promoters flanking a multiple cloning region within the α-peptide coding region of the enzyme β-galactosidase. Insertion inactivation of the α-peptide allows identification of recombinants by blue/white screening on indicator plates. Therefore, after the overnight incubation, selected the white colonies grown on the plates were selected and analyzed by Colony PCR after incubation at 37 °C overnight.

2.17 Site Direct Mutagenesis by PCR

This protocol was performed according to the method of Loening (2005). A pair of complemented mutagenic primers was designed containing the desired substitution of nucleotides flanked by ~ 20bp unmodified nucleotides on each side of the mutation site. Each 50 μl PCR reaction contains 5 μl 10× Pfu Ultra buffer (Stratagene), 2 μl of 25 ng/μl template plasmid DNA, 12.5 μl of 10 ng/μl of each
primer, 2.5 µl of 2.5mM dNTP, 1 µl Strategy Pfu Ultra polymerase (Stratagene) and 14.5 µl ddH₂O. PCR was carried out under the following cycles:

95 °C × 5 min - 1 cycle
95 °C× 50 sec, -30 cycles
60 °C× 50 sec,
68 °C× 1 min+1 min/1 kb template
7 min × 68 °C - 1 cycle

The PCR product was immediately digested with Dpn I before transforming into *E.coli* DH5α and then plated out on the LB plate with appropriate antibiotics. Successful mutated plasmids was purified from cultured single colony and verified by sequencing. A schematic picture of this procedure is shown in Figure 2.2 below.

![Figure 2.2 Schematic picture of site-directed mutagenesis (Loening, 2005)](image)

**2.18 Generalized Transduction of *E. coli***

Phage P1 was used in strain construction to move mutations from donor to recipient strains.

**2.18.1 Preparation of P1 lysate**

Two sterile universal bottles containing 5 ml LB broth supplemented with 0.2% sucrose, 5 mM CaCl₂, and antibiotics as appropriate, were inoculated with 200 µl of an overnight culture of the donor strain. Each culture was serially diluted into three further fresh universals (10⁻¹ to 10⁻³) containing supplemented LB as
described above. All eight cultures were incubated at 37 °C with shaking at 200 rpm for 30 min. After this time, one dilution series was inoculated with 100 μl P1 labeled as ‘donor’, whilst nothing was added to the second series labeled as ‘control’. All cultures were incubated at 37 °C with shaking for 2 to 3 h to allow infection and lysis of the growing culture. The donor culture showing most clearing compared to the corresponding control was thoroughly mixed with 100 μl of chloroform, and dispensed into micro-centrifuge tubes. Cell debris was removed from the phage suspension by centrifugation at 10,000 rpm for 10 min in a desktop centrifuge. 20 μl chloroform was mixed with the resulting phage suspension decanted in fresh tube, and stored at 4 °C.

2.18.2 Determination of P1 Titre
A P1-sensitive strain of E.coli was grown to mid-exponential phase (OD600nm = 0.4 - 0.6) at 37 °C in LB supplemented with 5 mM CaCl2. A series of 10-fold dilutions of P1 in LB broth supplemented with 5 mM CaCl2 was set up, and 100 μl aliquots of each dilution were mixed with 100 μl of bacteria and incubated at 37 °C for 20 min. Soft-top agar (10% w/v tryptone, 10% w/v NaCl, 5% w/v yeast extract, 0.75% w/v agar) was then added to each dilution and the entire mixture poured onto LB agar plates containing 10 mM CaCl2. The plates were incubated overnight at 37 °C and the number of plaques on each plate was counted to determine the phage titre in terms of plaque-forming units per ml (pfu/ml). If the phage titre was low, a second round of lysis was performed on the donor strain and the titre re-determined.

2.18.3 Transduction
An overnight culture of the recipient strain was diluted 1:100 in fresh LB supplemented with 2.5 mM CaCl2 and antibiotics as appropriated, and grown to an OD600 ≥ 0.8. Two 1 ml aliquots of the recipient strain were transferred to 15 ml Falcon tubes. One aliquot was incubated with 100 μl of P1 lysate generated from the donor strain whilst the other served as a negative control, and both tubes were incubated at 30 °C for 30 min without shaking. The cultures were then mixed with 4 ml LB supplemented with 5 mM sodium citrate, and incubated at 30 °C for 1h with shaking before being collected by centrifugation at 3000 g for 10 min. The
cells were resuspended in 100 μl LB supplemented with 5 mM sodium citrate, and plated on LB agar with appropriate selection. Following overnight incubation at 37°C, any resulting transductants were subjected to two single-colony isolations on selective plates before being screened for the appropriate mutation by colony PCR.

2.19 Protein Purification and Analysis

2.19.1 IHF over-expression and collection
This protocol was performed according to the method of Nash et al. (1987) and Filutowicz et al. (1994). The strain K5746 containing plasmid pP_liphimA-5 (table 2.2) was inoculated into 50 ml LB broth with Ampicillin and incubated at 30°C with shaking overnight. Overnight culture was then diluted 1:100 into flask containing 500 ml LB Broth supplemented with Ampicillin and grown until to mid-log phase. Once the culture were grown to the desired OD value (OD_{650nm}=0.9), the cells were immediately shifted to a 42°C water bath and shaking was continued 2 hours to achieve overexpression of IHF. Cells were collected by centrifugation at 8,000 g in a Sorvall RC5BPLUS swing-arm rotor for 20 min and keep the pellet at -20°C.

2.19.2 Sonication cell lysis
After being thawed and resuspended in 25 ml Buffer A (25 mM Tris.Cl, 50 mM NaCl, 1 mM EDTA, pH=7.4) supplemented with 250 μl of 20 mg/ml Protease Inhibitor in DMSO and 25 μl of 1 M DTT. The suspension was aliquoted into 2 × 50 ml falcon tubes and added 50 μl of 5 mg/ml DNase I to each falcon tube. The cells were disrupted with six bursts of 3 min of sonication, with 90 s between each burst. During the sonication, the sample must be kept cooled. Keep the samples of sonic extract on ice for 30 min and the cell debris was removed by centrifugation at 19,000 rpm for 35 min at 4°C. The supernatant was decanted and stored on ice.

2.19.3 Ammonium sulfate precipitation and dialysis
After centrifugation, it was yield 29 ml supernatant of clarified extract. The supernatant was adjust to ~60% saturation by the addition of 10 g ammonium sulfate, keep constantly stirring over 20 min and centrifuge for 20 min at 19,000 rpm at 4°C. The resulting pellet was dissolved in around 9 ml Buffer A with 10 μl of
1 M DTT to obtain 10 ml suspension and dialyzed by 5000 MWCO (Spectrum Medical Industries) dialysis tube against Buffer A (with DTT) at 4°C overnight. Kept 200 µl dialysis for later analysis and the rest of overnight dialysis needs to filter by passing through 0.22 µm filter and also kept 200 µl dialyzed solution for later analysis.

2.19.4 Affinity chromatography purification
IHF was purified using a HiTrapTM Heparin HP 1ml column (GE Healthcare) according to the manufacturer’s instructions. After equilibrating the column with 5 ml degassed washing buffer (Buffer A), the clarified cell-free extract was loaded and the column was washed with 5 ml washing buffer A. Protein was eluted by adding 1ml of the elution buffer- Buffer B (25 mM Tris-Cl, 1.5 M NaCl, 1 mM EDTA, pH=7.4) each with increasing salt concentration consecutively.

2.19.5 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)
The Mini Protean 3 system (Bio-Rad) was used to set up denaturing polyacrylamide gels. The gel was composed of a 5% stacking gel and 15% resolving gel shown as in Table 2.4 below. The resolving gel and stacking gels were thoroughly mixed with N,N,N',N', tetramethylethylenediamine (TEMED) to catalyse the polymerization of the acrylamide gel, respectively, before poured into the glass plates. The propan-2-ol was added after the resolving gel poured to the glass plates and then rinsed off with sterile distilled water. Then the stacking gel was adding on the top of resolving gel. Once set, the stacking gel was washed with sterile distilled water before the entire gel was submerged into SDS electrophoresis buffer (Table 2.4). 10 µl protein samples were incubated for 5 min at 95°C in 10 µl 2 × SDS loading buffer (Table 2.4). Samples were loaded onto the gel alongside molecular weight markers (Bio-Rad). Samples were pre-electrophoresis at 80 V for 10 min before being electrophoresis at 180 V until they has migrated to the end of the gel. SDS-PAGE gels were visualized by staining in Coomassie Instant Blue (Expedeon) at room temperature for 15 min followed by destaining with sterile distilled water for several times. For the XCell SureLock Mini-Cell system (Life Technologies), the preset NuPAGE 10% Bis-Tris Mini Gels was used for protein analysis. Protein
samples were mixed with 5% of the reducing agent (2-mercaptoethanol) and 4 × LDS sample Buffer (see below), then incubated at 95 °C for 10 min. The appropriate concentration of proteins sample were loaded on the gel and electrophoresis in the NuPAGE MES running buffer at 200 V for 35 min. Protein sizes were estimated by comparison with molecular weight.

2.19.6 Protein Quantification

The BioRad protein assay was performed to determine protein concentration according to the manufacturer’s instruction. Biovine serum albumin (BSA) was used as a calibration standard and results were determined by measuring absorbance at 595 nm in a spectrophotometer.

<table>
<thead>
<tr>
<th>Component</th>
<th>2 × SDS Loading buffer</th>
<th>4 × LDS sample Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>91 Mm Tris pH 6.8</td>
<td>106 mM Tris HCl pH 8.5</td>
</tr>
<tr>
<td></td>
<td>0.2% (w/v) SDS</td>
<td>141 mM Tris Base</td>
</tr>
<tr>
<td></td>
<td>0.2% (v/v) Glycerol</td>
<td>2% LDS</td>
</tr>
<tr>
<td></td>
<td>0.2% (w/v) Bromophenol blue</td>
<td>10% Glycerol</td>
</tr>
<tr>
<td></td>
<td>7% (v/v) 2-mercaptoethanol (BME)</td>
<td>0.51 mM EDTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.22 mM SERVA Blue G250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.175 mM Phenol Red</td>
</tr>
</tbody>
</table>

**Table 2.4 Constituents of polyacrylamide gel**

<table>
<thead>
<tr>
<th>Component</th>
<th>Tris-glycine SDS-PAGE</th>
<th>10% resolving gel (ml)</th>
<th>15% resolving gel (ml)</th>
<th>5% continuous gel (ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30% acrylamide (29:1 acrylamide:bis) mQH2O</td>
<td>0.17</td>
<td>1.7</td>
<td>2.5</td>
<td>0.83</td>
</tr>
<tr>
<td>1 M Tris (pH6.8)</td>
<td>0.68</td>
<td>1.9</td>
<td>1.1</td>
<td>2.92</td>
</tr>
<tr>
<td>1.5 M Tris (pH8.8)</td>
<td>0.13</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10% SDS</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>5×TBE</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>50% glycerol</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10% APS</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>TEMED</td>
<td>0.001</td>
<td>0.002</td>
<td>0.002</td>
<td>0.005</td>
</tr>
</tbody>
</table>
2.20 Electrophoretic Mobility Shift Assay (EMSA)
Competitor DNA and the promoter regions fragment were amplified from pBluescript or the promoter constructs with primers designed to anneal to the plasmid backbone either side of the cloning site. For each construct, PCR amplification products were checked by agarose gel electrophoresis and purified by MiniElute PCR purification kit (Qiagen). For IHF EMSA assay, equal quantities (25 ng) of the target and competitor DNA were mixed and added to binding reactions containing reaction buffer (5×Reaction binding buffer: 10 nM TrisCl pH=7.8; 0.5 mM DTT; 1.25 % (v/v) glycerol; 25 µM spermidine) and varying concentrations of purified IHF which had been pre-equilibrated for 10 min. For H-NS EMSA assay, equal quantities (20 ng) of the DNA mixed to varied concentrations of protein H-NS to binding reactions containing reaction buffer (5 × Reaction binding buffer: 50 nM TrisCl pH=7.8, 250 mM KCl, and 0.5 % Triton X-100) in a final volume of 10 µl.

Following incubation at 37 °C for 30 min, the entire reaction volume (10 µl) was resolved on 5% non-denaturing Tris-borate-EDTA (TBE: 446 mM Tris; 445 mM boric acid; 10 mM EDTA, pH=8) polyacrylamide gels (Table 2.4) made using the Bio-Rad Mini-Protein 3 system. Polyacrylamide gels underwent preliminary electrophoresis in 1×TBE at 90 V for 90 min before samples loading. Samples were run into the gel at 250 V for 3 min before being separated at 120 V for 60 min. One lane containing DNA loading buffer was used as an indicator for migration of DNA fragments through the gel. Following electrophoretic fractionation of samples, gels were stained with 0.5 µg/ml Ethidium Bromide for 30 min and visualized under UV light.

2.21 Quantitative Real-time Polymerase Chain Reaction (qRTPCR)
Total RNA was prepared from 1 ml of OD_{600nm}=0.5 samples of cell cultures. RNAprotectTM bacteria reagent and RNeasy mini kits, with off-column DNase treatment, were used according to the manufacturer’s protocols (Qiagen) described as above in section 2.13. Specific designed primers were used to amplify around 100 – 200 bp fragment from the variable region of target gene. Reverse
transcription was performed using QuantiTect Reverse Transcription Kit for fast cDNA synthesis enabling sensitive real-time two step RT-PCR for gene expression analysis (Qiagen) with 0.5 – 1 µg total RNA. Quantitative polymerase chain reaction (QPCR) was performed on an ABI Prism sequence detector (Applied Biosystems) with 20 µl FAST SYBR Green Master Mix (Applied Biosystems), 2 µl of 0.25 µM primers mix (Forward and Reverse) and 20 µl of 1:5 or 1:10 diluted reverse transcribed products for 4 × 10 µl reactions. The thermal cycling conditions were as follows: A 20s holding step at 95 °C was followed by 40 cycles of a 1s denaturation step at 95 °C and a 30s annealing/polymerisation step at 60 °C. For each set of primer, a standard 1/10 series dilution (highest con. is 100 pM) of a target amplified DNA fragment was performed for quantification. A linear relationship was obtained by plotting the threshold cycle against the logarithm of a known amount of copy numbers. The quantities of target copies contained in an unknown sample were determined by extrapolation from the linear regression of the standard curve. Calculate the copy number per 1 µg total RNA in was shown as below.

\[
\text{Copy Number/µg RNA} = X \times 10^{-15} \times M \times 2 \times V_0 \times L \times N_A
\]

Where the X is the reading of X-axis from standard curve by plotting the threshold cycle against the logarithm; \(V_0\) = volume of diluted reverse transcription products (cDNA) from 1 µg total RNA and \(N_A\) = 6.02 × 10^{23} copy number/mol. A negative control (just RNA) that had not undergone the reverse transcription step was also included in each run. For each sample’s cDNA copy number quantification, the number of copies was determined in triplicate and the mean value was used for analysis. In addition, each set of primers were also checked every time to make sure no primer-dimers occurred and no other DNA contamination in reaction by carrying out melting curve stage after the total amplification cycle (rapid heating up to 95 °C for 15s to denature the DNA, followed by cooling to 60 °C for 1min and increasing 0.3°C/s to 95 °C for 15s).
2.22 Immunofluorescent staining and imaging

For visualization of bacteria, strain UTI89 was grown to mid-log and 500 µl of LB-broth cultures were pelleted by centrifuging at 13000 rpm for 5 min. If take cells from different time points of growth curve, the early time point pellet can be snap freeze in liquid nitrogen and keep at - 80 °C for the later analysis. The thawed pellet was resuspended in 500 µl PBST buffer (0.05 % Tween 20 in PBS buffer) and incubated with 500 µl 4 % paraformaldehyde (PFA) at room temperature for 10 min followed by 50 min on ice. The cells were washed two times in PBST (0.05 % Tween 20 in PBS buffer) at room temperature by centrifugation and resuspended in 500 µl of PBST. Loaded 10 µl of the cells suspension on to 0.1 % Poly-L-Lysine-HB treated slides and air-dried at room temperature. Specific primary and secondary antibodies were diluted depending on the requirement of each antibody in 1 mg/ml Bovine serum albumin (BSA) in PBS right before labelling. Samples were incubated with primary antibodies at RT for 20 min, washed 10 times with PBST, followed by incubation with mixture of secondary antibodies and 0.5 µg/µl DAPI (Sigma) for 20 min and 10 times washed with PBST. The slides were then mounted in mounting agent Slow Fade ® Gold Antifade Reagent (Life Technologies), adding a coverslip and sealing the slide in nail varnish. Slides were visualized using wide-field fluorescent microscope. Images were collected on a [Olympus BX51] upright microscope using a [10x/ 0.30 Plan Fln] objective and captured using a [Coolsnap ES camera (Photometrics)] through [MetaVue Software (Molecular Devices)]. Specific band pass filter sets for [DAPI, FITC and Texas red] were used to prevent bleed through from one channel to the next. Images were then processed and analysed using ImageJ (http://rsb.info.nih.gov/ij).

The PFA solution was prepared by dissolving 6 g PFA in 160 ml of PBS at 80 °C, followed by addition of 200 µl 0.1 M CaCl₂ and 200 µl 0.1 M MgCl₂ while the solution is warm, and adjusted to pH=7.4 with a final volume of 200 ml. Aliquots of the PFA solution were stored at 20 °C and kept at 4 °C for up to 1 week.
Chapter 3. Identification and characterization of multiple promoters at promoter 1 region (PR1) of the E. coli K5 antigen gene cluster

3.1 Three additional promoters were determined in the 5’ Untranslated Region (5’ UTR) at PR1

3.1.1 Introduction

Previously three additional transcriptional start sites had been identified in the 5’ UTR of the PR1 promoter in the K1 capsule gene cluster (Cieslewicz & Vimr, 1996). Since both the K1 and K5 are Group 2 capsules and share 98.7% similarity in 5’ UTR region, it is reasonable to assume that there will also be additional transcription start sites in the PR1 promoter in the 5’ UTR of K5 capsule gene cluster. In this chapter, a variety of reporter gene fusions and 5’ RACE analyses were performed in order to determine whether there are other putative promoters in the 5’ UTR region of PR1 in K5 capsule gene cluster.

3.1.2 Generating E. coli K5 PR1 promoter region, 5’ UTR region and upstream promoter region lacZ fusions

In order to study PR1 promoter of the E. coli K5 capsule gene cluster, four different PR1 lacZ transcriptional fusions were generated. To generate a lacZ transcriptional fusion of PR1, the 873 bp intact PR1 region from -645 to +228, including the first ATG of kpsF gene was released from BamH1 digested pHA2, and ligated into lacZ transcriptional fusion vector pRS415 (Figure 3.1). Colony PCR using primers pJJ2-F-EcoR1 and pRS415-lacZ-SEQ (Table 2.3) were used to identify clones containing the 1063 bp amplicon. This plasmid was called pJPR1.

A second PR1 lacZ transcriptional fusion was made by amplifying 869 bp PR1 region from -645 to +224 to exclude the first ATG of kpsF gene with primers pJPR2-F and pJPR2-R-BamH1 (Table 2.3) using pHA2 as template (Figure 3.1). The
amplified fragment was digested with BamH1 and ligated into the vector pRS415. The successful clones were identified by colony PCR using primers pJJ1-F-EcoR1 and pRS415-lacZ-SEQ with 1060 bp amplicon. The identified plasmid was termed pJPR2.

In order to assess the level of transcription driven from the 5’ UTR region, the intact PR1 region was separated into 5’ UTR region and upstream promoter region. A 218 bp 5’ UTR fragment from +1 to +218 (F1) was generated from pHA2 by PCR amplification using primers pJJ1-F-EcoR1 and pJJ1-R-BamH1 (Figure 3.1). A 660 bp fragment containing the upstream promoter binding region from -645 to +15 (F2) was amplified by PCR amplification from the plasmid pHA2 using primers pJJ2-F-EcoR1 and pJJ2-R-BamH1 (Figure 3.1). Fragments F1 and F2 were then digested with BamH1 and EcoR1 simultaneously and ligated into vector pRS415, respectively. The successful clones containing the 5’ UTR (F1) and the upstream promoter binding region (F2) were identified by colony PCR using two sets of primers, pJJ1-F-EcoR1/ pRS415-lacZ-SEQ and pJJ2-F-EcoR1/pRS415-lacZ-SEQ (Table 2.3), respectively. The successful F1 construct had 218 bp amplified product and was named pJJ1, while the F2 construct had 660 bp amplified product and named pJJ2 (Figure 3.1 A). Plasmid DNA was purified from the selected transformants and the presence of the inserted DNA fragments confirmed by restriction enzyme digestion and agarose gel electrophoresis. The plasmid pJJ1 when digested by the enzymes EcoR1 and EcoRV released two fragments of 9500 bp and 1473 bp (Figure 3.2). When plasmid pJJ2 was digested with EcoR1 and BamH1 two fragments of 10751 bp and 655 bp were generated (Figure 3.2).

The nucleotide sequence of the cloned fragments in plasmids pJPR1, pJPR2, pJJ1 and pJJ2 were determined by nucleotide sequencing using primer pRS415-lacZ-SEQ and P1-internal-SEQ (Table 2.3).
Figure 3.1 Generation of PR1 transcription fusion pJPR1, pJPR2, 5′ UTR region lacZ fusions pJJ1 and upstream promoter region lacZ transcriptional fusion pJJ2.

(A) Plasmid pJPR1 contained 873 bp of PR1 fragment. Plasmid pJPR2 contained 869 bp of PR1 fragment, excluded the first start codon ATG of kpsF. The 218 bp 5′ UTR region was amplified by primers pJJ1-F-EcoR1 and pJJ1-R-BamH1 and a 660 bp upstream promoter region was amplified by primers pJJ2-F-EcoR1 and pJJ2-R-Bam H1, respectively. The black arrow indicated the original transcription start site (+1) and the grey arrow indicated the minor transcription start sites (+133 ‘A’ and +183 ‘G’) in the 5′ UTR of E. coli K1 capsule PR1 mapped by Cleslewica and Vimer (1996).

(B) A simplified map of the multi-copy transcriptional vector pRS415. The amplified fragments F1 and F2 were ligated into vector pRS415 at the EcoR1/BamH1 sites.
3.1.3 The transcription from 5' UTR showed modest transcription activity at 37 °C

The transcriptional activity in each of the constructs was measured by performing β-galactosidase assays on mid-exponential cultures (OD600nm=0.4-0.6) grown at 37 °C and 20 °C, respectively. Plasmid pRS415 was the vector control and allowed the basal level of β-galactosidase activity that was generated due to the presence of the vector to be measured. All constructs were transformed into host strain P90C (Table 2.1). Strain P90C (pJPR1) or P90C (pJPR2) displayed significant high levels of activity (Figure 3.3). As predicted, both of them were temperature regulated with ~2000 Miller Units at 37 °C (no significant difference) while approximately 10-fold reduced at 20 °C (Figure 3.3). According to the previous study by Cieslewicz and Vimr (1996) which demonstrated that there were additional transcription start sites located in the 5’ UTR of PR1 in the E. coli K1 capsule gene cluster, it was not surprising that the transcription from 5’ UTR region in P90C (pJJ1) presented modest transcription activity at 37 °C (204.3±30.8 Miller Units). But unexpectedly, the transcription in P90C (pJJ1) was still temperature regulated with significant lower activity at 20 °C (Figure 3.3). Moreover, the 5’ UTR in P90C (pJJ1) exhibited much lower transcriptional activity with approximately 9-fold lower at 37 °C and 10-fold decreased at 20 °C in comparison with the transcriptional activity driven from the
PR1 region in P90C (pJPR2) (Figure 3.3). Strain P90C (pJJ2) that lacks the UTR, displayed extremely high level of transcription activity and lost temperature regulation (Figure 3.3). These data indicated that there are additional temperature regulated promoters in the 5’ UTR region of PR1 in the K5 capsule gene cluster and the 5’ UTR region seems play an important role for the transcriptional thermoregulation at PR1.

![Figure 3.3 Beta-galactosidase activities of PR1 transcriptional fusions pJPR1, pJPR2, pJJ1 and pJJ2 grown at 37 °C and 20 °C.](image)

**Figure 3.3** β-galactosidase activities of PR1 transcriptional fusions pJPR1, pJPR2, pJJ1 and pJJ2 grown at 37 °C and 20 °C.

β-galactosidase activities generated by different PR1 transcriptional fusions in strain P90C at 37 °C and 20 °C. Plasmid pRS415 serves as negative control. Values are the means of three independent experiments performed in triplicate. Error bars represent standard error.
3.1.4 Determination of the potential transcription start sites in the 5' UTR of Region 1 Promoter by 5' RACE assay

The β-galactosidase activity of strain P90C (pJJ1) at 37 °C (Figure 3.3) indicated that there are additional transcription start site(s) within the 5’ UTR; therefore it is meaningful to investigate the transcription start points in the UTR. In an attempt to map the potential transcription start sites in the 5’ UTR, the mRNA was extracted from the strains grown at 37 °C and analysed by 5’ RACE kit (Roche) to determine the 5’ end of mRNA (Methods 2.14). A work flow chart of mapping transcription starts sites can be seen in Figure 3.4. Taking as an example of 5’ RACE in strain P90C (pJJ1), the products of all the 5’ RACE reactions were analysed by gel electrophoresis (Figure 3.5). The mRNA transcribed from the putative transcription start sites was reversed transcribed into different sizes of cDNA and amplified into double strand DNA in 5’ RACE assay. As seen from Figure 3.5, the second round PCR reaction (lane 9) yielded multiple products. The multiple sizes of amplified fragment indicated that there might be more than one transcription start site in the 5’ UTR region. In order to analyse the sequences of multiple products for 5’ RACE reactions and try to distinguish sites of transcription initiation in PR1 and the 5’ UTR region, the purified second round PCR products were therefore ligated into pGEM®-T easy vector system. The ligated constructs were transformed into high efficiency competent cells DH5α. The recombinant constructs, as identified by blue/white screening on media supplemented with X-gal, were checked by colony PCR with primers T7 and SP6 (Table 2.3). A number of different sizes of inserts were visualized on the agarose gel (Figure 3.6). The sequence of the inserts was determined by DNA sequencing of the representative plasmids with primer SP6. All primers used in 5’ RACE can be seen in Table 2.3. The same procedure was repeated for identifying transcription start sites in other strains P90C (pJPR1), P90C (pHA2) and MS101 (Table 2.1, Table 2.2).
Figure 3.4 Schematic representation of the strategy used for mapping transcription starts sites by 5’RACE assay.
Figure 3.5 The 5’ RACE reaction using mRNA extracted from strain P90C (pJJ1) grown at 37 °C.

Lane 1 to 2, the control of first-strand cDNA synthesis. The band around 170 bp indicated the cDNA synthesis has been successful. Lane 3, 4, 6 and 8, the negative control showed no DNA contamination of RNA sample and the 5’ RACE reactions. Lane 5, the positive control showed the 218 bp 5’ UTR region amplified by primers pJJ1-F-EcoR1 and pJJ1-R-BamH1 using plasmid pJJ1 as template. Lane 7, the first round PCR products of 5’RACE reaction. Lane 9, the second round PCR products of 5’RACE reaction. M, Hyperladder IV DNA markers (bp), Agarose gel, 2% (w/v)

Figure 3.6 Colony PCR for DH5α bearing plasmid pGEM-T easy ligated with P90C (pJJ1) second round PCR products in 5’ RACE assay.

The DH5α recombinants were screened by Colony PCR using primer T7/SP6. Lane 1- 35, the colony PCR products. Lane 36, the blank negative control showed no DNA contamination. M, Hyperladder IV DNA markers (bp). Agarose gel, 1% (w/v).
By analysing the details of putative transcription start sites (TSSs) discovered in four different strains P90C (pJPR1), P90C (pJJ1), MS101 and P90C (pHA2) by 5’ RACE assay (Figure 3.7 and Table 3.1). It was noticed that the 5’ end of transcripts (represented by stars in Figure 3.7) obtained from all sequenced samples in these four strains were clustered at position +1, +133, +142, +182 and the region between +133 and +142 (Figure 3.7).

In the strain P90C (pJPR1) containing the whole PR1 region, there were at least one sample (out of 17 sequenced samples) showing the transcripts exactly starting at the positions +1, +133, +142 and +182 (Figure 3.7; Table 3.1). Combining the Cleslewica and Vimers’ (1996) primer extension result (Figure 3.8) of *E. coli* K1 PR1 region and the 5’ RACE results of strain P90C (pJPR1) obtained in this study (Figure 3.7, Table 3.1), it was confirmed that the additional existence of TSSs at position +133 and +182 regarding to ‘A’ and ‘G’ residues, respectively. The identification of the TSS at +182 seemed more accurate compared with the TTS at +183 mapped by Cleslewica and Vimers (1996) using primer extension assay, since it was difficult to be unequivocal in the assignment of the first nucleotide from their study (Figure 3.8). In addition, the 5’ RACE results obtained in this study strongly suggesting there was another TSS at position +142, although this was not previously detected by the study done by Cleslewica and Vimers (1996).

To study this further, 5’ RACE analysis was performed on the 5’ UTR transcriptional fusion pJJ1, PR1 translational fusion pHA2 and *E. coli* K5 strain MS101 (Figure 3.7 and Table 3.1). Not all of the TSSs could be detected in all of the strains and this perhaps reflects the number of samples sequenced (Table 3.1). However, overall all of the TSSs were detected among these three additional strains. Analysis of all these data indicated that those sites (+133, +142 and +182) identified by 5’ RACE in the UTR may be the potential transcription start sites of the *E. coli* K5 PR1. All the DNA sequencing data for the 5’ RACE assay in strains P90C (pJPR1), P90C (pJJ1), MS102 and P90C (pHA2) are presented in Appendix I.
Figure 3.7 The detail of putative transcription starts sites analyzed by 5’ RACE.

The nucleotide sequence from -95 to +228 relative to the Region 1 promoter of *E. coli* K5 capsule gene cluster is shown above. The start codon of *kpsF* is indicated by black bold type. The original transcription start point is indicated by red bent arrow and the promoter binding sites proposed by Simpson *et al.* (1996) are boxed. Black bent arrows indicate the minor transcription start sites on the 5’ UTR of *E. coli* K1 capsule Promoter 1 region (Cleslewica and Vimer, 1996). Grey bent arrows indicate the additional minor transcription start sites on the 5’ UTR observed in this study. Black star indicate the putative transcription start sites on the 5’ UTR of *E. coli* K5 capsule Promoter 1 region which investigated by P90C (pJJ1) by 5’ RACE assay; Red stars indicate 5’ end of RNA mapped in strain P90C (pHA2) by 5’ RACE assay; Blue stars indicate 5’ end of RNA mapped in strain MS101 by 5’ RACE assay; Yellow stars indicate 5’ end of RNA mapped in strain P90C (pJPR1) by 5’ RACE assay. Positions in bold are those most commonly found for a given 5’ end of mRNA.

<table>
<thead>
<tr>
<th>TSS position</th>
<th>+1</th>
<th>+133</th>
<th>+142</th>
<th>+182</th>
<th>Total sequenced sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>P90C (pJPR1)</td>
<td>(3)</td>
<td>(2)</td>
<td>(2)</td>
<td>(1)</td>
<td>17</td>
</tr>
<tr>
<td>P90C (pJJ1)</td>
<td>(0)</td>
<td>(0)</td>
<td>(3)</td>
<td>(1)</td>
<td>8</td>
</tr>
<tr>
<td>MS101</td>
<td>(1)</td>
<td>(0)</td>
<td>(1)</td>
<td>(0)</td>
<td>8</td>
</tr>
<tr>
<td>P90C (pHA2)</td>
<td>(2)</td>
<td>(0)</td>
<td>(0)</td>
<td>(1)</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 3.1 Summarization of detected transcription start sites from different constructs used in 5’ RACE assay.

Bold numbers indicated the most commonly found putative transcription start sites (TSSs) in 5’ RACE. The number in parentheses indicated the number of cloned RACE products 5’ end of mRNA corresponding to the given site obtained.
In addition, the modified 5' RACE (Method 2.15) was also carried out in K5 strain MS101 to confirm the observation of potential transcription start sites based on the original 5' RACE assay. In this modified 5' RACE assay, the 5' end of SP3 primer was modified by fluorescent dye 6-FAM (Sigma). Therefore, the obtained PCR products labelled by 6-FAM then can be directly sequenced for the length of all amplified products of the second round PCR of 5' RACE. The length of amplified transcripts from the putative transcription start sites with forward poly-T anchor primer (39 bp) and reverse SP2-MS101-6FAM (Table 2.3) could be calculated out by adding the length of predicated amplified length starting from the corresponding transcription start sites. For instance, if the transcription started at position +1, the predicted transcript starting at +1 and end in the 3' of the reverse primer would transcribe into 295 bp length of mRNA, and adding 39 bp length of forward poly-T anchor primer 5' to the transcription start sites would result in transcript with 334 base pairs. The other transcript lengths can be similarly calculated as above, TTS at +133, +142 and +182 would amplify around 200 bp, 190 bp and 150 bp, respectively. In the output data of DNA sequencing, the fluorescently labelled DNA fragments were distinguished according to molecular
weight. The proportions of different size fragments were calculated out and plotted into Figure 3.9. It was shown that abundant different sizes of DNA fragments ranging from 100 bp to 340 bp were obtained and four significant peaks (proportion ≥0.2) were present at 150, 190, 200 and 330 bp. This would support the prediction of transcripts from the position +182, +142, +133 and +1, respectively. This information redefines the task of identifying the putative transcription start sites as a matter of implementing a method able to present transcription start signals within the PR1 region.

![Figure 3.9](image.png)

Figure 3.9 Portions of different sizes of amplified DNA fragments with 6-FAM modified SP2 primer in modified 5’ RACE assay.

Each point indicated the proportion of corresponding size of amplified fragments.

Based on the observation described above, I predict that there are three tandem promoters in 5’ UTR region initiating at +133, +142 and +182 (denoted PR1-2, PR1-3 and PR1-4, respectively) in addition to the previously identified promoter (hereafter denoted PR1-1) are present in the promoter region of region 1 operon of *E. coli* K5 capsule gene cluster. To test whether the multiple promoters may
have functional role for the total transcription from Region 1, therefore, the transcripts driven from different promoters need to be measured.

3.2 Evaluating the relative transcriptional activities of multiple promoters

Simpson et al. (1996) had already demonstrated that the PR1 promoter is $\sigma^{70}$ dependent and there was no other $\sigma$ factor polymerases involved in transcription initiation in the PR1 region. Thus, in order to further characterize the observed transcription start sites and to evaluate the relative transcriptional activities, the putative promoters regions were amplified at least 40 bp upstream of the putative transcription start sites that would provide enough upstream sequence for $\sigma^{70}$ RNA polymerase binding. In this section, the amplified minimal putative promoter regions were cloned into lacZ transcriptional vector and the relative transcriptional activity of each promoter measured by performing a $\beta$-galactosidase assay and qRTPCR analysis.

3.2.1 Construction of E. coli K5 region 1 minimal promoter lacZ fusions

In order to evaluate relative activities of the tandem Region 1 promoters, different promoter-lacZ fusions were generated by using the same method of constructing plasmid pJJ1 described above. Briefly, the target fragments were amplified by primers which forward primers modified by EcoR1 while reverse primer modified by BamH1 (Table 2.3) with plasmid pHA2 as template. The amplified fragments were digested by EcoR1/BamH1 before being ligated into vector pRS415 (Figure 3.1B). The successful transformants were screened by colony PCR and the constructs were checked by DNA sequencing. With respect of TSS $+133$, pJJ133 was generated by amplifying promoter PR1-2 region from $+1$ to $+140$ (Figure 3.10). Since the putative promoter region of $+133$ and $+142$ located overlapped leading to the difficulty of separating PR1-3 out of the promoter context, therefore, the plasmid pJJ133-142 which containing both promoters PR1-2 and PR1-3 together was made with amplifying the region from $+1$ to $+146$ (Figure 3.10). Regarding to the minimal PR1-4 promoter construction, two constructs (pJJ182 and pJJ182A)
only containing the putative PR1-4 promoter region were made (Figure 3.10). Plasmid pJJ182 contained the region from +146 to +218 while pJJ182A contained the upstream 53 bp of an IHF binding site from +93 to +218 (Figure 3.10).

Figure 3.10 Construction map of transcriptional lacZ fusion inserts.

The insert of pJPR1 was cut from pHA2 by BamH1 digestion. The insert of pJPR2 was amplified by PCR from pHA2 with primers both modified by BamH1 sites. All the other cloned fragments were amplified from template pHA2 by PCR and inserted into EcoR1-BamH1 sites of pRS415.
3.2.2 Transcriptional fusions confirm putative promoter functionality

The transcriptional fusions with minimal putative promoters were transformed into strain P90C and β-galactosidase assays were performed at 37 °C and 20 °C, respectively (Figure 3.11). A basal level of β-galactosidase activity was defined by the promoterless plasmid pRS415 and averaged around 1 Miller Unit either at 37 °C or 20 °C. As previously described, plasmid pJPR2 containing the whole PR1 region demonstrated temperature dependent expression, it displayed high level of β-galactosidase activity at 37 °C (1834.2 ± 286.70 Miller Units) and significantly reduced β-galactosidase activity at 20 °C (347.0 ± 72.2 Miller Units) (Figure 3.11). Strain P90C (pJJ2) lost temperature regulation with extremely high level of transcription activity both at 37 °C and 20 °C. The temperature regulated fusion pJJ1 containing the three putative promoters gave 237.2 ± 31.77 Miller Units at 37 °C and low β-galactosidase activity at 20 °C (32.0 ± 1.3 Miller Units) (Figure 3.11). The construct pJJ133 contained the single putative promoter PR1-2 showed very low activity (13.7 ± 3.0 Miller Units) at 37 °C while no activity (0.9 ± 0.04 Miller Units) at 20 °C, which indicated that PR1-2 is a functional temperature regulated promoter but with relatively low activity. In contrast, strain P90C (pJJ133-142) had 4.6-fold increased β-galactosidase activity as compared to strain P90C (pJJ133) at 37 °C (Figure 3.11). An explanation for this was the extra transcription activity contributed from the putative promoter PR1-3. Thus it was inferred that the promoter PR1-3 was also a functional promoter in the PR1 region. When the promoter PR1-4 minimal lacZ fusion P90C (pJJ182) was assayed, it lost temperature regulation and expressed relatively high β-galactosidase activity at 37 °C (504.5 ± 59.2 Miller Units) even 2-fold higher than the strain P90C (pJJ1) containing whole UTR region (Figure 3.11). The relatively higher activity generated from strain P90C (pJJ182) indicated that the putative promoter PR1-4 was a functional active promoter with higher activity than either PR1-2 or PR1-3. However, when the strain P90C (pJJ182A) bearing the construct containing the IHF consensus binding region was assayed, it had approximately 4-fold reduced
expression compared to pJJ182 (Figure 3.11). In addition, the plasmid pJJ182A with extra upstream region of promoter PR1-4 restored transcriptional temperature regulated for promoter PR1-4 (Figure 3.11). These data indicated that the transcription activity of promoter PR1-4 was inhibited by binding of IHF and that the binding of IHF may play a role in regulating the transcription from PR1-4.

Figure 3.11 β-galactosidase activities of different minimal putative promoter-lacZ fusions grown at 37 °C and 20 °C.

β-galactosidase activities generated by different minimal PR1 transcriptional fusions in strain P90C at 37 °C and 20 °C. Values are the means of three independent experiments performed in triplicate. Error bars represent standard error.

Taken together, the previously observed promoter PR1-1 was the strongest among the tandem region 1 promoters. The minimal promoter-lacZ fusion analysis confirmed that the 5’ UTR region contained three functional tandem promoters. Notably, two overlapping promoters PR1-2 and PR1-3 were much less active than promoter PR1-4. However, one cannot rule out the possibility that the low level of transcription activity might be due to some artefact effects when the small
fragment was removed from the intact promoter region. Therefore, carrying out experiment that can identify promoter motifs of the putative promoters seemed to be necessary.

3.3 Identification of functional elements of PR1 region tandem promoters

It is known that many *E. coli* promoters are completely inactivated by base changes at position 2 of the -10 hexamer (Feklistov & Darst 2012; Saecker et al. 2011). Therefore, based on the conserved $\sigma^{70}$ polymerase binding region, it can be feasible to further verify the existence of the proposed DNA target of RNA polymerase by mutating the promoter -10 hexamer.

3.3.1 Generating single or multiple sites mutation in the predicted motif of putative promoters

In order to identify sequence elements essential for the promoter, Site-Direct Mutagenesis PCR was used to generate PR1 promoter region with single or multiple nucleotide substitutions. Prior to generating mutations in the PR1 region, a new construct pBHA1 was made (Table 2.2). Plasmid pBHA1 contained the PRI region from -645 to +228 and was generated from pHA2 by BamH1 digested and cloning of this fragment onto the vector pSK+ (Table 2.2). Successful constructs were confirmed by sequencing with the T7 primer (Table 2.3). Site direct mutagenesis was performed as described previously (Method 2.17) and the whole plasmid pBHA1 was amplified up with a pair of complementary primers with mutated sites. All primers used for Site-direct mutagenesis can be seen in Table 2.3. The amplified plasmid was checked by agarose gel electrophoresis and immediately transformed into DH5α after treatment with enzyme *DpnI* (Methods 2.17). Following DNA sequencing to confirm mutagenesis, successfully mutated fragments were then cloned into transcriptional *lacZ* fusion pRS415 at the *BamH1* sites. The correct orientation of successful resulting recombinant plasmid was checked by colony PCR using the primers pJJ2-F-EcoR1 and pRS415-lacZ-SEQ. The
schematic representation of strategy used to generate mutation in region 1 promoters was shown on Figure 3.12.

Figure 3.12 Schematic representation of constructing Site-Direct Mutagenesis plasmids in PR1 region.
3.3.2 Mutation of the second T of PR1-1 promoter -10 Pribnow box could abolished 95% of transcription activity

To disrupt the PR1-1 promoter, the second position of the -10 hexamer was mutated. The PR1 whole region from -645 to +228 containing a single mutated site with second T to C substitution at -10 hexamer of PR1-1 (Table 2.2) was generated by SDM assay with primer SDMPR1-1_F and SDMPR1-1_R (Table 2.3) and cloned into pRS415, which was designated pJPR1-Δ+1. In order to test the inactivation occurred at PR1-1 promoter region only, promoter region from -645 to +15 with respect to TSS +1 was amplified from pJPR1-Δ+1 by using primer pJJ2-F-EcoR1 and pJJ2-R-BamH1 (Table 2.3) and cloned into pRS415, which was designated pJJ2-Δ+1. β-galactosidase assays were then performed to quantify the activity of PR1-1 in mutants by measuring the effect of the substitution on β-galactosidase expression at 37 °C. When the PR1 region carried the -12 T>C mutation in the PR1-1 -10 element in pJPR1-Δ+1, this reduced expression by over 95% compared to wild type strain P90C (pJPR1) (Figure 3.13). Furthermore, this mutation in strain P90C (pJJ2-Δ+1) decreased by 99% β-galactosidase activity compared to strain P90C (pJJ2). These observations were indicated that firstly, substitution at the second base pair of -10 hexamer of promoter could abolish the promoter activity significantly (99% reduced); secondly, transcription from PR1-1 may account for the majority (around 95%) of the total transcripts driven from PR1 region into kpsF gene.
Figure 3.13 Site direct mutation analysis of the promoter PR1-1 with second T to C substitution at -10 hexamer.

Cultures were grown at 37 °C in LB medium to mid-log phase. Values are the means of three independent experiments performed in triplicate. Error bars represent standard error.

3.3.3 The transcription activity with 165A>C and 138G>A mutation showed significant reduction at PR1-4 respectively

Based on the observation in Section 3.3.2, mutation at the second base pair of -10 hexamer was also performed to further analyse the predicted function elements of additional putative promoters. The promoter -10 and -35 hexamer were tentatively assigned shown in the Figure 3.14. As it shown, the putative -35 and -10 elements of the PR1-4 promoter, TTGTTA and TAAATT, are separated by suboptimal 22 bp spacer and deviate from the consensus -35 (5’-TTGACA-3’) and -10 (5’-TATAAT-3’) elements in 8 out of 12 positions (Figure 3.14; Lisser and Margalit, 1993). The putative -35 and -10 hexamers of the PR1-2 promoter, TAAGCA and TATATT, are also separated by a 17 bp spacer and deviate from the consensus in 8 out of 12 positions (Figure 3.14). And in the same manner, the putative -35 and -10 hexamers of the PR1-3 promoter, CTGACC and TAAAAA, are
separated by a 16 bp spacer and deviate from the consensus in 8 out of 12 positions (Figure 3.14).

Of the three putative promoters identified, the activity of overlapping promoters PR1-2 and PR1-3 exhibited extremely low activity and therefore I focused primarily on the promoter PR1-4. The measurements of the promoter activity of each mutation by β-galactosidase assay, together with controls, were showed in Table 3.2. As expected, when the second nucleotide (165A) of -10 hexamer of PR1-4 was substituted by C in pJPR1-182SDM2, the transcription activity was reduced from 3133.2±131.4 Miller Units to 2220.9±206.3 Miller Units. Whereas the measured promoter activity (3195.4±302.8 Miller Units) of the strain P90C (pJPR1-182SDM1) with 160T>C substitution of the second base pair of -10 hexamer of PR1-4 predicted by Cieslewicz and Vimr (1996) showed no significant difference compared to strain P90C (pJPR1), which indicated that the –10 hexamer predicted by Cieslewicz and Vimr (1996) is not correct. To prove our prediction further, pJPR1-182SDM3 carrying mutated -10 hexamer CCGATC instead of TAAATT was also measured by β-galactosidase assay, the significantly decreased promoter activity which was corresponding to the observation of pJPR1-182SDM2 as well. Furthermore, new construct pJJ182-SDM2 with 165A>C substitution was amplified out from pJPR1-182SDM2 using primer pJJ182-F-EcoR1 and PJJ1-R-BamH1 (Table 2.3) and generated as the same way as construct pJJ182 which only containing the PR1-4 promoter region from +146 to +218. The 60% reduction of PR1-4 promoter activity in pJPR1-182SDM2 compared to pJJ182 was confirmed the prediction of -10 hexamer element described above. Hence it was indicated the genuine -10 element of PR1-4 should be the TAAATT from the position +149 to +153.
Figure 3.14 Nucleotide sequence and map of the predicted promoter functional elements.

Nucleotides sequence of promoter region of region 1 from +1 to the first ATG of kpsF gene. -35 and -10 hexamer elements are underlined, and the transcription start site is indicated by a bent arrow. Mutations in the -35 and -10 promoter elements are indicated by vertical arrows. IHF consensus binding sequence is highlighted in green.

<table>
<thead>
<tr>
<th>Promoter PR1-4 SDM constructs</th>
<th>Promoter sequence from position +131 to 170</th>
<th>β-gal activity (Miller Units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pJPR1</td>
<td>AAATTTTGTTACACCCATTTGATTTGACATAAATAA</td>
<td>3133.2±131.4</td>
</tr>
<tr>
<td>pJPR1-182SDM1</td>
<td>AAATTTTGTTACACCCATTTGATTTGACACAATAAATTA</td>
<td>3195.4±302.8</td>
</tr>
<tr>
<td>pJPR1-182SDM2</td>
<td>AAATTTTGTTACACCCATTTGATTTGACAATTTAATTA</td>
<td>2220.9±206.3</td>
</tr>
<tr>
<td>pJPR1-182SDM3</td>
<td>AAATTTTGTTACACCCATTTGATTTGACAAATACG</td>
<td>2146.4±165.8</td>
</tr>
<tr>
<td>pJPR1-182SDM4</td>
<td>AAATTTTGTTACACCCATTTGATTTGACAAATACG</td>
<td>2005.1±17.3</td>
</tr>
<tr>
<td>pJJ182</td>
<td>AAATTTTGTTACACCCATTTGATTTGACAAATACG</td>
<td>568.8±29.1</td>
</tr>
<tr>
<td>pJJ182-SDM2</td>
<td>AAATTTTGTTACACCCATTTGATTTGACAAATACG</td>
<td>209.4±74.5</td>
</tr>
</tbody>
</table>

Table 3.2 Identification of the functional elements of the PR1-4 promoter.

Measuring β-galactosidase activities in cultures of P90C which carrying different constructs containing different mutations. Cultures were grown at 37 °C in LB to mid-log phase. Values are the means of three independent experiments performed in triplicate, giving a mean ± SD. The central part of the Figure shows the fragment base sequence from position +131 to +171, with the location of predicted PR1-4 promoter -10 and -35 hexamer elements underlined while -10 hexamer predicted by Cieslewicz and Vimr (1996) shaded in grey. Base substitutions in the different sites are highlighted in red.
In order to identify the -35 elements essential for PR1-4 promoter activity, the third G in the -35 hexamer was also substituted by A (138G>A) in the construct pJPR1-182SDM4. As expected, the activity of promoter was reduced 30% compared to wild type pJPR1 (Table 3.2) just by single base pair changed, which suggested that the motif TTGTTA at positions from +136 to +141 is likely the functional -35 hexamer element.

I sought to use this method to identify the functional elements for minor promoters PR1-2 and PR1-3. However, it was unsuccessful since the relatively low promoters’ activities (13.7±3.0 and 64.2±3.1 Miller Units, respectively) would not show obvious difference when compared to the wild type pJPR1 with extremely high activity (2908.5±124.7 Miller Units) (Data not shown).

In conclusion, promoter PR1-2 and PR1-3 may are the cryptic promoters while PR1-1 and PR1-4 are the major promoters contributes majority of the total transcripts. Nevertheless, it is still not clear how much transcripts come from these two major promoters individually. Thus, more experiments need to be carried out to quantify the transcripts come from each promoters.

3.4 Analysis of transcriptional level at PR1 from two major promoter PR1-1 and PR1-4

To gain more insight into the transcription in PR1 region, in this section, I began an analysis of the multiple promoters by qRTPCR in an effort to determine the relatively contribution between two major promoters PR1-1 and PR1-4.

3.4.1 Quantification of the relative contribution of each transcriptional start site to the total transcript at 37 °C

In order to quantify the actual transcripts driven from each promoter, qRTPCR assay (Method 2.21) was performed to quantify the absolute mRNA copy number by amplifying target region. Thus, three sets of primers used for amplifying the
different target regions located just downstream of the transcription start sites were designed (Figure 3.15). Also, 121 bp amplicon from housekeeping gene *rpoD* was amplified by primers (rpoD-qPCR-F and rpoD-qPCR-R) that served as an internal control. All primers were detailed in Table 2.3 and the corresponding amplicons in strain pJPR1 (P90C) were shown in Figure 3.15. The standard curve template for the these primers was a 1058 bp length of DNA fragment which was amplified from plasmid pJPR1 by PCR using primers pJJ2-F-EcoR1 and pRS415-LacZ-SEQ (Table 2.3) and then prepared by 1/10 series dilution from the starting concentration 100 pM. As shown from Figure 3.15, the transcripts only initiated from promoter PR1-1 was measured by amplifying the region from +22 to +101 with primers 1_1a_F and 1_1a_R. A second set of primers for detecting the mRNA transcribed from PR1-1 was also designed by amplifying the region from +38 to +214, which presented the actual contributing transcripts initiated from PR1-1 after running through 200 bp of UTR region. The total transcripts from PR1 going into *lacZ* gene were detected by amplifying 129 bp of reverse transcribed mRNA with primers 1_4_F and 1_4_R.

Total mRNA was extracted from strain pJPR1 (P90C) grown at 37 °C until to the mid-exponential phase (OD600nm=0.5) and reverse transcription was performed using 1 µg of total RNA. The qRT-PCR was then performed with 1:10 diluted reverse transcribed products (cDNA) per 10 µl reaction as described previously (Method 2.21). In this study, the absolute quantification method was used for quantification of transcription that allowed the precise determination of copy number per reaction. The absolute standard curves for each individual amplicon were shown in Figure 3.16 and no primer dimer artefacts were observed.
Figure 3.15 Illustration of qRTPCR primers and corresponding amplicons performed in qRTPCR assay of strain P90C (pJPR1).

873 bp of Promoter Region 1 plus downstream lacZ gene from construct pJPR1 were displayed. The name of the amplicon was shown on the left side while the corresponding lengths were shown on the right side.

Figure 3.16 Standard curves for 1-1a, 1-1, 1-4 and rpoD amplicons performed in qRTPCR assay.

A standard curve is generated from StepOne™ software v2.3 by plotting the Ct values against the logarithm of the initial copy numbers. Amplicons lengths are 80 bp (A), 177 bp (B), 129 bp (C) and 121 bp (D), respectively. Eff% presented the primers working efficiency. Linear regression equation was indicated by $Y=mX+b$ on each corresponding standard curve.
The copy numbers of experimental target RNAs calculated after real-time amplification from the linear regression of standard curves was summarized in Table 3.3. For each amplicon, the expression level was determined in three independent biological repeat experiments by normalizing against the reference gene rpoD mRNA copy numbers. For the strain P90C (pJPR1) grown at 37°C, the total number of mRNA copies transcribed from the Region 1 promoters was $2.75 \times 10^8 \pm 1.26 \times 10^7$ copy numbers (1-4 amplicon) (Table 3.3). For transcripts that only came from the promoter PR1-1 (1-1a amplicon) the value was $1.79 \times 10^8 \pm 9.52 \times 10^6$ copy numbers, whereas using the primer set 1_1_F and 1_1_R measuring transcription that goes through 200 bp of the UTR the transcription was reduced to the level of $1.37 \times 10^8 \pm 1.12 \times 10^7$ copy numbers (Table 3.3). The transcripts coming from promoter PR1-4 represented by PR1-4 amplicon was calculated by subtracting the copy number of 1-1 amplicon from the copy number of 1-4 amplicon for each single experiment, which were $1.38 \times 10^8 \pm 1.97 \times 10^6$ copy number and equal to 50% of the total transcripts (Table 3.3). Since the promoter PR1-2 and PR1-3 displayed very low β-galactosidase activity compared with the promoter PR1-1 and PR1-4 (Figure 3.11), therefore they were unlikely to be contributing significant transcription into kpsF.

All these results suggested that in the mid-exponential phase of P90C (pJPR1) grown at 37°C, promoter PR1-4 was activated and accounted for 50% of the total transcripts, and the transcription from PR1-1 was initially contributed 65% to the total transcripts in the PR1 region and then decreased to 50% after a 200 bp UTR region.
### Table 3.3 Transcripts copy number of the promoters PR1-1 and PR1-4 in pJPR1 (P90C) mid-log cultures grown at 37 °C.

<table>
<thead>
<tr>
<th>Amplicon</th>
<th>1-1a</th>
<th>1-1</th>
<th>PR1-4</th>
<th>1-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copy number/µg RNA</td>
<td>1.79E+8</td>
<td>1.37E+8</td>
<td>1.38E+8</td>
<td>2.75E+8</td>
</tr>
<tr>
<td>SE</td>
<td>9.52E+6</td>
<td>1.12E+7</td>
<td>1.97E+6</td>
<td>1.26E+7</td>
</tr>
<tr>
<td>%</td>
<td>65%</td>
<td>50%</td>
<td>50%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Each repeated results were normalized against the copy number of *rpoD*. Values are the means ± standard error of three independent experiments performed in quadruplex. % stand for the ratio of the corresponding transcripts to the Total Transcripts.

#### 3.4.2 Destruction of promoter PR1-1 caused attenuation of PR1-4 transcriptional level

The qRTPCR results suggested that around 50% of the total transcripts going into *kps* gene cluster was contributed by promoter PR1-4 in the mid-exponential phase at 37 °C (Table 3.3). However, in strain P90C (pJPR1-Δ+1), carrying the mutation in the PR1-1 -10 element (-12T>C), the β-galactosidase activity was reduced over 95% compared to the wild type P90C (pJPR1) (Figure 3.13). This suggested that when the promoter PR1-1 was inactivated, the activity of PR1-4 might be also affected.

To confirm whether the transcription from PR1-4 was dependent on the activation of PR1-1, the qRTPCR assay was carried out to assess of the activity of promoters PR1-1 and PR1-4 individually in plasmids pJPR1-Δ+1 and pJPR1-182SDM2, that carried a single base pair mutation at the -10 promoter element at PR1-1 and PR1-4 respectively. Total RNA was extracted from strains grown at 37 °C in the mid-exponential phase (OD600nm=0.5). 1 µg of extracted total RNA was reverse transcribed into cDNA and 1:10 diluted into 200 µl of cDNA products. The qRTPCR were repeated as described above and no primer dimer artefacts were observed. The total amount of transcripts from PR1 region was determined as the copy number of the 1-4 amplicon and the transcripts from PR1-1 and PR1-4 were identified by the 1-1a and PR1-4 amplicons respectively. The transcription copy
number of 1-1a amplicon in strain P90C (pJPR1-Δ+1) was reduced to $1.92 \times 10^7 \pm 8.72 \times 10^5$ [copy number/μg RNA] as compared to the level of transcription detected in the wild type P90C (pJPR1) of $1.79 \times 10^8 \pm 9.52 \times 10^6$ [copy number/μg RNA] (Figure 3.17). This indicated that the promoter PR1-1 in strain P90C (pJPR1-Δ+1) was inactivated by ~90% and this result was consistent with the previous β-galactosidase results described above (Figure 3.13 and Figure 3.17). It was notable that the copy number of 1-4 amplicon in the strain P90C (pJPR1-Δ+1) were decreased by 92% from $2.75 \times 10^8 \pm 1.26 \times 10^7$ to $2.10 \times 10^7 \pm 1.14 \times 10^6$ [copy number/μg RNA] in comparison with the strain P90C (pJPR1) (Figure 3.17), indicating that the total transcript from PR1 region of strain P90C (pJPR1-Δ+1) was reduced 92% compared to wild type P90C (pJPR1). In the meantime, the transcript copy number of PR1-4 in strain P90C (pJPR1-Δ+1) was also significant reduced to $5.59 \times 10^6 \pm 7.53 \times 10^5$ [copy number/μg RNA], which was a reduction of ~96% when compared to wild type strain P90C (pJPR1) (Figure 3.17). Therefore, it would appear that the destruction of promoter PR1-1 caused a proportional attenuation of transcription from PR1-4. This result confirmed the previous prediction that the transcription activation of promoter PR1-4 was dependent on the transcription initiation of PR1-1.

A 30% reduction in the copy number of 1-4 amplicon in strain P90C (pJPR1-182SDM2) compared to strain P90C (pJPR1) (Figure 3.17) was consistent with the previous observation that the β-galactosidase activity of P90C (pJPR1-182SDM2) dropped 30% when compared to the wild type P90C (pJPR1) (Table 3.2). However, the transcript copy number from PR1-1 (copy number of 1-1a and 1-1 amplicons) showed no significant difference between the wild type P90C (pJPR1) and the mutant P90C (pJPR1-182SDM2) (Figure 3.17), indicating that the promoter activity of PR1-1 was not affected by the inactivation of promoter PR1-4. Therefore, these results suggested that the transcription from PR1-4 is dependent on the activation of PR1-1 whereas the transcription of PR1-1 is context independent.
Figure 3.17 The transcript copy number of promoters PR1-1 and PR1-4 in promoter destruction mutants grown at 37 °C.

The corresponding amplicons were shown in above. Amplicon 1-1a represented the transcript just transcribed from the promoter PR1-1. Amplicon 1-1 represented the transcript from promoter PR1-1 after running through 200 bp of UTR. Amplicon 1-4 represented the total transcripts come from PR1 region. The transcript copy number from PR1-4 was represented by PR1-4 that was calculated by subtracting the copy number of 1-1 amplicon from the copy number of 1-4 amplicon. Each repeated results were normalized against the copy number of rpoD. Values are the means ± standard error of three independent experiments performed in triplicate.
3.5 Discussion

3.5.1 Observation the multiple tandem promoters at PR1 5’ UTR region

In the previous study, the transcription start sites (TTSs) mapping of *kpsF* gene in *E. coli* group 2 capsule gene cluster was performed in primer extension assay (Cieslewicz & Vimr, 1996). In order to identify the *E. coli* K5 region 1 promoter, primer extension assay was performed to detect the transcription start sites in the PR1 region. There were also three additional potential transcriptional start sites detected which approximately located around 40, 90, 220 bp upstream of *kpsF* gene (Simpson *et al*.,1996). This result was consistent to the identification of TSSs at position +1, +133 and +183 in the *E. coli* K1 capsule gene cluster (Cieslewicz & Vimr, 1996). Both of their experiments definitely found there were three potential TSSs, however, they only focus on the most 5’ of *kpsF* gene and no further investigation was carried on to study the additional promoters involved in the modulating the *kpsF* expression. In this study, we used a well-established and widely used method 5’ RACE to specifically amplify the 5’ end of a transcript and facilitating mapping of the TSS and the approximate location of promoter elements. There were three additional putative transcription start sites were tentatively discovered by 5’ RACE assay, which located positions +133, +142 and +182 in the 5’ UTR with respect to the original transcription start point of the *kpsF* mapped at +1 (Figure 3.7).

To evaluate the relative strengths of tandem *kpsF* promoters, different minimal promoter-*lacZ* fusion constructs containing the promoter region corresponding to promoter PR1-1, PR1-2, PR1-3 and PR1-4 were made and the transcription activity was measured by β-galactosidase assay and qRTPCR assay. All of these four tandem promoters were functional in the minimal promoter-*lacZ* fusions with different levels of transcriptional activity. Two overlapping temperature regulated promoters PR1-2 and PR1-3 were cryptic promoters which displayed extremely low activity on their minimal constructs pJJ133 and pJJ133-142 at 37 °C, compared to the intact PR1 *lacZ* fusion pJPR2 (Figure 3.11). Therefore, this poor activity coming
from PR1-2 and PR1-3 seemed neglected to the contribution of the total transcripts into *kpsF*.

Promoter PR1-4 *lacZ* minimal fusion pJJ182 lost temperature regulation and exhibited moderate transcriptional activity with comparison to pJPR2 at 37 °C. Also, the minimal promoter PR1-1 *lacZ* fusion pJJ2 lost its temperature regulation and showed the highest activity (Figure 3.11) both at 37 °C and 20 °C. Therefore, it was reasonable to consider promoter PR1-1 and PR1-4 were the principle promoters in the PR1 region, which PR1-1 was the most powerful promoter whereas PR1-4 had relative less transcriptional activity. However, by quantification of the relative transcripts coming from promoter PR1-1 and PR1-4 in qRT PCR assay, it was shown that both PR1-1 and PR1-4 contributed evenly to the total transcripts of *kpsF* gene in the exponential phase at 37 °C. The transcription efficiency will decrease 15% off after running 200 bp of UTR region eventually, even though there were 65% of the total transcripts just transcribed from PR1-1 (Table 3.3). This result was corresponding to the β-galactosidase assay performed in strain P90C (pJJ2). The observation that strain P90C (pJJ2) displayed dramatically increased β-galactosidase activity suggested that 225 bp 5’ UTR plays a negative role in moderating the transcription from PR1 (Figure 3.3).

The minimal promoter-*lacZ* fusion such as P90C (pJJ133), P90C (pJJ133-142), P90C (pJJ182A) were temperature regulated with significant reduced expression at 20 °C, which was in agreement that the expression of *kpsF* was transcriptionally silent at the lower temperature and no unique start sites were detected with mRNA isolated from cells grown at 20 °C (Cieslewicz & Vimr, 1996). However, what is clear is that when the 5’ UTR is separated away from the rest of the region 1 promoter, P90C (pJJ2) is no longer temperature-regulated. This implied that 225 bp 5’ UTR may involve in temperature regulation of promoter PR1-1. It has shown previously that in the case of PR3 the large 741 bp UTR acts to reduce the level of transcription that reaches the first gene *kpsM* (Xue et al., 2009). Therefore, in the case of PR1, the smaller 225 bp UTR could also be acting in a similar way to reduce the amount of transcription that reaches the *kpsF*. However as well as having a *cis-*
activity the 5’ UTR could also be the site for trans-acting factors that moderate transcription from PR1-1, such as the binding of nucleoid associated protein IHF at +130 will hinder the transcription elongation from PR1-1, or H-NS may bind at downstream of +1 in the 5’ UTR forming a bridge and trap the RNAP at PR1-1. However, it is not possible to exclude the possibility that the UTR is also having a posttranscriptional effect, possibly affecting the efficiency of translation. It was demonstrated that the expression of a number of genes encoded outer membrane proteins, which is regulated by untranslated small RNAs (sRNAs) in E. coli (Guillier et al., 2006). The mode of action of sRNAs is to base pair with the target mRNA, usually around the ribosome binding site, and thereby affect the stability and translatability of the mRNA (Guillier et al., 2006). To determine if there are sequences in the E. coli genome homologous to the 5’ UTR that could potentially function as sRNAs, a BLASTN analysis was carried out and there was no significant sequence identity between this region and sequences in E. coli genomes (data not shown). This suggests that a trans-acting sRNA is unlikely to be interacting with the UTR to affect the translation of the mRNA. The mRNA secondary structures of 5’ UTR fragments may play an important role in regulating the efficiency of translation. It had been demonstrated that the bubble-like region of the UTR secondary structure might have a direct effect on the efficiency of mRNA translation, or via interactions with other regulators (Jovcic et al., 2008). For example, the expression of transcription activator for virulence gene in Listeria monocytogenes was inhibited by its UTR mRNA second structure at 30 °C. The formation of the mRNA second structure at UTR region of gene prfA will mask the ribosome binding site and inhibiting the translation process of prfA at 30 °C (Johansson et al., 2002). Also, same thermoregulation mechanism by 5’ UTR RNA thermosensor was also observed in controlling the expression of capsule biosynthesis gene cssA in Neisseria meningitidis (Loh et al., 2013). But in the 5’ UTR of PR1 of E. coli K5 capsule gene cluster, there was no high probability (>60%) to form an RNA step-loop structure that including the ribosome binding site in the UTR was predicted by a web Serves for RNA Secondary Structure Prediction - “RNAstructure”(http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict
Therefore, it excluded the possibility that RNA thermosensor in the 5’ UTR control the posttranscriptional regulation of K5 capsule gene in *E. coli*. In addition, previous work showed in other systems that the presence of a long region of 5’ untranslated mRNA is important in mRNA stability affecting decay and increasing messenger half-life (Hillmann *et al.*, 2007). But in the *E. coli* K5 capsule gene, it already confirmed that the presence of the UTR of PR3 had no effect on the half-life of the *lacZ* mRNA (Xue *et al.*, 2009).

### 3.5.2 Identification of the promoter elements for multiple promoters

A typical *E. coli* promoter contains a 4/6 match to the consensus within both the -10 and -35 hexamers (Mitchell *et al.*, 2003). It is supporting the identification of the putative promoter PR1-3 and PR1-4 in the UTR might be genuine since they both contains 8 out of 12 consensus nucleotides. Unfortunately, it was failure to identify the exactly location of the promoter element of promoter PR1-2 and PR1-3 by the second base pair mutation in the -10 element, because of the difficulty of detecting the small difference between the wild type pJPR2 and the its mutants of promoter PR1-2 and PR1-3. But the extremely low activity of minimal promoter-*lacZ* fusion pJJ133 and pJJ133-142 indicated that they might be cryptic promoters with very low activity. The existing of the additional promoters PR1-2 and PR1-3 at PR1 may hold the RNAP in transcriptionally inactive “closed” complex and thereby increase the RNAP concentration and have higher binding affinity towards the real principle promoter (Huerta *et al.*, 2006). On the other hand, these overlapping promoters may be mutually exclusive and the interaction of several RNAP within the overlapping promoter-like sites can interfere with each other. Hence it was speculated that these two overlapping promoters PR1-2 and PR1-3 may compete and mutually exclusive based on the previous observations that the enterohaemorrhagic *E. coli* LEE1 operon has two overlapping promoters and the RNAP binding region of upstream cryptic promoter P1A is sequestered by the -35 elements of downstream promoter P1 (Islam *et al.*, 2011). What is more, the IHF binding sites previously was found binding at +130 and IHF protected at least 25 bp on the DNA template (Goodrich *et al.*, 1990). At the PR1 region, the binding of IHF
at +130 just covered the -10 promoter elements of promoter PR1-2 and PR1-3. This might be explained by that the binding of IHF would obstruct the binding of the RNAP and thus repress the transcription from promoter PR1-2 and PR1-3.

In this study we confirmed that there are three additional promoter within the K5 5’ UTR and that this corresponds to that previously mapped in the K1 5’ UTR. Computational analysis of *E. coli* sequences had revealed that existence of potential of -10 and -35 elements displayed higher frequencies than the expected frequencies in intragenic regulatory regions (Huerta & Collado-Vides, 2003). The number of predicted transcription start sites greater than the typical number of functional promoters and clustered of promoter-like signals were found for more than 80% of genes (Huerta & Collado-Vides, 2003). In *E. coli*, there were 78 regions with an extremely large number of potential transcription start points; termed promoter islands have been found. It was hypothesized that the transcription from these promoter islands are halted by transcriptional regulator covering (such as H-NS), and only short abortive mRNA can be synthesised (Panyukov & Ozoline, 2013).

In this study, the 5’ RACE assay results showed there were many putative transcription start sites clustered within the AT rich region around PR1-2 and PR1-3 (Figure 3.7), which implied that there might have other promoter-like sites within this region. However whether these transcription start sites are the degraded mRNA product from the upstream promoters or there are other putative promoter-like signals, still awaits further elucidation. It was proposed previously that these additional promoter PR1-2 and PR1-3 or other promoter-like sites are cryptic promoter which almost silent in a given genetic context, however, these functional elements could be defunct or can be activated by mutation for helping the adaption of bacterial populations to environmental changes (Huerta & Collado-Vides *et al*., 2006).

In conclusion, there were three additional temperature regulated promoters, PR1-2, PR1-3 and PR1-4 that were identified in respect to the original mapped promoter PR1-1 at +1, which located at site +133, +142 and +182, respectively. Promoters PR1-1 and PR1-4 were the major promoter with equivalent contribution
to the total transcript of *kps* gene cluster in the mid-exponential phase at 37 °C. Promoters PR1-2 and PR1-3 had relative low level of transcription activity that was suggested they may be cryptic promoters. Additionally, transcription from PR1-4 was dependent on the activation of PR1-1.
Chapter 4. Investigating how the multiple promoters at PR1 of the *E. coli* K5 gene cluster are regulated by a transcriptional regulator(s)

4.1 Role of H-NS in transcription of PR1 region of K5 capsule gene

Previous studies (Rowe *et al*., 2000; Corbett *et al*., 2007) demonstrated that H-NS appears to have a dual role in the thermoregulation of the *E. coli* K5 capsule gene cluster. The observation made by Corbett *et al*. (2007) indicated H-NS was required for maximal transcription at 37 °C while repressing transcription at 20 °C (Rowe *et al*., 2000; Corbett *et al*., 2007). In this section, the binding of H-NS to the UTR region was further determined by electrophoretic mobility shift assay (EMSA), and a variety of *lacZ* reporter gene fusions were generated to investigate the whether H-NS regulates other additional promoters at PR1.

4.1.1 H-NS as a negative regulator of the capsule gene cluster

H-NS generally acts as a repressor of virulence gene expression rather than a transcriptional activator (Fang & Rimsky, 2008). It is involved in the thermoregulation of many genes in *Enterobacteriaceae*, typically repressing their expressions at low temperature (Atlung and Ingmer, 1997). In respect to the PR1 region, with its multiple promoters the mechanism by which H-NS exerts its regulatory control remains to be determined. It is worthy to investigate whether H-NS binds to the PR1 region and regulates the PR1 multiple promoters’ transcription directly or perhaps through other indirect means.

4.1.1.1 H-NS regulates transcription at PR1 negatively and is important for temperature regulation at PR1

To determine the functional significance of potential H-NS regulating PR1 transcription *in vivo*, β-galactosidase assays were performed on cultures of wild type strain P90C (Table 2.1) and its *hns::kan* derivative. The strain P90C *hns::kan* was generated from strain MS101 *hns::kan* (Table 2.1) by P1 transduction and the successful transductants were checked by colony PCR as described above by using
primer HNS-F and HNS-R (Table 2.3). The expected size of amplicon of *hns::kan* mutants with the kanamycin cassette should around 2 kb while wild type strain P90C will have an amplicon of 481 bp, and an agarose gel of one such screening is shown in Figure 4.1.

**Figure 4.1 Colony PCR screening of P90C and P90C hns::kan transductants.**

Lane 1-5 represents the P90C hns::kan transductants screened by primer HNS-F and HNS-R. Lane 6-8 represents wild type P90C screened by primer HNS-F and HNS-R. Lane 9 represents negative control without template. M, Hyperladder I DNA markers and molecular weight indicated on the left of the gel.

Two different transcriptional PR1-lacZ fusions pJPR2 and pDSHcH (Table 2.2) were transformed into strain P90C and P90C hns::kan respectively and β-galactosidase assays were performed at 37 °C and 20 °C (Figure 4.2 and Figure 4.3). Plasmid pJPR2 is a transcriptional fusion carrying 869 bp of PR1 region from -645 to +224 at the *BamH1* site of vector pRS415. A 1.1 kb PR1 region fragment cut by *HincII/HindIII* from plasmid pCB191 was inserted into the transcriptional fusion vector pCB192 to generate plasmid pDSHcH. As shown from Figure 4.2, in comparison to the strain P90C (pJPR2), the level of transcription activity from PR1 region in strain P90C hns::kan (pJPR2) was increased 1.3-fold at 37 °C (from 2036.8 ± 168.8 to 2637.4 ± 89.1 Miller Units), suggesting that H-NS may play a role in repressing region 1 transcription at 37 °C. At 20 °C, a significantly increase (around 5-fold increase) of transcriptional activity in strain P90C hns::kan (pJPR2) was observed which indicated that H-NS may also play a critical role by repressing the
transcription at 20 °C. In addition, as seen in Figure 4.3, the pattern of β-galactosidase activity expressed in pDSHcH was essentially shown the same trends as observed in pJPR2 that the level of transcription activity was increased in the hns mutants both at 37 °C and 20 °C, although it was found significantly lower (around 100 – 150 fold) compared with pJPR2 in both null and mutant strains (P90C and P90C hns::kan) at both temperatures. The significant increase of β-galactosidase activities observed in hns::kan mutants at 20 °C indicated that H-NS was responsible for the thermoregulation of PR1 transcription in vivo since the transcription of PR1 lost its temperature-dependent manner in the hns::kan mutants.
Figure 4.2 The effect of the hns:kan mutants on Region 1 promoter activity at 37 °C and 20 °C in the vector pRS415 background.

Strains P90C and its hns::kan mutant derivative carrying reporter vector pRS415 and PR1 region fusion pJPR2 were grown to mid-log phase at 37 °C and 20 °C and the promoter activity was measured by β-galactosidase assay. Values are the means of three independent experiments performed in triplicate. Error bars represent standard error.

Figure 4.3 The effect of the hns:kan mutants on region 1 promoter activity at 37 °C and 20 °C in the vector pCB192 background.

Strains P90C and its hns::kan mutant derivative carrying reporter vector pCB192 and PR1 region fusion pDSHcH were grown to mid-log phase at 37 °C and 20 °C and the promoter activity was measured by β-galactosidase assay. Values are the means of three independent experiments performed in triplicate. Error bars represent standard error.
To further confirm the negative role of H-NS in regulating transcription of the K5 capsule gene cluster, qRTPCR assay were performed using strain MS101 and MS101 *hns::kan* (Table 2.1) grown at 37 °C and 20 °C respectively. Total RNA was extracted from strain MS101 and MS101 *hns::kan* grown at 37 °C or 20 °C until mid-exponential phase (OD600_l=0.5) and reverse transcription was performed using 1 µg of total RNA, respectively. The qRTPCR was then performed with 1:5 diluted reverse transcribed products (cDNA) per 10 µl reaction. The primer set MS101-qPCR-F/MS101-qPCR-R would give a 126 bp amplicon from the 5’ coding region of *kpsF* of MS101 K5 gene cluster and the standard curve template for primer MS101-qPCR-F/MS101-qPCR-R was a 978 bp DNA fragment which amplified from strain MS101 by colony PCR using primer pJJ2-F-EcoR1 and SP1-MS101 (Table 2.3). The template was then prepared by 1/10 series dilution from the starting concentration 100 pM and the standard curve was shown in Figure 4.4. Also, a 121 bp amplicon from housekeeping gene *rpoD* was amplified by primers rpoD-qPCR-F and rpoD-qPCR-R which served as an internal control (Table 2.3), and the standard curve was kept using the same as shown in Figure 3.17. D.

As shown in Figure 4.5, the copy number of *kpsF* transcript in MS101 *hns::kan* was increased more than 2-fold compared to wild type strain MS101 at 37 °C, further suggesting that H-NS plays a negative role in regulating transcription at PR1 region at 37 °C. What is more, at 20 °C, there was a significant increase (P<0.05) in transcript copy number of the *kpsF*-specific transcript in MS101 *hns::kan* that confirms that H-NS was required for repression of expression from PR1 region at 20 °C.

Taken together, all data indicated that H-NS regulated the transcription at PR1 negatively both at permissive and non-permissive temperature and was important for temperature regulation at PR1.
Figure 4.4 Standard curves for MS101-kpsf amplicon performed in qRT PCR assay.

A standard curve is generated from StepOne™ software v2.3 by plotting the Ct values against the logarithm of the initial copy numbers. Amplicons length is 126 bp. Eff% presented the primers working efficiency. Linear regression equation was indicated by Y=mX+b on each corresponding standard curve.

Figure 4.5 H-NS represses the transcription of PR1 promoter region of E. coli K5 capsule gene cluster.

RNA were extracted when strains grown to mid-log phase at 37 °C and 20 °C respectively. Each repeated results were normalized against the copy number of rpoD. Values are the means ± standard error of three independent experiments performed with technical quadruplicate.
4.1.1.2 *In vivo* transcription of PR1 tandem promoters were influenced negatively by H-NS

As the data shown above in Figure 4.2, the transcriptional activity of plasmid pJPR2 containing the whole PR1 region was significantly increased in strain P90C *hns::kan* suggesting that H-NS played a role in repressing region 1 transcription both at the permissive and non-permissive temperatures. Therefore, it was worthy to investigate how the individual promoters in the PR1 region was affected by H-NS. Thus, different minimal promoter-*lacZ* fusions (Figure 4.6) were introduced into strain P90C *hns::kan* and β-galactosidase assays were performed at 37 °C and 20 °C.

As seen in Figure 4.7, H-NS exerted significant inhibition of transcription at promoter PR1-1 in plasmids pJJ2 and pJJ2A, similar to effects on the level of transcription at PR1 region (plasmid pJPR2). The whole UTR region in pJJ1 was also repressed significantly by H-NS and 1-fold repressed was also observed at promoter PR1-4 (in plasmid pJJ182A and pJJ182) (Figure 4.7). The observation that a *hns* mutant resulted in a 2-fold increase in β-galactosidase activity with plasmid pJJ133-142 (Figure 4.7) but had no effect at 37 °C on β-galactosidase activity with plasmid pJJ133 indicated that H-NS repressed the transcription at promoter PR1-3 but not PR1-2 at 37 °C (Figure 4.7). At 20 °C, the β-galactosidase activity of all the promoter-*lacZ* fusion was increased significantly in the mutant strain P90C *hns::kan* (Figure 4.8). Overall, the extent of the H-NS repression seen in this experiment closely corresponds to that found in previous observation described above, suggesting that the transcription repression exerted by H-NS at the multiple promoters, especially at promoter PR1-1 and PR1-4, resulted in a significant repression at PR1 both at 37 °C and 20 °C.
Figure 4.6 Different minimal promoter-lacZ fusion inserts.
Figure 4.7 H-NS repressed transcription at PR1 in vivo at 37 °C.

β-galactosidase activities of wild type strain P90C and P90C hns::kan bearing different minimal putative promoter-lacZ fusions grown at 37 °C. Values are the means of three independent experiments performed in triplicate. Error bars represent standard error.
Figure 4. 8 H-NS repressed transcription at PR1 in vivo at 20 °C.

β-galactosidase activities of wild type strain P90C and P90C hns::kan bearing different minimal putative promoter-lacZ fusions grown at 20 °C. Values are the means of three independent experiments performed in triplicate. Error bars represent standard error.
4.1.2 EMSA analysis of H-NS binding region at PR1

H-NS has been identified interacting directly with PR1 of the K5 capsule gene cluster and DNaseI footprinting assay has been used to map the binding sites for H-NS 5’ to the transcriptional start site at PR1-1 (Corbett et al., 2007). Given the previous results for H-NS regulating transcription at the PR1 negatively (Figure 4.7 and 4.8), therefore, it was necessary to determine whether H-NS was able to bind directly to the PR1 5’ UTR region.

The hns gene was cloned into plasmid pET22b yielding plasmid pET22BH-NS (Table 2.2), encoding a C-terminally His6-tagged H-NS protein. The recombinant protein was over-expressed in E. coli BL-21 (Table 2.1) and purified on a 1ml Ni-NTA resin column and following washing eluted gradient with imidazole in chromatography systems (performed by Protein Expression Facility in University of Manchester).

DNaseI footprinting assay had been done by Corbett et al. (2007) shown that H-NS protected three regions spanning through PR1 promoter region from position -224 to -134, -121 to -79 and +1 to +32. According to this observation, two PCR fragments H-NS_F1 and H-NS_F2 (Figure 4.9) were generated by two sets of primers HNS01_F and HNS01_R, HNS02_F and H-NS02_R (Table 2.3), in order to identify whether the UTR region of PR1 region bound to H-NS at 37 °C. H-NS protein and purified DNA fragments were incubated at 37 °C for 30 min in vitro and resolved on non-denaturing 5% TBE-polyacrylamide gels, and the H-NS:DNA complexes were identified as bands whose electrophoretic mobility has been reduced relative to no protein (0 µM) control. The results shown in Figure 4.10A indicated that H-NS specifically bound to the F1 fragment that contained the region from -338 to +34 shown previously to bind H-NS (Corbett et al., 2007). H-NS binding started retarding the mobility of the F1 fragment from the concentration of 2.4 µM and no free fragment was detectable at 4.2 µM (Figure 4.10A). PCR fragment F0 was amplified by primer M13-F and M13-R (Table 2.3) from pSK+ (Table 2.2) and served as free DNA which should not have binding affinity to protein H-NS. This fragment showed no binding by H-NS under the conditions used (Figure 4.10A) indicating a specific binding of H-NS to fragment F1. In comparison
using the same concentration range of H-NS no specific binding was detectable to fragment F2 (Figure 4.10B). These data indicate that H-NS is not binding in vitro to the UTR region from +39 to +208 and suggest that the inhibitory effects on PR1-2, PR1-3 and PR1-4 are likely to be indirect.
Figure 4.9 DNA fragments used in analysis of H-NS binding at PR1 region.

Fragments were amplified by PCR and used in H-NS EMSA analysis. Scale bar represents 100 bp.

Figure 4.10 H-NS binds to PR1 upstream region but not UTR region in vitro.

EMSA of purified H-NS incubated with PCR fragments shown in Figure 4.1, performed at 37 °C. (A) H-NS binds to the fragment F1 that containing the PR1 region from -338 to +34. (B) H-NS had no binding affinity to the fragment F2 that containing the PR1 region from +39 to +208 in vitro. H-NS concentrations are indicated above each lane. M: Hyper ladder 100 bp DNA marker, molecular weight indicated on the left of the gel. Protein:DNA complex are indicated on the right by black bars. Free negative control DNA presented by F0.
4.1.3 The integrity of PR1 region is important for promoter PR1-1 temperature regulation which may related to H-NS binding

Analysis of the transcription activity of promoter PR1-1 indicated that when the promoter PR1-1 itself was separated away from the UTR, such as in plasmid pJJ2 (Figure 4.11), it lost temperature regulation with no significant repression of transcription at 20 °C (P value=0.22) (Figure 4.12). In contrast, plasmid pJJ2A that retained up to position +125 of the UTR displayed temperature regulation with significant decrease of 67% in transcription at 20 °C (P value=0.002). Since three H-NS binding sites (H-NS binding sites I II and III) overlapped the promoter PR1-1 region (Corbett et al., 2007), it is reasonable to assume that these three H-NS binding sites are with the cooperation among these H-NS binding sites might occlude the access of RNA polymerase to the -35 and -10 elements of the PR1-1 promoter. Therefore, to determine whether the upstream of H-NS binding site I and II are also required for repression, an additional constructs, pJJ2B which lacks the upstream H-NS binding sites was made (Figure 4.11). As predicted, pJJ2B lacking the upstream region and H-NS binding sites I and II of the PR1-1 promoter region also lost temperature regulation (Figure 4.12), with no significant difference between 37 °C and 20 °C (P value=0.06). Therefore, this confirms that the temperature regulation of PR1-1 required the integrity of PR1 region both upstream and downstream region of the TSS at +1. It was hypothesized that the regulation at PR1-1 by H-NS may relate to the H-NS binding sites over-spanning the PR1-1 region and the H-NS binding may form an unstable nucleoprotein complex which can resist the competition with the RNA polymerase binding at PR1-1 promoter region at 20 °C.
Figure 4.11 Comparison the different PR1-1 transcriptional lacZ fusion inserts.

The cloned fragments were amplified from template pJPR2 by PCR and inserted into EcoR1-BamH1 sites of vector pRS415.

Figure 4.12 Temperature regulation of PR1-1 require both upstream and downstream region of PR1-1.

Strains P90C carrying different PR1-1 region fusions were grown to mid-log phase at 37 °C and 20 °C and the promoter activity was measured by β-galactosidase assay. Values are the means of three independent experiments performed in triplicate. Error bars represent standard error.
4.2 The role of IHF in regulating transcription at PR1

Previous gel shift assays had demonstrated that IHF binding at +130 region where an IHF consensus binding region was present and that a 13-bp deletion of IHF consensus binding site around +130 abolished IHF binding (Rowe et al., 2000). It was speculated that binding of IHF to the downstream of promoter PR1-1 would induce DNA bending perhaps bringing upstream sequences closer to the RNA polymerase or other regulatory proteins (Rowe et al., 2000). Since in my study I have identified additional putative promoters in PR1, it was therefore important to re-examine and understand how IHF-mediated transcription regulation at PR1 multiple promoters.

4.2.1 IHF play a positive role in regulating the transcription at PR1-1 but repressing the transcription from PR1-4

In order to determine if global regulator IHF was also involved in other promoters’ activation, the promoter activity of different minimal promoter-lacZ transcriptional fusions in wild type P90C and also their Δihf derivative mutant strains were analysed by β-galactosidase assay (Figure 4.13). To generate an IHF inactive background in P90C, the ihfB::cm mutation of strain MS105 (Table 2.1) was moved to P90C by P1 transduction. The successful transductants were screened by the colony PCR using primers IHF-F and IHF-R (Table 2.3) for identifying the presence of chloramphenicol resistant gene. The expected size of the PCR product from ihfB::cm mutants was around 2 kb and from wild type strains P90C 846 bp (data not shown). In the ΔihfB mutants, a significant decrease in pJPR2 was observed, indicating that IHF is essential for the activation of promoters of Region 1 gene cluster at 37 °C (Figure 4.13). This result supported previous observations whereby expression of the KpsE protein was reduced in a himA mutant (Simpson et al., 1996). With respect to regulation of PR1-1, around 40% reduction of transcriptional activity was both observed in strain ΔihfB mutants bearing plasmids pJJ2 and pJJ2A (Figure 4.13), indicating that IHF regulated positively on the expression of PR1-1 promoter in the absence of the IHF binding site at +130. In
contrast, very significant increase was found in pJJ1 and pJJ182A in mutant strain P90C *ihfB::cm* compared to their wild type strains. No difference was observed in pJJ133-142 between the strain P90C and P90C *ihfB::cm* at 37 °C (Figure 4.13). It was indicating that IHF regulated the expression of promoter PR1-4 negatively but no effect on PR1-2 and PR1-3. Since pJJ182 lacks the IHF binding region, thus it was reasonable that no significant difference between wild type and ΔihfB mutants (Figure 4.13).

![Figure 4.13](image)

**Figure 4.13 IHF is essential for the transcription at PR1 in vivo at 37 °C.**

β-galactosidase activities of wild type strain P90C and *ihfB::cm* P90C bearing different minimal putative promoter-*lacZ* fusions grown at 37 °C. Values are the means of three independent experiments performed in triplicate. Error bars represent standard error.
4.2.2 Purification of IHF

IHF was purified from *E. coli* strain K5746 (Table 2.1) as described in Section 2.19. In the plasmids pP₁hip.himA-5, both subunits gene of *himA* (IHF-α) and *hip* (IHF-β) of IHF were placed under the control of a regulatable promoter. When the strain was shifted to 42 °C, the regulatable promoter started producing both subunits each around 10 kDa (Nash *et al.*, 1987). IHF was purified using a HiTrap™ Heparin HP 1ml column (GE Healthcare) according to the manufacturer’s instructions. Protein was eluted with a 50 mM – 1.5 M NaCl gradients in Fast Protein Liquid Chromatography (FPLC) system (AKTA Purifier) (Figure 4.14), and several stages during protein purification; samples were withheld for analysis by SDS-PAGE shown on Figure 4.15. Fractions from #29 – #31 (Figure 4.14) were supplemented with glycerol to final concentration 50% and stored at -20 °C. The concentration of the purified IHF protein was determined to be 6.67 μM using the Bio-Rad bovine serum albumin as a standard (data not shown). The eluted IHF finally was assessed for purity by SDS-PAGE. Under this condition α and β subunits of IHF cannot be completely separated (Figure 4.16).
10 ml of sample injection followed with 20 ml washing by Buffer A and then eluted by 30 ml of Buffer B. 1 ml/min flow rate with linear gradient from 50 mM to 1.5 M NaCl. IHF pooled at fraction #29 – #31. Chromatography was done using AKTA Purifier system.

![Figure 4.14 IHF elution profiles of HiTrap Heparin HP (1 ml) fractions.](image)

Figure 4.14 IHF elution profiles of HiTrap Heparin HP (1 ml) fractions.

M, prestained Marker (Precision Plus Protein Standard, BioRad); lane 1, cell lysate; lane 2, clarified extract; lane 3, dissolved pellet by Buffer A; lane 4, dialysis; lane 5, filtered; lane 6-13, flow through from selected column fractions (#5, #10, #15, #20, #25, #30, #35, #40). The arrow denotes the induced IHF protein. 15% denaturing gel was used. Molecular weights (kDa) are indicated.

![Figure 4.15 Coomassie blue stained SDS-PAGE analysis of IHF protein induction.](image)

Figure 4.15 Coomassie blue stained SDS-PAGE analysis of IHF protein induction.

M1, prestained Marker (Precision Plus Protein Standard, BioRad); M2, Polypeptide SDS-PAGE standards (BioRad). Lane 1-2, 8 µl of 3:1 mix of purified IHF and SDS loading dye. Lane 3, 5 µl of 3:1 mix of purified IHF and SDS loading dye. NuPAGE 10% Bis-Tris Mini Gels was used. Molecular weights (kDa) are indicated.

![Figure 4.16 Coomassie blue stained SDS-PAGE indicated purified IHF protein.](image)

Figure 4.16 Coomassie blue stained SDS-PAGE indicated purified IHF protein.
4.2.3 IHF is also directly involved in the transcriptional regulation at PR1 promoters

Previous results had demonstrated that IHF played a positive role indirectly in transcriptional regulation of whole PR1 region of kps operon at 37°C (Simpson et al., 1996). Interestingly, it was required for promoter PR1-1 activation but repressing the transcription of PR1-4 at 37°C. Therefore, in this section, further analysis of the physiological relevance of the IHF binding at +130 to transcriptional regulation of kpsF operon was performed by in vitro gel shift, site-directed mutagenesis, reporter fusion activity analysis and qRTPCR analysis.

4.2.3.1 EMSA analysis of IHF binding at PR1

Using EMSA, the purified E. coli IHF protein was tested for functionality by binding to the +130 region that is known to have an IHF binding site. A PCR fragment was amplified from pJPR1 or pCBIHF-1 (Table 2.2) using primer EMSA_IHF_F and EMSA_IHF_R (Table 2.3; Figure 4.17) and incubated with the purified protein IHF, respectively (Figure 4.18). Free DNA fragment F0 was amplified by primer M13-F and M13-R (Table 2.3) from pSK+ (Table 2.2) that has no binding affinity for the IHF protein. In the case of PCR fragment F1 (F1 containing the PR1 region from +45 to +223) incubated with increasing concentration of IHF, it was shown that IHF binding affected the mobility of the F1 fragment from the concentration of 0.24 µM with no free fragment detectable at 0.48 µM at 37°C (Figure 4.18A). In addition, PCR fragment F2 with a 13 bp deletion of IHF binding site was tested as negative control by incubating with IHF at 37°C. The failure of IHF bind to the F0 fragment (Figure 4.18 A, B) confirms that the purified IHF is functional and that binding is specific for the IHF binding site.
Figure 4.17 IHF sequence for EMSA analysis.

Sequence of region 1 promoter from +1 to +225 was shown. Transcription start site was indicated by broken arrows. IHF consensus binding sequence was highlighted in dark green and yellow highlighted sequence indicated the 13-bp deletion of IHF consensus binding sequence in plasmid pCBIHF-2 (Simpson et al., 1996). Red nucleotides indicated the substitutions for generating plasmid pJPR1-IHF-BSM containing mutated IHF consensus binding sequence by site-direct mutagenesis. Underlined sequence indicated the fragment that used for EMSA of IHF amplified by primer EMSA_IHF_F and EMSA_IHF_R.

Figure 4.18 EMSA analysis of purified IHF.

(A) IHF binds to the fragment F1. F1 containing the PR1 region from +45 to +223 amplified by primer EMSA_IHF_F and EMSA_IHF_R from pJPR1. (B) IHF has no binding affinity to the fragment F2. F2 containing the PR1 region from +45 to +223 amplified by primer EMSA_IHF_F and EMSA_IHF_R from pCBIHF-1 with 13-bp deletion of IHF consensus binding sequence. EMSA of purified IHF incubated with the two PR1 fragments (25 ng/µl) performed at 37 °C. Free negative control DNA presented by F0. IHF concentrations are indicated above each lane.
4.2.3.2 Destruction of IHF binding site at +130 of PR1 region by Site-direct Mutagenesis

To demonstrate in a more direct manner that the binding of IHF is mediating the transcription driven from PR1-4, an IHF binding site mutation was constructed by the Site-Direct Mutagenesis PCR. The IHF site at +130 (TTACAACCCATTG) [conserved nucleotides are underlined] has 7 out of 13 base pairs similarity with the reported *E. coli* consensus sequence with the conserved nucleotides (Rice *et al.*, 1996). Therefore, Site-Direct Mutagenesis PCR was used to abolish IHF binding site with multiple nucleotide substitution mutations by using primer IHF_BSM_F and IHF_BSM_R (Table 2.3) as descried above (section 2.11).

Eleven bases out of the consensus IHF binding sequence 5’-TTACAACCCATTG [the mutated nucleotides are underlined] were replaced by sequence 5’-GCATGTGACGGAC from region +139 to +152 of PR1 region (Figure 4.17). The altered sequence in PR1 region was checked by DNA sequencing and ligated into pRS415 *BamH1/EcoR1* site to generate PR1-*lacZ* transcriptional fusion pJPR1-IHF-BSM. To address whether IHF binds to these mutated IHF binding sites, EMSA was carried out with increasing concentration of the purified IHF protein with two PCR fragments (F1 and F2) which amplified from pJPR1 and pJPR1-IHF-BSM using primer EMSA_IHF_F and EMSA_IHF_R, respectively (Figure 4.19). It was observing that at the same concentration of IHF (0.23 µM), there was a significant band shift when IHF incubated with F1 fragment (Figure 4.19A) whereas no band shift was detected when incubated with mutated DNA fragment F2 (Figure 4.19B). Therefore, it was confirmed that the disruption of IHF binding site in pJPR1-IHF-BSM would completely prevent formation of the IHF-DNA complex.
Figure 4.19 Inactivation of IHF binding sites by site-direct mutagenesis at PR1 +130 region.

(A) IHF binds to fragment F1. F1 containing the PR1 region from +45 to +223 amplified by primer EMSA_IHF_F and EMSA_IHF_R from pJPR1. (B) IHF has no binding affinity to fragment F2. F2 containing the PR1 region from +45 to +223 amplified by primer EMSA_IHF_F and EMSA_IHF_R from plasmid pJPR1-IHF-BSM. EMSA of purified IHF incubated with the two fragments (20ng/µl) performed at 37 °C. Free negative control DNA presented by F0. IHF concentrations are indicated above each lane. M: Hyperladder 100 bp DNA markers, molecular length indicated on the left of the gel. Protein:DNA complex are indicated on the right by black bar.
4.2.3.2 The binding of IHF at +130 directly represses the transcription of PR1 region at 37 °C

In order to investigate the direct role by binding of IHF in regulating PR1 region promoters, plasmid pJPR1-IHF-BSM was introduced into strain P90C and followed with β-galactosidase assay along with strain pJPR1 (P90C) performed at 37 °C and 20 °C, respectively (Figure 4.20). Mutation of the IHF binding site in pJPR1-IHF-BSM led to a significantly increased (P value=0.001) β-galactosidase activity compared to the wild type reporter fusion pJPR1 (Figure 4.20), suggesting that the binding of IHF at +130 of PR1 region would effectively repress the transcription from PR1 region at 37 °C. The strain P90C (pJPR1-IHF-BSM) was still temperature regulated as same as P90C (pJPR1) with extremely low activity at 20 °C, indicating that binding of IHF was not involved into the transcriptional regulation of PR1 region at 20 °C. Overall, the reduction in transcription seen in an ihf mutant (Figure 4.13) coupled with the increase in transcription seen here when the IHF consensus site is abolished would suggest a complex role for IHF in regulating transcription from the PR1 promoter region. It would appear that acting indirectly it is needed for maximal transcription from PR1, but acts directly to reduce transcription by binding to the IHF binding site at +130.

Figure 4.20 Transcriptional activity of PR1 region was significantly repressed by the binding of IHF at 37 °C.

β-galactosidase assay were performed when strain pJPR1(P90C) and pJPR1_IHF_BSM grown to mid-log phase at 37 °C. Values are the means of three independent experiments performed in triplicate. Error bars represent standard error.
In addition, it had been shown that there was 63.8% increase of the β-galactosidase activity in pJJ182A was observed in the ΔihfB mutant compared to the wild type whereas no significant difference was found in pJJ182 that lacks the IHF binding sites (Figure 4.13). Thus it led to the hypothesis that the binding of IHF could play a negative role in regulating promoter PR1-4 transcription by binding upstream of PR1-4 at 37 °C. In which case it was predicted that the transcription activity of promoter PR1-4 in plasmid pJPR1-Δ+1 should be increased in the absence of IHF binding. Therefore, in a further attempt to determine whether the expression of PR1-4 was substantially affected by binding of IHF, new construct pJPR1-SDM+1-IHF-BSM was generated by site direct mutagenesis in plasmid pJPR1-IHF-BSM with primers SDMPR1-1_F and SDMPR1-1_R (Table 2.3). The site-mutated PR1 region was then ligated into the Bam H1/EcoR1 site in pRS415. Successful PR1-{}-lacZ transcriptional fusions consisting PR1 site-mutation were checked by DNA sequencing and transformed into strain P90C. Strains P90C (pJPR1-Δ+1) and P90C (pJPR1-SDM+1-IHF-BSM) were grown to mid-exponential phase (OD600 nm=0.4-0.6) and assayed for β-galactosidase activity (Figure 4.21). As predicted, loss of IHF binding site in P90C (pJPR1-SDM+1-IHF-BSM) resulted in a significantly (P value=0.04) increase in the level of transcription from PR1-4 compared with P90C (pJPR1-Δ+1) (Figure 4.21). This result further implied that IHF mediated regulation of transcription of PR1-4 was a direct consequence of transcriptional control by binding upstream of PR1-4.
In addition, in order to investigate whether the binding of IHF at PR1 region was involved in regulating the expression from individual promoter, qRTPCR assay was performed in the strains P90C (pJPR1), P90C (pJPR1-IHF-BSM), P90C (pJPR1-Δ+1) and P90C (pJPR1-SDM+1-IHF-BSM), respectively (Figure 4.22). As previously described, three sets of primers (1_1, 1_1a and 1_4) were used for amplifying the different target region were used and detailed in Figure 3.15. A 121 bp amplicon from housekeeping gene *rpoD* served as an internal control. All standard curves for each set of primer were summarized in Figure 3.16. Total mRNA was extracted from strains grown at 37 °C until to the mid-exponential phase (OD$_{600nm}$=0.5) and reverse transcription was performed using 1 μg of total RNA. The qRTPCR was then performed with 1:10 diluted into 200 μl of reverse transcribed products (cDNA). As shown in Figure 4.22, total transcripts from PR1 region presented by copy number of Amplicon 1-4 in strain pJPR1-IHF-BSM was significantly increased (P<0.01) with comparison of wild type strain P90C (pJPR1), which was consistent with previous β-galactosidase activity results. In agreement with previous results, total transcripts from P90C (pJPR1-SDM+1-IHF-BSM) were also increased notably (P<0.01) compared to strain P90C (pJPR1-Δ+1). Regarding to the transcripts copy

Figure 4.21 β-galactosidase activities of PR1 transcriptional fusions pJPR1-Δ+1 and pJPR1-SDM+1-IHF-BSM grown at 37 °C.

Values are the means of three independent experiments performed in triplicate. Error bars represent standard error.
number driven from PR1-4, inactivation of IHF binding site in strain P90C (pJPR1-IHF-BSM) resulted in a dramatic increase (P<0.05) compared to strain P90C (pJPR1), and a similar effect (P<0.01) to that also observed in strain P90C (pJPR1-SDM+1-IHF-BSM) compared with P90C (pJPR1-Δ+1) (Figure 4.22). These results provided additional evidence that the binding of IHF is acting as a negative regulatory element on PR1-4 transcription at 37 °C.

The copy number of Amplicon 1-1a and 1-1 that detects transcription from promoter PR1-1 in strain P90C (pJPR1-IHF-BSM) was also significant higher than from the wild-type strain P90C (pJPR1) (Figure 4.22), which indicating that the binding of IHF at +130 will also play a negative role in regulating transcription from promoter PR1-1 at 37 °C. Strain P90C (pJPR1-Δ+1) and P90C (pJPR1-SDM+1-IHF-BSM) consisting inactive promoter PR1-1 displayed negligible effect in the transcript copy number of Amplicon 1-1a and 1-1.

Overall, it was concluded that the binding of IHF could repress the level of transcription effectively from promoter PR1-1 and PR1-4 at 37 °C.
4.2.3.3 The binding of IHF may affect the temperature regulation at PR1-4 when the promoter PR1-4 was separated out from the PR1 region

To further analyse the role of IHF on regulating the expression of PR1-4, plasmid pJJ182A-IHF-BSM was generated by cloning a PCR fragment amplified from construct pJPR1-IHF-BSM with primer pJJ182A-F-EcoR1 and pJJ1-R-BamH1 (Table 2.3) into EcoR1/BamH1 site of pRS415. The resulting constructs were checked by DNA sequencing. This plasmid together with plasmid pJJ182 and pJJ182A were introduced into strain P90C and grown to mid-exponential phase (OD600nm=0.4-0.6) at 37 °C and 20 °C, and a β-galactosidase assay was performed respectively (Figure 4.23). Strain P90C (pJJ182) has no temperature regulation of PR1-4 with no significant difference of β-galactosidase activity between 37 °C and 20 °C, while strain P90C (pJJ182A) was temperature regulated with significant lower level of transcriptional activity at 20 °C (Figure 4.23). These data indicate that the loss of
temperature regulation of PR1-4 in strain P90C (pJJ182) was due to the loss of the IHF binding site.

The observation that strain P90C (pJJ182A-IHF-BSM) lost its temperature regulation of PR1-4 confirms that inactivation of IHF binding could potentially relieve the repression exerted from IHF at 20 °C. In another words, inactivation of IHF binding site resulted in a similar phenotype both at the capsule permissive and non-permissive temperature, significant increased transcription from PR1-4.

**Figure 4.23 PR1-4 lost temperature regulation in the absence of IHF binding.**

β-galactosidase assay were performed when strain P90C (pJJ182), P90C (pJJ182A) and P90C (pJJ182A-IHF-BSM) grown to mid-log phase at 37 °C. Values are the means of three independent experiments performed in triplicate. Error bars represent standard error.

### 4.2.4 Analysis the activation of promoter PR1-4

The mutation analysis presented in Figure 3.17 indicated that the inactivation of PR1-1 by single site mutation in strain P90C (pJPR1-Δ+1) decreased the activity of PR1-4 proportionally indicating that the promoter PR1-1 and PR1-4 may somehow interact with each other. There are two possible hypotheses: the activation of promoter PR1-4 may rely on the activation of promoter PR1-1; or alternatively, the
elongation of RNAP from PR1-1 may alter the DNA conformation by binding of IHF at PR1-4 upstream which could release the repression exerted from IHF.

To test the latter hypothesis that the elongation of RNAP driven from PR1-1 could displace IHF and release the transcription from PR1-4, the transcription copy number from PR1-4 should decreased when the elongation of RNAP was stopped just upstream of IHF binding site. To test this hypothesis, a construct carrying a Rho-independent transcription terminator sequence downstream of the PR1-1 transcription start site and upstream of the PR1-4 promoter motifs was generated in this study.

4.2.4.1 Design and characterization of artificial Rho-independent transcriptional terminators

A Rho-independent terminator was designed based on previous research of terminator structure and termination efficiency. The Rho-independent terminator sequence of MraY gene scored -9.71 (scores less than – 4.0, a more negative score is better) was selected according to the RNAMotif algorithm in the E. coli K12 genome (Lesnik et al., 2001) as the model terminator sequence. The transcription of MraY also was regulated by the σ^70 factor as same as PR1-1. Based on the transcription terminator sequence of MraY, a new artificial designed terminator for PR1-1 was designed by modifying few nucleotides to optimizing the terminate efficiency (Table 4.1). The principle of designing intrinsic terminator was based on that of d’Aubenton Carafa et al. (1990) and Lesnik et al. (2001), an 11 nt A-riched region upstream of a stable tetra-loop UUCG and followed by a 9 bp poly-T tail, and additional 50 bp of random nucleotides was added at the 3’ end of the terminator (Table 2.3) (d’Aubenton Carafa et al., 1990; Lesnik et al., 2001). The terminator efficiency was predicted to be 95% using a formula developed by d’Aubenton Carafa et al. (1990).
The artificial designed terminator was generated by a synthesized double strand DNA fragment (GeneArt Strings™, Invitrogen by Life Technology) with two ends modified by restriction enzyme *NcoI* binding site. In order to generate replicable copy of terminator for the future analysis, this newly synthesized DNA fragment was A-tailed at both ends and ligated into pGEMT-easy vector to generate construct pJJTNc13 (Figure 4.24). The successful construct pJJNc13 with terminator was checked by DNA sequencing and then digested by enzyme *Ncol/Sacl/SacII* (NEB) simultaneously before ligated into *NcoI* site in plasmid pBHA1 (Figure 4.24). Colony PCR was used to screening the successful ligated construct pBTNc13 using primer P1T and M13-R (Table 2.3). The successful construct pBTNc13 was checked by DNA sequencing and the PR1 region with 112 bp of terminator inserted downstream of promoter PR1-1 was ligated into transcriptional vector pRS415 at *BamH1* site to generate pJP1T. The correct orientation of insert in pJP1T was checked by colony PCR using primer pJJ1-F-EcoR1 and pRS415-LacZ-SEQ (Table 2.3) and DNA sequencing. A schematic of generating plasmid pJP1T can be seen in Figure 4.24 and same procedure was used for generating construct pJP1T-IHF-BSM. A nucleotide sequence of the newly synthesis construct insert was shown in Figure 4.25.

**Table 4.1 Terminator Sequence used for this study.**

<table>
<thead>
<tr>
<th>Name</th>
<th>A-rich region</th>
<th>Stem-loop</th>
<th>Loop</th>
<th>Poly-T Tail</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>MraY</em></td>
<td>TGAAGGGCTGG</td>
<td>CCCGAAACGCGCTCTATTGTGCCTTCTGG</td>
<td>TCAT</td>
<td>ATTATTTGCGTGTAT</td>
</tr>
<tr>
<td>PR1-1</td>
<td>AAATAGACTGT</td>
<td>CCCGAAACGCGCGTCTCCGCTTCTGG</td>
<td>TTCG</td>
<td>TTATTTTTGCGTGTAT</td>
</tr>
</tbody>
</table>
Figure 4. Schematic representation of the strategy used for constructing PR1 region inserted with Rho-independent terminator.
Figure 4.25 Nucleotide sequence of the 5’ end upstream region of the \textit{kpsF} gene with Rho-independent terminator at \textit{NcoI} site.

The \textit{NcoI} site was indicated by yellow highlight and \textit{BamH1} sites in grey highlight. Transcription start sites were indicated in bent arrows and their promoter elements were underlined. IHF consensus binding site was highlighted in dark green. The \textit{Shine-Dalgarno} box was colored in orange. The start codon ATG was in bold. The sequence of Rho-independent terminator was in Italic.
4.2.4.2 Transcription at promoter PR1-1 can simulate downstream promoter PR1-4 transcription activation

Three sets of primers (1_1a, 1_1 and 1_4) were used for qRT-PCR in strains pJP1T (P90C) and pJP1T-IHF-BSM (P90C) (Figure 4.26). A 121 bp amplicon from rpoD served as an internal control. The standard curve of 1_1 primer for amplicon 1-1’ with terminator was shown in Figure 4.27. The qRT-PCR was then performed with 1:10 diluted into 200 μl of reverse transcribed products (cDNA) from 1 μg of total extracted RNA. In comparison with strain P90C (pJPR1), total transcripts from PR1 region presented by copy number of Amplicon 1-4 in strain P90C (pJP1T) and P90C (pJP1T-IHF-BSM) was significantly decreased (P value = 0.0038 and P value=0.001, respectively) (Figure 4.28), which was consistent with β-galactosidase activity results with around 50% and 65% reduction respectively (data not shown).

The successful designed Rho-independent transcription terminator downstream of PR1-1 with predicted 95% terminating efficiency should abolish PR1-1 transcription proceeding through to PR1-4. It was observed that insertion of terminator just downstream of promoter in the strain P90C (pJP1T) could cause dramatically 93% decrease in the copy number of Amplicon 1-1’ compared to wild type P90C (pJPR1) (Figure 4.28). Regarding to the transcripts copy number driven from PR1-4, a significant increase (P<0.01) in the copy number of amplicon PR1-4 can be seen in strain P90C (pJPR1-IHF-BSM) was due to the released repression effect of IHF (Figure 4.28). However, no significant difference of amplicon PR1-4 copy number was observed in strain P90C (pJP1T) compared to P90C (pJPR1) (Figure 4.28), suggesting that the promoter PR1-4 was active even though the transcription from PR1-1 was terminated. Thus this result would contradict the hypothesis that RNAP initiated from PR1-1 and running through PR1 region would unblock the promoter PR1-4 by pushing the IHF out. It lead to another hypothesis that the activation of downstream promoter PR1-4 was dependent on the activation of promoter PR1-1 in vivo and the failure of open complex formation and/or following transcription steps of PR1-1 may alter the transcription activity at PR1-4.
Figure 4.26 Illustration of qRTPCR primers and corresponding amplicons performed in qRTPCR assay of strain P90C (pJP1T) and P90C (pJP1T-IHF-BSM).

873 bp of Promoter Region 1 plus downstream lacZ gene from construct pJPR1 were displayed. The name of the amplicon was shown on the left side while the corresponding lengths were shown on the right side.

Figure 4.27 Standard curves for 1_1’ amplicon in pJP1T and pJP1T-IHF-BSM performed in qRTPCR assay.

A standard curve is generated from StepOne™ software v2.3 by plotting the Ct values against the logarithm of the initial copy numbers. Amplicon length is 126 bp. Eff% presented the primers working efficiency. Linear regression equation was indicated by Y=mX+b on each corresponding standard curve. Template is a 283 bp DNA fragment amplified from pJP1T using qRTPCR primer set 1.1.
Figure 4.28 Transcription initiation from PR1-4 was dependent on the activation of promoter P1-1.

Each repeated results were normalized against the copy number of \textit{rpoD}. Values are the means ± standard error of three independent experiments performed in triplicate.
4.3 Role of SlyA in transcription of PR1 region of K5 capsule gene

Different minimal promoter-\textit{lacZ} fusions were also transformed into \textit{slyA::kan} P90C strain and \(\beta\)-galactosidase assays were performed at 37 \(\degree\)C and 20 \(\degree\)C, respectively (Figure 4.29 and Figure 4.30). At 37 \(\degree\)C, the intact PR1-\textit{lacZ} transcriptional fusion pJPR2 displayed significant decrease from 1173.1 ± 240.0 Miller Units to 499.6 ± 29.2 Miller Units (Figure 4.29), which corresponding to the previous finding that SlyA was required for transcription from PR1 at 37 \(\degree\)C (Corbett \textit{et al.}, 2007). Similar effect was also observed in construct pJJ2 and pJJ2A; however, no significant differences were found in the construct pJJ1, pJJ133-142 and pJJ182A between the wild type strain and \textit{slyA::kan} mutants. This suggested that SlyA was specific required for transcription from PR1-1 but not working on the other promoters at 37 \(\degree\)C.

At 20 \(\degree\)C, the levels of transcription activity from all constructs were very low and there was no difference of \(\beta\)-galactosidase activity between the wild type P90C and the mutant P90C \textit{slyA::kan} for all promoter-\textit{lacZ} fusions (Figure 4.30). It was indicated that SlyA did not affect the transcription at PR1 at 20 \(\degree\)C.
Figure 4.29 SlyA is required for the transcription from PR1-1 in vivo at 37 °C.

β-galactosidase activities of wild type strain P90C and P90C slyA::kan bearing different minimal putative promoter-lacZ fusions grown at 37 °C. Values are the means of three independent experiments performed in triplicate. Error bars represent standard error.
Figure 4.30 SlyA has no effect on the transcription at PR1 region in vivo at 20 °C.

β-galactosidase activities of wild type strain P90C and P90C slyA::kan bearing different minimal putative promoter-lacZ fusions grown at 20 °C. Values are the means of three independent experiments performed in triplicate. Error bars represent standard error.
4.4 Discussion

4.4.1 H-NS negatively regulating transcription from PR1

Previous results in this study indicated that 5’ UTR plays a negative role in moderating the transcription from PR1 and being important for temperature regulation of expression of promoter PR1-1. One possibility is that this region may contain cis-regulatory element for the regulatory proteins binding that are important for transcription regulation at PR1 region.

In this study, the PR1-lacZ transcription fusion in the P90C hns::kan mutant increased transcription at both the permissive and non-permissive temperatures to comparable levels, indicating that H-NS played a negative role in regulating transcription from PR1 (Figure 4.2 and 4.3). What is more, performing qRTPCR assay in the E. coli K5 strain MS101 and its hns::kan mutant further confirmed that H-NS negatively regulating PR1 transcription at both temperatures. This study was the first time revealed that H-NS repressed the transcription at PR1 of E. coli K5 gene cluster both at 37 °C and 20 °C but contradictory to the previous studies. Previous studies revealed that H-NS has a dual role, being required in the activation of Group 2 capsule gene expression at 37 °C while repressing transcription at 20 °C. Besides, the level of transcription activity from PR1 was reduced significantly in the hns::kan mutant at 37 °C, indicating that H-NS played a positive role in regulating transcription from PR1 at 37 °C (Rowe et al., 2000; Corbett et al., 2007).

In the study done by Rowe et al. (2000), the hns mutation could reduce transcription from PR1 region in pDSHcH more then 2-fold compared to wild type at 37 °C. In their study, plasmid pDSHcH served as a PR1-lacZ transcriptional fusion and transformed into strain MS152. The plasmid pDSHcH containing 1.1-kb HincII-HindIII fragment containing region 1 promoter cloned upstream of lacZ gene from promoter probe plasmid pCB192 (Schneider & Beck, 1986). In our study, we also examined the transcription activity in plasmid pDSHcH between the wild type strain P90C and its hns mutant, interestingly, the level of transcription activity at PR1 in the hns mutants was increased and further confirmed that H-NS repressed...
the transcription at PR1 (Figure 4.2 and 4.3). With comparison to the level of transcription from PR1 in strain P90C (pJPR2), it was obvious that plasmid pDSHcH displayed extremely lower activity (Figure 4.3). The reason for explaining this difference is may be due to the different strength of Shine-Dalgarno (SD) sequence of lacZ gene between the vector pDSHcH and pJPR2. Plasmid pDSHcH was constructed from the promoter-probe vector pCB192 (Schneider & Beck, 1986; Rowe et al., 2000). The SD sequence of pCB192 is [AGAGGG], while the SD sequence of pRS415 is engineered as [AGGAAA] (Simons et al., 1987). The SD sequence of pRS415 seems more efficient since there was much higher transcriptional activity was observed in the pRS415 background compared to the pDSHcH.

In the study done by Corbett et al. (2007), it was found that in the strain HA1 hns::kan contained a PR1-lacZ fusion in the chromosome, that the transcription from PR1 was reduced 50% relative to wild type. However in the strain HA1, the PR1-lacZ fusion was in λ attachment site not at the kps locus (Askar Ph.D Thesis, 2004). Thus the different chromosomal location of kps gene cluster could be critical if DNA supercoiling is important in thermoregulation and the input of H-NS. It is known that DNA supercoiling also affected the expression of some genes in response to temperature (Chen & Wu, 2003), so plasmids carrying the capsule genes may not exhibit the same thermoregulation as the chromosomal genes. H-NS itself is known to affect DNA supercoiling and to condense DNA in vitro and in vivo (Lim et al., 2014). Thus, the difference in the superhelicity between chromosomal and plasmid DNA may also affect the action of H-NS or other regulators acting by altering DNA topology. However, the qRTPCR experiments reported here looking at expression directly from the chromosome would seem to be less open to extraneous effects, and the qRTPCR results were also consistent with the results obtained in the plasmid experiments in this study.

The binding of H-NS at binding site III, downstream of PR1-1 has a more repressing effect on PR1-1 promoter activity since the transcriptional activity of pJJ2A was significantly lower than in pJJ2 (Figure 4.12). Thus in the absence of H-NS downstream binding site III, pJJ2 will cause less increase of promoter activity
compared to pJJ2A (Figure 4.7). In addition, at 20 °C, it was shown that in pJJ2A had significantly increase of β-galactosidase activity in the P90C *hns::kan* mutant compared with wild type P90C (Figure 4.8). This indicates that the binding of H-NS at the downstream region of +1 at PR1-1 will exert a greater repressing effect on PR1-1 at 20 °C. Also it was observed that either the upstream H-NS binding site or downstream H-NS binding site is absent (pJJ2 and pJJ2B), the PR1 is no longer repressed by H-NS at 20 °C (Figure 4.12). It was suggested that the temperature regulation at PR1-1 by H-NS might need both upstream and downstream region of promoter PR1-1. It is known that H-NS binds preferentially to AT-rich intrinsically curved DNA sequences located either upstream or downstream of promoters (Singh & Grainger, 2013). These sequences serve as nucleation sites, leading to the polymerization of H-NS on the DNA template, and the formation of higher order nucleoprotein complexes results in the repression of the target promoters (Dame *et al.*, 2000; Rimsky, 2004). If this is the case, one could hypothesize that the binding of H-NS to regions up- and downstream- of the transcription start site of PR1-1 may enable bridge formation, DNA looping and hence inhibit transcription (Becker *et al.*, 2007). Thus it was speculated that the H-NS loop formation can trap the RNAP at PR1-1. Binding of the H-NS at upstream region of transcription start site, cooperatively stabilized by another H-NS dimer bound to downstream binding site, resulting in a repression by blocking the RNAP in the loop (Shin *et al.*, 2005). Or alternatively, the loop which was stabilized by the interaction between H-NS molecules forms an independent topological domain and the changed conformation of the promoter region may be altered in such a way that it is no longer able to drive transcription initiation.

The β-galactosidase activity from minimal promoter-*lacZ* fusions pJJ182, pJJ182A and pJJ133-142 in the *hns::kan* mutants were all significantly increased compared to the wild type both at 37 °C and 20 °C (Figure 4.7 and 4.8), indicating that H-NS might also negatively regulate the transcription from these additional promoter as well. However, the EMSA assay shown there was no H-NS binding sites at UTR region (Figure 4.10), which indicating that the negatively regulation by H-NS at the additional downstream promoters were indirectly effect. Based on the previous
hypothesis that promoter PR1-4 was a supercoiling-sensitive promoter. It was speculated that the binding of H-NS may induce or change the supercoiling of downstream DNA region and then inhibit the transcription indirectly.

H-NS is a global negative regulator that represses the transcription of more than 200 genes transcription in *E. coli* (Dame *et al.*, 2002; Rimsky, 2004; Nagarajavel *et al.*, 2007). Therefore, it was suggested that H-NS more likely acted as a transcription silencer rather than an activator at the *E. coli* K5 PR1 region. H-NS is known to affect DNA topology as well as DNA structure and it would not be surprising if the role of H-NS in PR1 repression is in both functions. Taken together, these data provide compelling evidence that H-NS is repressed the transcription from the region 1 promoters both at 37 °C and 20 °C directly at promoter PR1-1 but indirectly at promoter PR1-3 and PR1-4.

### 4.4.2 SlyA may antagonizes H-NS-mediated silencing at PR1

Previous study have been identified that SlyA functions as a transcriptional activator from PR1 region in *E. coli* K5 capsule gene cluster at 37 °C (Corbett *et al.*, 2007). The observation that the transcription from PR1-1 was reduced 50% in the slyA mutants compared to wild type at 37 °C (Figure 4.29) confirms this hypothesis. However, SlyA had no effect on the other additional promoter downstream of PR1-1 at 37 °C (Figure 4.29) suggesting that SlyA acts on PR1-1 alone in the PR1 promoter region. There was no difference in the transcription activity of PR1 region between the wild type strain and slyA::kan mutant at 20 °C might due to SlyA was self-regulated with much less expression at 20 °C keeping with previous studies (Corbett *et al.*, 2007).

Based on the previous study done by Corbett *et al.* (2007), a potential mechanism was proposed by which the regulation of PR1 dependent on the concentration ratio between the SlyA and H-NS. This hypothesis of H-NS role in capsule gene regulation in *E. coli* would seem to fit the proposed of SlyA/RovA behaving as an H-NS anti-repressor. The functional interplay between H-NS and SlyA is similar to that observed at other promoters regulated by SlyA and its *Yersinia spp.*
homologue RovA. RovA/SlyA has been postulated to alleviate H-NS-mediated repression through competition for binding sites (Heroven et al., 2004; Wyborn et al., 2004). SlyA activates the hlyE gene expression by antagonizing H-NS repression of hlyE expression, apparently by steric hindrance when the two proteins compete for overlapping operators (Wyborn et al., 2004). SlyA can also activate fimB gene expression by antagonizing H-NS repression in E. coli K12 (McVicker et al., 2011).

The multiple copy of slyA was able to increase PR1 transcription 12-fold at the capsule non-permissive temperature (Corbett et al., 2007). Increasing of the H-NS concentration in the presence of high concentration of SlyA affected DNaseI sensitivity, indicating that the concentration ratio between H-NS and SlyA did influence the local promoter architecture and resulting in a structurally altered nucleoprotein complex at PR1-1 (Corbett et al., 2007). In addition, competitive footprinting studies performed on the hlyE promoter done by Lithgow et al. (2007) suggested that SlyA could displace H-NS at high enough concentrations. It is known that H-NS is very abundant while the levels of SlyA have not been quantified but are likely to be much lower. Therefore, in the model it was proposed at 37 °C, SlyA would preferably bind to the region where H-NS occupied and counteract with H-NS and play an anti-repression role. At least two models have been proposed for how SlyA antagonizes H-NS binding that SlyA may disrupt the H-NS–DNA silencer complex by altering the local DNA conformation and/or activate transcription through displacement of H-NS from the promoters (McVicker et al., 2011).

With respect to temperature regulation, the observation that H-NS bound PR1 equally at both temperatures (Corbett et al., 007) would suggest that the reduced expression of SlyA at 20 °C will alter this ratio and H-NS predominantly binding at PR1 hence the expression of kpsF is silenced at 20 °C.

**4.4.3 The role of IHF in regulating PR1 transcription**

The 5’ UTR region of PR1 is also a target for regulation by the global regulator IHF. It is known that IHF can bind to the 5’ UTR region at +130 and is essential for transcription from PR1 under wild type conditions, but that in the absence of H-NS,
IHF is largely redundant, having only a modest effect on transcription (Corbett Ph.D thesis, 2007). Thus it seems like the principle role of IHF appears to be one of the H-NS anti-repression, but the mechanism by which IHF activates transcription at PR1 is unclear.

IHF was initially found to be required for maximum transcription from PR1 by performing β-galactosidase assay between the wild type and ihf mutants. It was shown that in the absence of IHF, transcription from PR1 was reduced significantly compared to wild type (Figure 4.13). In addition, when the IHF binding sites was disrupted by site direct mutagenesis, either the β-galactosidase activity from the PR1-4 or the transcript copy number both coming from PR1-1 and PR1-4 was increased significantly (Figure 4.21 and Figure 4.22). All of data suggested that IHF plays a dual role in regulating the transcription from PR1 of E. coli K5 gene cluster, which being positively involved in the regulating of transcription of PR1 indirectly but repressing the transcription directly by binding at +130.

If IHF represses the transcription from PR1-4 directly, when IHF binding site was abolished in the strain P90C (pJPR1-IHF-BSM), the transcription activity of pJPR1-IHF-BSM should also be increased at 20 °C. However, the transcription from PR1 of strain P90C (pJPR1-IHF-BSM) was still being repressed at 20 °C with extremely low activity (Figure 4.20). One of the explanations would be related to the transcription activation of PR1-4 was coupled with PR1-1 transcription initiation. PR1-4 was silenced in the IHF binding site mutant since the upstream promoter PR1-1 was still repressed at 20 °C. In addition, inactivation of IHF binding site at +130 resulted in promoter PR1-4 lost temperature regulation (Figure 4.23), which indicated that temperature regulation of promoter PR1-4 required IHF.

How does the binding of IHF regulate the transcription at PR1 region? Firstly, IHF binding overlap with the consensus promoter elements of the PR1-4 promoter, thus it was reasonable that the binding of IHF could occlude the binding of RNAP at PR1-4 and inhibit the transcription at PR1-4 in sterical hindrance effect (Tsui et al., 1991). Secondly, it is already known that the IHF protein introduce a 160° – 180° bend into DNA upon binding (Rice et al., 1996). Thus it is speculated that the
purpose of IHF binding at downstream of PR1-1 is to bring regulatory elements present the downstream region into close apposition with the promoter PR1-1. In this loop formation model of H-NS, the IHF is likely to involve in the repression activity exerted by H-NS. The DNA bending protein IHF seems necessary to enable loop forming by inducing DNA curvature or may be additional regulatory elements modulating loop stability. Therefore, the binding of IHF at +130 may induce such a sharp bend at PR1 region and promote the H-NS:DNA bridge/loop formation at promoter PR1-1, and hence inhibit transcription from RP1-1 and subsequently from PR1-4. Lastly, explanation for IHF negatively regulating transcription from PR1-1 and PR1-4 could involve the IHF may affect the DNA supercoiling. It was suggested that some nucleoprotein complexes can form topological barriers that prevent the diffusion and merger of chromosomal supercoil domain (Fulcrand et al., 2013). The topological barrier could block supercoil diffusion and divided the DNA molecule into two independent topological domains (Leng et al., 2011). Thus it is possible that the binding of IHF functions as a topological barrier to modulate localized DNA supercoiling and hence IHF may prevent transcription initiated supercoiling diffusing downstream to activate supercoiling-sensitive promoter PR1-4.

4.4.4 Transcription from promoter PR1-4 is dependent on promoter PR1-1 activation

In this study, it was observed that the transcription from promoter PR1-4 was dependent on the activation of PR1-1 since the destruction of promoter PR1-1 would cause a proportional attenuation of PR1-4 transcription level (Figure 3.17). Later on, experiments were designed to investigate whether this promoter activation resulted from the repression effect of IHF. It was speculated that the RNAP elongation complex from PR1-1 may run through downstream of UTR region to displace the binding of IHF at +130 and release the inhibitory effect from IHF, and thereby activate the transcription from PR1-4. However, it was shown that the transcriptional activity from PR1-4 displayed no difference between the wild type P90C (pJPR1) and the strain with terminator P90C (pJP1T) (Figure 4.28), suggested
that the activation of PR1-4 may dependent on the activation of PR1-1, such as the open complex formation, or the following elongation complex moving downstream towards PR1-4, rather than the IHF repressed effect.

Two tandem promoters are regulated in such way may potentially subject to the kinds of effects that strong promoter generate significant over-supercoiling and such topological coupling between promoters might generate promoter interference leading to co-operatively between promoter pairs (Lilley & Higgins, 1991). Being considered the fact that twin-domain model of Liu and Wang (1987), which demonstrated that closely spaced divergent promoters must be transcriptionally coupled. The transcription-induced DNA supercoiling may be the source of the local DNA supercoiling for supercoiling sensitive promoter activation. The best illustrated studies are the topological coupling at the leu-500 promoter mutation of the leuABCD operon in S. typhimurium topA mutant (Fang & Wu, 1998b) and the promoter PilviH activates an upstream tandem cryptic promoter (PleuO) of leuO gene (Fang & Wu, 1998a). Previous studies have demonstrated that promoter-promoter interaction via localized DNA supercoiling generated by RNA transcriptional processes is normally short-range (around 250 bp) (Tan et al., 1994). The distance between PR1-1 and PR1-4 is around 180 bp, and the transcription from PR1-1 can influence the activity of adjacent promoter PR1-4 on the same DNA chain though local supercoiling is therefore reasonable. Thus, topological coupling between promoters may be biological important. Two tandem promoters whose activity is stimulated by supercoiling will act co-operatively. The promoter PR1-4 located downstream of PR1-1 might be activated by positive supercoiling induced by transcription from PR1-1. However, transcription activation of PR1-1 was independent on the transcription activation from PR1-4 since there was no significant difference of promoter PR1-1 activity between the wild type and the mutant with mutation at promoter PR1-4 -10 hexamer (Figure 3.18). One possibility may be explained that transcription from PR1-4 did not significantly affect positive supercoils generated by PR1-1 activity. Since the promoter PR1-1 is a very strong promoter, therefore the negative supercoiling
generated by transcription from PR1-4 will not significantly diminish the number of supercoils generated by transcription from PR1-1.

In addition, the open complex formation involves DNA untwisting, bending or wrapping around RNA polymerase. It is proposed that DNA supercoiling changes the structure of the promoter DNA to a form that can be recognized by RNA polymerase and then easily melted. The binding of RNAP to promoter DNA would be very sensitive to the relative orientation of the -10 and -35 regions, which is itself dependent on the (supercoiling-sensitive) twist of the spacer DNA (deHaseth and Helmann, 1995). In the lac P5 promoter the dependence on negative superhelicity is reduced by gain-of-function mutations in the -10 or -35 hexamers and also the spacer length (Borowiec & Gralla, 1987). There are classes of promoter, such as the MerR-dependent promoter, where recognition of these promoters by RNAP is exquisitely sensitive to the exact angular orientation of the two promoter elements (Brown et al., 2003). The promoter PR1-4 contains an unusual 22 base pair spacer, which makes it a rather inefficient promoter. It was speculated that the binding of transcription regulator might change the DNA twist between -10 and -35 elements by overwinding of the spacer DNA, resulting the elements much closer to the consensus promoter. This allosteric overwind of the spacer sequence brings the phases of the -10 and -35 elements into a location comparable with the consensus promoter structure with normal spacer length.

Based on this, a hypothesized model of transcription activation of PR1-4 is proposed in Figure 4.31.

Taken together, the transcriptional activation itself of PR1-1 is important in activating the downstream PR1-4, perhaps via a mechanism of transcription-driven supercoiling as in the short-range promoter-promoter interaction.

In conclusion, in this chapter, we demonstrated that H-NS is a transcription repressor that represses the transcription from PR1 both at 37 °C and 20 °C. H-NS represses transcription from PR1 region directly, which required H-NS binding both upstream and downstream of promoter PR1-1. However, H-NS is not binding in vitro at UTR region from +39 to +208 and the inhibitory effects on promoter PR1-3
and PR1-4 are likely to be indirect. The H-NS anti-repressor SlyA functions as a transcriptional activator from PR1 region but acts on PR1-1 alone at PR1. IHF plays a dual role in regulating the transcription from PR1 of *E. coli* K5 gene cluster, which being positively involved in the regulating of transcription of PR1 indirectly but repressing the transcription directly by binding at +130. Additionally, we observe that the transcription from promoter PR1-4 is dependent on the activation of PR1-1. Further prove this promoter dependent manner is not related to the relief of IHF repression.

![Diagram](image)

**Figure 4.31** A hypothesized model of activation of PR1-4 is dependent on the transcription activation of PR1-1.

The non-optimum promoter spacer region of PR1-4 was not recognized by RNAP efficiently. Once the transcription initiated at PR1-1 and the template DNA melting occurred, resulted in increasing of downstream DNA template superhelical density. This increased superhelical density of a template directly affects DNA twist or topology. Thereby allow the promoter element orientation to a closer match to the consensus promoter elements at PR1-4 and maximum the transcription activation from PR1-4.
Chapter 5. Studying the transcription from multiple promoters at PR1 during growth phase

5.1 Analysis the time courses of multiple promoters transcriptional activity at PR1 in E. coli K5 strain

Previous results in this study identified three additional promoters (PR1-2, PR1-3 and PR1-4) in the UTR region of kpsF. Additionally, it was shown that PR1-4 was a major promoter that contributed 50% of the total transcripts into kps gene in the mid-log phase at 37 °C. However, why these tandem multiple promoters exist and how they play a physiological role for bacteria growth are still unclear. Therefore, we first speculated that these multiple promoters might function at different growth time points. To study the time courses of the transcriptional activity of different promoters at PR1, strain P90C (pJJ2A), P90C (pJJ133-142) and P90C (pJJ182) (Figure 3.10, Table 2.2) were grown at 37 °C in LB-broth overnight and 1:100 dilute into pre-warmed 50 ml fresh LB-broth. Cultures were taken out at specific time points and snap frozen in liquid nitrogen by mixing with Z-buffer without chloroform. After collecting all the time points, β-galactosidase activity were assayed after adding chloroform for each reaction. Figure 5.1 shown that the transcriptional activity increased sharply when strains P90C (pJJ2A) reached late exponential phase (after time point=3hrs, OD600_{nm}=1.4) and increased until early-stationary phase of growth (time point=6hrs, OD600_{nm}=2.0) at 37 °C. In strain P90C (pJJ182A) β-galactosidase activity increased steadily during all the time points (Figure 5.1) in contrast to strain P90C (pJJ133-142) showed an increase at 6 hours with a modest rise to the end of the experiment (Figure 5.1). Both strains expressed significantly less activity than P90C (pJJ2A) (Figure 5.1).

The control strain P90C (pRS415), expressed very low levels of β-galactosidase activity through the whole growth curve indicating that the empty vector yielded no β-galactosidase activity. The β-galactosidase activity of each time point in each strain was significantly higher than the corresponding negative control’s value in P90C (pRS415). These results indicated that when E. coli bacteria were grown at 37
at 37°C continuously, the promoter PR1-1 maintained high activity during the growth, while the transcription activity of PR1-2, RP1-3 and PR1-4 retained at low level for all time points.

Figure 5.1 Analysis of multiple promoters at PR1 transcriptional activity during growth at 37°C.

(A) β-galactosidase activity generated by, and (B) growth curve of, P90C (pRS415), P90C (pJJ2A), P90C (pJJ133-142) and P90C (pJJ182A) at 37°C. Values are means of three independent experiments performed in triplicate. Error bars represent standard error.
5.2 Examining transcriptional level of PR1-1 and PR1-4 at PR1 in strain P90C (pJPR1) during growth phase at 37 °C by qRTPCR

The results detailed in Figure 5.1 presented that the promoters start working at different time point. However, since the accumulation of β-galactosidase in the cells, this result could not present the changing of transcriptional activity from each promoter in real time. Therefore, as an alternative procedure, we use qRTPCR to verify the transcription level changes in tandem promoter PR1-1 and PR1-4 during the growth curve.

Same as previous described, four sets of primers (1_1, 1_1a, 1_4 and rpoD) used for amplifying the different target region were used and are detailed in Figure 5.2. All standard curves for each set of primer were summarized in Figure 3.16. Overnight culture of pJPR1 (P90C) was 1:100 diluted into pre-warmed 50 ml of fresh LB-Broth supplemented with Ampicillin and grown at 37 °C continuously. Total mRNA was started harvesting at 30 min intervals after inoculation and the reverse transcription was performed using 1 µg of total RNA. The qRTPCR assay was then performed 10 µl per reaction with 1:10 diluted into 200 µl of reverse transcribed products (cDNA). As shown in Figure 5.2, the detected transcriptional copy number from PR1-1, PR1-4 and total transcript into kpsF gene were presented along with corresponding growth point at 37 °C. The copy number of amplicon 1-1a represented the transcripts just transcribed from promoter PR1-1 while amplicon 1-1 represented the transcripts from promoter PR1-1 running after 200 bp downstream of PR1-1. Total transcripts from PR1 region was represented by amplicon of kpsF that detected by primer set 1_4 in qRTPCR assay. Transcripts from promoter PR1-4 were obtained by subtracting the copy number of 1-1 from the copy number of 1-4.
Figure 5.2 Transcripts level of PR1-1 and PR1-4 measured by qRTPCR in strain pJPR1 (P90C) during growth phase at 37 °C.

The corresponding primer sets were shown above. Total RNA was isolated from cell grown at 37 °C and remove after 0.5h (OD=0.028), 1h (OD=0.096), 1.5h (OD=0.198), 2h (OD=0.392), 2.5h (OD=0.637) and 7h (OD=2.191). Data are representative of one independent experiment performed in triplicate.

As seen from the Figure 5.2, the transcripts copy number of amplicon 1-1a always higher than amplicion 1-1, which was in agreement with previous observation that 250 bp UTR downstream of PR1-1 would lower the amount of transcription that exits the UTR. The transcription from promoter PR1-1 (represented by 1-1) was predominantly expressed through the whole growth curve. In the mid-exponential phase (OD=0.637), promoter PR1-1 (represented by 1-1) and PR1-4 shown equivalently transcriptional activity which was consistent to with the previous conclusion that promoter PR1-1 and PR1-4 contributed evenly in the mid-
exponential phase at 37 °C (Table 3.3; Figure 5.2). The transcript copy number of 1-1a and 1-1 was dropped simultaneously from 1.5 hours to 2.5 hours, which indicating that the transcriptional activity of PR1-1 was decreased when the cell grown into mid-exponential phase (Figure 5.2). In the stationary phase (OD600nm=2.194), the level of total transcripts was the lowest point during the growth curve and the majority of total transcript was contributed from promoter PR1-1.

Taken together, it was indicated that the level of transcription of different promoters in the UTR region presented different patterns along the growth course (Figure 5.2). For instance, the total transcripts of kps started decreasing after the mid-exponential phase. There was a clear time-related difference in the pattern of PR1-4 transcripts, being a delay in the onset of increased expression levels compared to the expression of PR1-1. However, the actual early transcription level of new generation of *E. coli* was hard to measure due to the more diffuse and variable expression with the mixture culture of overnight culture and new generation of *E. coli* in fresh media.

### 5.3 Examining the transcriptional activity of PR1-1 and PR1-4 at PR1 during the growth phase following a temperature shift from 20 °C to 37 °C

Next I wanted to explore how these multiple promoters were modulated following temperature upshift from the capsule non-permissive temperature of 20 °C to the 37 °C, the temperature experienced when growing in the host. This involved using qRT-PCR analysis of PR1 activity in the *E. coli* strain UTI89 grown under a variety of conditions.

Three sets of primers used for amplifying the different target region which are located just downstream of the transcription start sites were designed (Figure 5.3). One 169 bp of amplicon was amplified by primers (16s-qPCR-F and 16s-qPCR-R) from housekeeping gene 16s which served as an internal control. All primers were detailed in Table 2.3 and the corresponding amplicons were shown in Figure 5.3. 188 bp length of DNA fragment which was amplified from UTI89 by PCR using
primers UTI89_1_1a_F and UTI89_1_1R (Table 2.3) as template of qRTPCR primer sets UTI891_1a and UTI891_1, while 168 bp of DNA amplified by primer UTI89_kpsf_F and UTI89_kpsf_R (Table 2.3) was used as template of qRTPCR primer sets UTI89_kpsf. The qRTPCR was then performed with 1:10 diluted into 200 μl of reverse transcribed products (cDNA) as described previously. The absolute standard curves for each individual amplicon were shown in Figure 5.4 and no primer dimer artefacts were observed.

Single colony grown at 37 °C and 20 °C on LB-Agar plate was inoculated in 50 ml LB-Broth respectively and parallel cultures were grown at 37 °C and 20 °C respectively. For the cells grown at 37 °C, RNA was extracted once the cell density reached OD600 nm=0.1 and then samples were taken at intervals of 30 min (Figure 5.5). To determine the temperature regulatory effect on promoter activity, the levels of transcriptional activity were also measured when the cell cultures were shifted up to 37 °C from 20 °C by qRTPCR (Figure 5.6). In the temperature upshift assay, the culture was grown at 20 °C until OD600 nm=0.1 as the first time point (T0) and immediately shifted to pre-warmed 37 °C water bath before then taking samples at 30 min intervals.

As seen in Figure 5.5, extremely low transcript copy number was detected when cells grown at 20 °C with comparison of bacteria grown at 37 °C (Figure 5.6), which matched the previous results that significant lower β-galactosidase activity was observed in strain P90C (pJPR1) at 20 °C.

The total transcription reaching the kpsF gene (amplicon UTI89-kpsF) was maximal at the mid-exponential phase (Figure 5.6). The same results were obtained for the bacteria grown in the 37 °C shift up experiment (Figure 5.7). However, the promoters responsible for transcribing the kps gene were significantly enhanced more than 3-fold when grown continuously at 37 °C compared to when the culture was upshifted 37 °C assay.
Figure 5.3 Illustration of qRTPCR primers and corresponding amplicons performed in qRTPCR assay of strain UTI89.

871 bp of Promoter Region 1 plus downstream kpsF gene in UTI89 was displayed. The name of the amplicon was shown on the left side while the corresponding lengths were shown on the right side.

Figure 5.4 Standard curves for 1-1a, 1-1, 1-4 and 16s amplicons performed in qRTPCR assay.

A standard curve is generated from StepOne™ software v2.3 by plotting the Ct values against the logarithm of the initial copy numbers. Amplicons lengths are 76 bp (A), 188 bp (B), 168 bp (C) and 169 bp (D), respectively. X-axis Unit is nM/reaction. Eff% presented the primers working efficiency. Linear regression equation was indicated by Y=mX+b on each corresponding standard curve.
When the cells grown in 37 °C continuously, the transcription of PR1-1 increased sharply during early to mid-exponential phase but then decreased steeply when the cells went into early stationary phase, whereas the PR1-4 promoter activity just started increasing at the time point when the PR1-1 transcription getting decreased at the OD=0.548 (Figure 5.6). Interestingly, same increasing trends for each promoter PR1-1 and PR1-4 could also be observed with temperature upshift assay in Figure 5.7. Except the fact that the promoter PR1-4 was start expressing immediately after cells were shift up from 20 °C to 37 °C (Figure 5.7) while PR1-4 was kept inactive until mid-exponential phase when bacteria continuously grown at 37 °C (Figure 5.6). This data indicated that the promoter PR1-4 might be stimulated to be functional by extracellular signals such as temperature upshift from 20 °C to 37 °C.

The qRTPCR revealed a fluctuation in expression levels at different time points of kps genes along with bacteria growth, indicating that different parameter modes of promoter activity related to growth processes are operating at different times. Overall, the direct quantitative transcript analysis of different time points from strain UTI89 implied that the virulence gene kps expression was differentially maintained over long-term growth at a given temperature; the transcription at PR1 was most active in the exponential phase and start decreasing to a steady low level when cells growing into stationary phase; PR1-4 can initiate mRNA synthesis at the first stage when the bacteria upshift to 37 °C.

What is more, the transcripts copy number detected by qRTPCR in upshift assay (Figure 5.7) was nearly 3-fold lower than the bacteria grown at 37 °C continuously (Figure 5.6). It was indicated that when the bacteria moved from non-permissive temperature to permissive temperature, the initial expression of first generation of activated capsule expression was folds lower than the bacteria had been adapted to physical environment at 37 °C for generations.
Figure 5.5 Time course of PR1 promoter transcription when the cells grown at 20 °C continuously.

Total RNA was isolated from cell grown at 20 °C continuously until OD=0.115 and then remove after 3h (OD=0.648), 6h (OD=1.813) and 8h (OD=2.009). Data are representative of a single experiment with technical quadruplicate; the second biological repeat shows the same trend.

---

Figure 5.6 Time course of PR1 promoter transcription when the cells grown at 37 °C continuously.

Total RNA was isolated from cell grown at 37 °C until OD=0.098 and then remove after 30min (OD=0.252), 60min (OD=0.548), 90min (OD=1.086), 120min (OD=1.564) and 6h (OD=2.168). Data are representative of a single experiment with technical quadruplicate; the second biological repeat shows the same trend.
Figure 5.7 Time course of PR1 promoter transcription following a temperature upshift from 20 °C to 37 °C.

Total RNA was isolated from the cells grown at 20 °C until OD=0.102 and then up-shifted to 37 °C and removed after 30min (OD=0.34), 60min (OD=0.635), 90min (OD=1.19), 120min (OD=1.635) and 6h (OD=2.158). Data are representative of a single experiment with technical quadruplicate; the second biological repeat shows the same trend.

In addition, as shown from Figure 5.7, transcription of kpsF gene was immediately driven by PR1-1 and PR1-4 when bacteria up shift to 37 °C growth condition, implying that E. coli start expressing capsule gene once the cell up shift from non-permissive temperature to permissive temperature. To support this finding, the expression of capsule gene was examined by immunofluorescent microscopy of E. coli K1 strain UTI89 using monoclonal K1 – specific antibody following growth in LB-Broth. It is known that no capsule biosynthesis takes place at 20 °C (Roberts, 1996), but following upshift to 37 °C, capsule biosynthesis occurs. Therefore, cultures were grown at 20 °C until OD600ₙₘ=0.1 and then extracted following upshift to 37 °C at every 30 min intervals which corresponding to the qRTPCR time points. As seen from Figure 5.8, just 30 min after upshift to 37 °C (T1), the first cell surface polysaccharide (stained in red) was detectable surrounding partially of the cell surface, but still a number of bacteria were not expressed capsules. After 30 min later (T2), K1 polysaccharide was observed in much more number of cells to
be expressed on the surface of the bacteria. More and more bright red fluoresce signals were observed from time point T3 to T5, indicating that expression of K1 polysaccharide was more likely enriched on the surface of bacteria.

The detected immunochemistry signal at T1 after 30 min upshift to 37 °C indicated that there was sufficient capsule expressed effectively and rapidly once the cell moved into 37 °C, which corresponding to the previous qRT-PCR result that kps operon was immediately transcribed once the cells were upshifted to 37 °C (Figure 5.7).
Figure 5.8 Immunofluorescent microscopy of strain UTI89 using K1-specific monoclonal antibody following a temperature upshift from 20 °C to 37 °C.

(A) Growth curve of UTI89 were grown at 20 °C (T0) and then shifted to 37 °C from T1. (B) Cells were removed at T0 and upshift to 37 °C after 30 min (T1), 60 min (T2), 90 min (T3), 120 min (T4) and 7 h (T5) before being fixed and analysed by immunofluorescent microscopy. DNA was stained with DAPI which appeared blue. K1 capsule was stained with monoclonal K1 antibody 735 and followed with AlexFluro Donkey-αMouse-TexasRed (Abcam) that appeared red. Photos were taken at 600× magnification. The scale bar was indicated in 10 μm.
5.4 Discussion

The first attempt trying to understand the time course of transcriptional activity of these multiple promoters was carried out by measuring the β-galactosidase activity from each minimal promoter-lacZ transcriptional fusion (Figure 5.1). However, β-galactosidase is a particularly stable enzyme. In Bacillus circulans, the half-life of β-galactosidase could last for 13 h at 40°C (Warmerdam et al., 2013). Therefore, the β-galactosidase assay is not an ideal method to measure the transcriptional activity during time course since the accumulation of β-galactosidase in the cell. From the β-galactosidase activity during the growth curve, the transcriptional activity increase sharply when the cells going to the stationary phase after 4 hours (Figure 5.1), implying that maybe some regulatory cis-acting factor start functional at that time point. The expression of SlyA increased rapidly during growth, with maximal level at the late exponential phase but start declining though stationary phase (Corbett et al., 2007). Thus it was suggested that the increased transcriptional activity might be related to the molecule number of SlyA in the cell. Therefore, β-galactosidase assays were performed in the strain P90C slyA::kan and P90C ihf::cm containing different minimal PR1-lacZ promoter fusions. The levels of transcription activity at late-exponential and stationary phase were measured along the growth courses both in wild type and mutants, respectively. There were no significant difference between the wild type and mutants (data not shown). It was known that β-galactosidase could accumulate in the cell along the growth phase. Hence, even though there was reduced level of transcription in the mutants, it was still difficult to detect the difference in the late-exponential or stationary growth phase.

The transcription activity from individual promoters was also measured from the strain P90C (pJPR1) by qRT-PCR with sub-cultured 37°C overnight cultures into fresh LB-Broth (Figure 5.2). Interestingly, during the first 30 min, a rapid decrease of the level of transcription from promoter PR1-1 and PR1-4 was observed (Figure 5.2). When the cells encountering the nutritional shift-up from the overnight culture into fresh media, the nutritional shift can cause a very rapid (within 2-3 min)
increase of DNA supercoiling followed by a slow relaxation (Balke & Gralla, 1987). Besides, the promoter activity is dependent both on the topological state of the promoter and the ability of the polymerase recognition (Travers & Muskhelishvili, 2005). Thus this unusual reduction from PR1 may be due to the changing of DNA conformation and the variation in the DNA topology could alter the efficiency of transcription initiation from PR1 region.

It was obvious that a clear time-related difference in the pattern of PR1-4 transcripts, being a delay in the onset of increased expression levels compared to the expression of PR1-1 either in the plasmid or in the chromosome (Figure 5.2, 5.6 and 5.7). One explanation for this phenomenon is may be due to the activation of PR1-4 is dependent on PR1-1 transcription activation and this is consistent with the previous observation in chapter 4. In addition, during the early exponential phase, the transcription from PR1-1 was extremely higher and contributed the most majority of the total transcripts into kpsF. Promoter interference can also occur with tandem promoters when a high initiation frequency of an upstream promoter ‘occludes’ a downstream promoter (Adhta and Gottesman, 1982; Zhang and Bremer, 1996). Thus it was reasonable to consider that the activation of PR1-1 stimulate the transcription initiation of promoter PR1-4 but the high initiation frequency of upstream promoter PR1-1 could somehow inhibit the transcription efficiency of promoter PR1-4 and once the transcription efficiency of PR1-1 was reduced, the ‘occluded’ promoter PR1-4 then start increasing its transcription.

After transcription from promoter PR1-4 was initiated, in the mid-log phase, PR1-4 had equal activity to PR1-1 on a plasmid pJPR1 (Figure 5.2), but it was silent in the mid-log phase in the UTI89 chromosome at 37 °C (Figure 5.6). Combined with the previous study demonstrated that activation of Pleu-500 is lost when the promoter is removed from its original chromosomal location (Richardson et al., 1988) suggesting that local supercoiling may be responsible for Pleu-500 activation. Therefore, one possibility to explain this difference is that chromosome supercoiling dynamics may play a role in this type of gene expression regulation. In the 37 °C upshift assay, it was notable that when the bacteria moved from non-
permissive temperature to permissive temperature, the initial expression of first generation of activated capsule expression was about 3-fold lower than the bacteria had been adapted to physical environment at 37 °C for generations (Figure 5.6 and 5.7).

Of particular interest was the observation that the PR1 promoters altered in transcription along with the growth stages and the total transcripts were start decreasing after it had maximum level of transcription in the mid-exponential phase (Figure 5.6 and 5.7) no matter the strain was constantly grown at 37 °C or in the 37 °C upshift condition. The fluctuation in expression levels at different time points of kps genes along with bacteria growth, indicating that different parameter modes of promoter activity related to growth processes are operating at different times (Travers & Muskhelishvili, 2005). It is possible that the different concentration of transcription regulator, such as the nucleotide-associated protein present during the growth curve in the cells may affect the level of transcription. For example, the molecule number of IHF started to increase when the cell started growing and got maximum in the early stationary phase (55,000 monomers per cell) (Ali Azam et al., 1999); H-NS had maximum expression (20,000 molecules per cell) in the exponential phase but decreased to 40% of the maximum at the late stationary phase (Ali Azam et al., 1999). In E. coli, the average negative superhelical density highest in early exponential phase and then gradually declines in late exponential phase and transition to stationary phase and it was proposed that the supercoiling dynamics of chromosomal DNA are governed by the relative abundances of the nucleoid-associated proteins (Travers & Muskhelishvili, 2005). Hence, the variable nucleoprotein complexes can modulate DNA topology during the growth cycle (Sobetzko et al., 2012) and the bacterial gene transcription is specifically affected by changes in supercoiling (Free & Dorman, 1994; El Hanafi & Bossi, 2000; Chen & Wu, 2003). Therefore, this phase-dependent gene transcription from PR1 region may relate to the general negative superhelicity and this supercoiling level are also affected by growth phase and a variety of environmental stimuli. Transcription regulation of the PR1 promoters is extremely complex, but the results described here agree well with the notion that PR1-1
plays an important role in transcription of \textit{kpsF} gene in cells growth phase. It was hypothesized that the transcription-driven supercoiling from the PR1-1 served as a signal to turn on the promoter PR1-4. This increased transcription from PR1-4 might subsequently attempt to maintain the level of total transcripts due to the levels of transcription from promoter PR1-1 decreasing steeply.

The transcription of \textit{kpsF} gene was immediately driven by PR1-1 and PR1-4 when bacteria up shift to 37 °C growth condition, implying that \textit{E. coli} start expressing capsule gene cluster that responsible for polysaccharide transport once the cell up shift from non-permissive temperature to permissive temperature. The immunofluorescent microscope results further demonstrated that after 37 °C shift up, the K1 polysaccharide immediately appeared on the surface of bacteria (Figure 5.8). Thus it was implied that once the pathogen \textit{E. coli} infected the host body it would sense the changing environment rapidly and thus activating related genes to survive and infect within the human body. The increased temperature could act as a ‘danger signal’ and enhancing its defence against human immune killing. It is advantageous to many bacteria to make their capsule expression immediately so that ensure at least a percentage of any given population of cells is capable of adhering to host cell tissues. No mechanism for how this might be achieved is apparent at this stage, but it would be interesting to investigate whether the same phenomenon occurs during growth in artificial urine medium or during infection with cultured epithelial cells. In addition, the capsular polysaccharide biosynthesis and export was demonstrated that localized at the poles of the cell in strain MS101 (K5⁺) (McNulty \textit{et al.}, 2006). In this study, it was first evidence that the localization of K1 capsular polysaccharide was rapid diffusion and spread over the cell surface of UTI89 (Figure 5.8).

Overall, in this chapter, we revealed a fluctuation in expression levels at different time points of \textit{kps} genes along with bacteria growth and the different temporal pattern of promoter PR1-1 and PR1-4 transcription coordinated the timing of bacterial growth cycle.
Chapter 6. General Discussion

The *E. coli* Group 2 polysaccharide capsule production is temperature-regulated at the level of transcription, principally by the global regulator H-NS, SlyA and IHF. The data indicated a complex pattern of regulation of transcription from PR1 with a number of interesting features. In this thesis, the multiple tandem promoters (PR1-1, PR1-2, PR1-3 and PR1-4) were located at the PR1 region, their regulation studied in detail and their relative contributions to the total transcripts of *kps* were determined.

*E. coli* is living in many places, such as the animal intestines or aquatic environments outside of the host. As such they need to adapt to extreme changes in moving from one environment such as the host, which may be relatively stable, to that of an external environment in which a greater diversity of environmental insults may be experienced. For most bacteria, survival depends on the selective expression of gene products to cope with the fluctuating environment. Therefore, it is no surprise that *E. coli* has evolved sophisticated systems to control gene expression. Most of the regulation involved multiple transcription factors or even multiple promoters interacting with each other. The identification of multiple transcription sites in PR1 and their differential expression raises interesting questions about their role; in particular the definition of what is a cryptic promoter. The permanence of cryptic promoters in the regulatory regions of bacteria could be facilitated by different kinds of evolutionary processes. For example, they could establish a group of “back up” promoter sequences, maintained by selection for robustness (Huerta *et al*., 2006). The $\sigma^{70}$ promoters are located within the zone with high densities of promoter-like signals in *E. coli* (Huerta & Collado-Vides, 2003). The existence of multiple potential promoters could minimize the harmful effects of genetic mutations at the gene expression that is critical for their surviving. In addition, the existence of multiple promoters of different strength could also allow the bacteria population rapidly adapting to the changing environment (Huerta *et al*., 2006).
The multiple promoters at PR1 region add significantly to our understanding of the regulation of Group 2 capsule gene expression. The clusters of potential promoter elements in regulatory regions lead to complex patterns of regulation. In *E. coli*, there were substantial differences in the number of families of promoters or/and the number of promoters in families even among the closely related strains (Matus-Garcia *et al.*, 2012). This evolutionary modification to transcriptional elements may be a critical aspect for the efficient and rapid rewiring of a gene’s transcriptional regulation. For example, during the promoter propagation in prokaryotes, the evolution of promoter can evolve through the recruitment or mobilization of already existing putative mobile promoters from elsewhere in the genome (Matus-Garcia *et al.*, 2012; Nijveen *et al.*, 2012). It is speculated that these cryptic promoter elements could remain inactive in the defunct regulatory systems in the evolution, or be waiting for future adoption in adaptive evolution (Islam *et al.*, 2011).

The transcription activation of promoter PR1-4 was shown to be dependent on the transcription initiation from promoter PR1-1 and speculated to mediate by a supercoiling dependent manner. Transcription mediated changes in supercoiling seems important in regulating genes that provide a way for transcription of one promoter to affect adjacent promoter via DNA topology. The transcription of *Pleu500* can be activated or inhibited depends on whether the insert elements orientation is same as or opposite to *leu* operon (El Hanafi & Bossi, 2000). It was suggested that promoters might control one promoter via local effects on DNA supercoiling, in turn, control the activity of a second promoter. Therefore, efficient recognition by RNAP at PR1-4 may require supercoiling-induced deformation of the promoter elements. The open complex formation at PR1-1 may affect a transition of conformational changes at the topological coupling promoter PR1-4 and thence activate the transcription from PR1-4.

In this study, H-NS was proven to negatively regulate the transcription from PR1 both at 37 °C and 20 °C. By performing the EMSA assay shown that H-NS had no binding affinity at the UTR region under 37 °C. Previous study demonstrated that SlyA was an anti-repressor that could counteract the negative effect from the
repressor H-NS and hence activating the transcription from PR1-1 by binding upstream of the transcription start site of PR1-1 (Corbett et al., 2007). But this positive regulation of SlyA was only targeted on the promoter PR1-1 but not others at PR1 region since no significant differences were observed for the other minimal promoter-lacZ fusions between the wild type strain and slyA::kan mutants both at 37 °C and 20 °C. It was proposed that the transcription repressor H-NS and its anti-repressor SlyA could mediate a switch between activation and repression of promoter PR1-1 at permissive temperature and non-permissive temperature, respectively. But the precise mechanism by how SlyA and H-NS participate in the activation of kpsF gene expression is still unclear. Lithgow and co-workers (2007) proposed a model for SlyA/H-NS regulation of hlyE expression in E. coli K12. When H-NS is binding dominantly at RNAP binding sites, the transcription will be inhibited and hlyE expression is silenced; when SlyA is dominant, SlyA prevents the binding of low levels of H-NS and allowing RNAP access to the promoter facilitating hlyE transcription (Lithgow et al., 2007). In E. coli, the expression of SlyA and H-NS are both temperatures dependent, which H-NS expression is increased at 20 °C relative to 37 °C while SlyA reduced at 20 °C relative to 37 °C (La Teana et al., 1991; Corbett et al., 2007). Thus it was suggested that the prevailing nucleoprotein complex that is formed at PR1 would partly depend on the relative amounts of H-NS and SlyA present in the cell and their relative affinities for the promoter at different temperatures. The interaction between repressor H-NS and transcription activator may be due to the DNA conformation changes from transcription-proficient structure to transcription inefficient structure by altering the DNA structure in opposing way; or alternatively, involved in the protein-protein competition and replacement between the activating and repressing regulator. It has been shown that the temperature shift can induce structural alteration, and these changes can reduce the interaction of H-NS, which disrupted the formation of a protein-DNA complex and allows the expression of virulence genes at higher temperatures (Madrid et al., 2002; Prosseda et al., 2004). It was proposed that the transcription switches on and off relied on a critical thresholds a ‘see-saw’ mechanism operates in which SlyA antagonizes H-NS interaction and H-NS
antagonizes SlyA among these different nucleoprotein complexes (H-NS:promoter; H-NS:SlyA:promoter and SlyA:promoter) (Lithgow et al., 2007). However, we could not exclude the possibility that there is some other transcriptional activator also involved in the regulation of PR1-1. One model had been proposed that the binding of SlyA might remodel the local nucleoprotein structure of the bacterial chromosome by counteracting bound H-NS and enabling activation of genes by other activators. The binding of SlyA likely antagonized H-NS and facilitated the interaction of protein PhoP to its own promoter and subsequently activating the phoPQ operon (Navarre et al., 2005; Ellison & Miller, 2006b).

About 1000 specific IHF binding sites have been identified in the E. coli chromosome and most of these sites are located in close vicinity upstream of promoters that are probably involved in transcription activation (Goosen & van de Putte, 1995; Ussery et al., 2001). However, in this study, IHF was shown to bind at the UTR region and repress the transcription from both promoters PR1-1 and PR1-4. It was proposed that IHF binding at the -35 hexamer region of PR1-4 could cause a U-turn as the DNA wrapped around the protein (Rice et al., 1996), which may prevent the RNAP recognizing the promoter element of PR1-4 efficiently, or render the DNA to conform in a way that precluded the stable binding of RNAP or later steps of the initiation cycle. The binding of IHF at downstream of PR1-1 is proposed to bring regulatory elements present the downstream region into close position with the promoter PR1-1, which may be involved in the H-NS-DNA-H-NS bridge formation at PR1-1. In addition, the DNA-binding proteins are important components of topological barriers (Fulcrand et al., 2013), and it is possible that the nucleoprotein complexes generated from IHF may serve as general topological barriers to modulate localized DNA supercoiling and hence affect the transcription from PR1-1 and PR1-4.

The activated transcription at 37 °C of kps genes in E. coli K5 capsule gene cluster was an important signal that trigger enhanced capsule expression of E. coli and aid resistance to the immune system in the human body. An understanding of the temperature regulation of transcription from PR1 by monitoring how the multiple
promoters at PR1 being transcribed during the temperature shift-up experiment can provide a reference to explore the environmental regulation of capsule gene expression in a more biological relevant context, for example under conditions similar to those encountered by the bacterium infected the human from the external environment. This study has provided a real-time expression of kps genes during the 37 °C temperature shift from the 20 °C, showing that promoters PR1-1 and PR1-4 at PR1 were turned on sequentially and the expression of them were fluctuated during the whole growth curve. This real-time detection of the transcription activity from the multiple promoters at PR1 during growth curve by qRT-PCR, illuminating the different temporal pattern of promoter transcription coordinated the timing of bacterial growth cycle, which might be involved in the variation of DNA superhelicity during the growth cycle (Free & Dorman, 1994; El Hanafi & Bossi, 2000; Chen & Wu, 2003; Sobetzko et al., 2012). However, the temperature dependent sensor responsible for switching between the activation to repression is still unclear. H-NS itself can function as a temperature sensor (Amit et al., 2003). H-NS can bind at AT-rich DNA template and polymerize along DNA to form a complex of higher bending rigidity at lower temperature but not above 32 °C, which indicated that H-NS can act directly as a temperature sensor that control of gene expression (Amit et al., 2003). What is more, thermoregulated capsule genes have been shown regulated by the two-component systems in bacteria (Clavel et al., 1996; Hagiwara et al., 2003; Kang et al., 2012). Thus it was possible that the two-component system may also involve in mediating the transcription from PR1.

To study the mechanism of transcription in more detail, more future works seem demanding to elusive the molecular mechanism for the regulation of E. coli K5 capsule gene expression: (1) Further investigating on the activation of promoter PR1-4 is supercoiling sensitive and its activation dependent on the transcription initiation from promoter PR1-1. The DNA supercoiling level is primarily set by opposing actions of DNA topoisomerase I and gyrase (Zechiedrich et al., 2000; Champoux, 2001). Therefore, expressing inhibitors of DNA gyrase or by introducing topA or gyr mutations can be feasible to test whether the promoter PR1-4 is
supercoiling dependent. A lot of studies had been performed by inhibiting the topoisomerase expression or function to examine how the supercoiling affects the transcription in the cells, such as measuring the transcriptional activity being supplemented with novobiocin, a DNA gyrase inhibitor (Rhee et al., 1999; Bordes et al., 2003; Gudlavalleti et al., 2004) or in the topA mutant strains (Qi et al., 1997; Fang & Wu, 1998b; El Hanafi & Bossi, 2000). What is more, in order to study how transcription induced supercoiling from PR1-1 affects downstream DNA topology and hence activates transcription from PR1-4, an experiment needs to be performed to measure the local topological changes resulted from transcriptional modulation of a transcription unit. Based on the method demonstrated by Moulin et al. (2005), a topological reporter based on the promoter of gyrA, P_{gyrA} (Menzel & Gellert, 1987) could be used to analyse how transcription affects downstream DNA topology. A supercoiling inducer promoter P_{BAD} (Guzman et al., 1995) fused to the E. coli uidA gene encoding β-glucoronidase and followed with a Rho-independent terminator served as inducible unit. A supercoiling probe which containing the P_{gyrA} fused to the E. coli lacZ gene located downstream of the inducible unit in the same orientation. Thus the increased β-galactosidase activity was strictly dependent on the presence of an inducible transcription unit and did not result from transcriptional read-through across the intervening terminator. Therefore, the potential supercoiling sensitive promoter PR1-4 transcription activity can be examined in the same way by substituting the promoter P_{gyrA} in the inducible unit. However, the plasmid-based study has an intrinsic limitation is that, because of the circular shape of the plasmid molecule. If negative and positive domains of supercoiling were generated by transcription elsewhere on the plasmid, these could diffuse around the circle and cancel each other by rotation about the duplex axis. Thus studying the promoter activation at PR1 in the chromosome seems necessary. (2) Further investigating the mechanism of H-NS repressing effect on promoter PR1-1. It is known that a short segment of DNA can resist torsional changes (Cloutier & Widom, 2004; Du et al., 2005) and that the proper angular orientation of two binding sites is necessary for the looping of DNA fragment by binding of regulatory protein (Choy et al., 1995; Bhat et al., 2014). Therefore, by
insertion a short DNA segment (few base pairs) to affect the orientation of H-NS binding site may be necessary to explore the possibility that H-NS proteins bound to the sites upstream and downstream of promoter PR1-1 transcription start site associate to induce DNA looping. To further confirm H-NS can prevent RNAP from proceeding to transcription elongation by forming a bridge and trapped in the open complex, the abortive transcription assays (Goldman et al., 2009) may also need to be employed. (3) Further explore the interaction between the regulators at PR1. It was known that H-NS preferentially binding at DNA at lower temperature (Lang et al., 2007) and thus it is possible that H-NS may binding at UTR region at lower temperature. Therefore, performing the EMSA assay by incubating the protein H-NS and UTR region at 20 °C seems necessary. Additionally, the effects that IHF have on the effects induced by H-NS could be supplemented with more extensive footprinting analysis of simultaneous H-NS and IHF binding the UTR region of PR1. (4) To understand transcription from PR1 promoter by measuring Gfp expression when the E. coli K5 PR1::gfp strain is grown under a number of different conditions. To test whether these multiple promoters may have distinct physiological roles, we could investigate how the other environmental conditions will affect transcriptions in the PR1, including oxygen concentration, osmolality, PH or during biofilm formation. Also, it might be interesting to investigate the activation of multiple promoters during growth in urine medium or during infection with cultured epithelial cells. These investigations may provide deeper insights into the biological functions of multiple promoters at PR1. (5) Investigating other regulators may involve in the regulation of capsule expression. We could mutate the E. coli K5 PR1::gfp strain with miniTn5 and look to identify regulator mutants that are defective in capsule (Gfp) expression under different growth conditions. Overall, the future works will help us to understand the mechanism of regulation of PR1 mutiple promoters of the K5 capsule genes. Further investigating the molecular level of the environmental regulation of capsule gene expression in pathogenic E. coli, in particular how multiple promoters are regulated during growth in the host?
E. coli is a model organism for studying bacterial genetics and microbial pathogenesis. The expanding understands of the capsule expression in the model organism E. coli can also provide much more information for other virulence gene expression in other bacteria. Inhibition of capsule biosynthesis may confer a valuable strategy for novel anti-bacteria treatment (Schneider & Sahl, 2010), which can avoid the biological hazard of human health from antimicrobial-resistant bacteria by using traditional antibiotics (Hammerum & Heuer, 2009). Small molecule inhibitors of K-antigen synthesis have been developed and tested in E. coli expressing K1 and K5 capsule (Goller & Seed, 2010; Noah et al., 2013). Recently, a large scale screening for the compound that inhibit the Uropathogenic Escherichia coli (UPEC) capsule biosynthesis had been performed to extend the discovery process for new capsule small molecule inhibitors (Goller et al., 2014). The E. coli K5 capsular polysaccharide have been reported as a precursor for the generation of heparin (Lindahl et al., 2005; Bhaskar et al., 2012). The polysaccharide of E. coli Group 2 capsules could prevent the biofilm formation by a wide range of Gram-negative and Gram-positive bacteria, which being significant important for designing a new strategy to limit biofilm formation on medical devices (Valle et al., 2006). Clearly, a better understanding of the regulation of the K5 capsule gene cluster promoters provide the potential to allow us to manipulate capsule gene expression and maximise production of the polysaccharide for the commercial or medical use.

In summary, regulation of the E. coli K5 genes involves a highly complex regulatory system and a schematic of the current understanding of transcription regulation of the K5 region 1 capsule gene is displayed in Figure 6.1. Overall, in this study, we identified in addition to previously characterized PR1-1 promoter mapped at +1, at least three tandem promoters PR1-2, PR1-3 and PR1-4 transcribing the kpsF direction mapped at +133, +142 and +182, respectively. H-NS needs the downstream binding site of PR1-1 to act as a transcription repressor that inhibited the transcription from PR1 both at 37 °C and 20 °C. The H-NS anti-repressor SlyA was a transcription activator that activated the transcription from promoter PR1-1. IHF played a dual role that was required for maximizing the transcription from the
PR1 but repressed the transcription from PR1-1 and PR1-4 directly. Additionally, IHF may also mediate the temperature regulation at PR1-4. We demonstrated that the activation of promoter PR1-4 rely on the transcriptional activation of promoter PR1-1. Lastly, it was first time that revealed a fluctuation in expression levels at different time points of kps genes along with bacteria growth and the different temporal pattern of promoter PR1-1 and PR1-4 transcription coordinated the timing of bacterial growth cycle. Altogether, this study gave us a better understanding on the transcriptional regulation of E. coli Group 2 capsule expression.

Figure 6. 1 Regulation of the Region 1 Promoter of Escherichia coli K5 at 37 °C.

The predicted regulation of PR1 is shown in schematic form. There are four tandem promoters PR1-1, PR1-2, PR1-3 and PR1-4 that transcribed mRNA in the same direction toward kpsF gene. Promoter PR1-1 and PR1-4 are the major promoters contributed majority of the total transcripts of kps genes. H-NS protected three regions spanning through PR1 promoter region from position -224 to -134, -121 to -79 and +1 to +32. SlyA and H-NS bind to multiple overlapping sites upstream of PR1-1 and the binding is not mutually exclusive (Corbett et al. 2007). At 37 °C, the model of regulation posits that SlyA prevents H-NS forming a nucleoprotein complex capable of repressing transcription. Question mark (?) indicates unknown regulator at PR1 that down regulated by BipA and IHF. The activation of promoter PR1-4 is positively correlated with the transcription initiation from PR1-1. IHF is required for the maximum transcription at 37 °C but the binding of IHF at +130 represses the transcription from PR1-1 and PR1-4.
References


Barne, K.A. et al., 1997. Region 2.5 of the *Escherichia coli* RNA polymerase σ 70 subunit is responsible for the recognition of the “extended – 10” motif at promoters. , **16**(13), pp.4034–4040.


---

211


Lang, B. et al., 2007. High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. *Nucleic acids research*, **35**(18), pp.6330–7.


Merrick, M.J., 1993b. MicroReview In a class of its own — the RNA polymerase sigma factor, 10, pp.903–909.


Svetlov, V. et al., 2007. Allosteric control of the RNA polymerase by the elongation factor RfaH. *Nucleic acids research, 35*(17), pp.5694–705.


Yuzenkova, Y. et al., 2011. A new basal promoter element recognized by RNA polymerase core enzyme. The EMBO journal, 30(18), pp.3766–75.


Appendix

The sequencing results below presented part of the sequenced data that between 10 bp upstream and 10 bp downstream of 5’ RACE products inserted into pGEMT-easy vector. Blue highlighted sequences indicate the pGEMT-easy vector flanking region. Grey highlighted sequences indicate the 5’ RACE anchor primer. Yellow highlighted sequences indicate the 5’ RACE specific reverse primer. Green highlighted sequences indicate the reverse transcribed mRNA in the UTR region of PR1 in 5’ RACE assay.

1. DNA sequencing data of 5’ RACE in strain P90C (pJPR1)

P90C (pJPR1) sequenced sample 1 (5’ mRNA start from ‘T’ at +139)

ACTAGTGATTGACCACCGGTATCGATGCGACTTTTTTTTTTTTTTTG
TACACCAACTTGAATTAGCACAAATAAAG
ATAGTGGGTTCGGGTTTGTGIGIGIGIGIGIGIGIGICIGICG
GATTATTTCGTCAAAGGGGATCTCGATTATCGAATTC

P90C (pJPR1) sequenced sample 2 (5’ mRNA start from ‘A’ at +126)

ACTAGTGATTGACCACCGGTATCGATGCGACTTTTTTTTTTTTTTTTT
ATATAAAATTTTTAGACACCAAAATAATTAAAC
GATCACCGGTTTGTGIGIGIGIGIGIGIGIGICIGICG
GATTATTTCGTCAAAGGGGATCTCGATTATCGAATTC

P90C (pJPR1) sequenced sample 3 (5’ mRNA start from ‘T’ at +139)

ACTAGTGATTGACCACCGGTATCGATGCGACTTTTTTTTTTTTTTTTG
TTACAACCCATTGATTTAGCACAAATAAATTA
ATTATAGTGGGTTCGGGTTTGTGIGIGIGIGIGIGIGIGICIGIC
GATTATTTCGTCAAAGGGGATCTCGATTATCGAATTC

P90C (pJPR1) sequenced sample 4 (5’ mRNA start from ‘G’ at +10)

ACTAGTGATTGACCACCGGTATCGATGCGACTTTTTTTTTTTTTGT
GCACCTCTGAGACTGACCTCAATATCGACACCAAAATAATTAAAC
GATCACCGGTTTGTGIGIGIGIGIGIGIGICIGICG
GATTATTTCGTCAAAGGGGATCTCGATTATCGAATTC

P90C (pJPR1) sequenced sample 5 (5’ mRNA start from ‘C’ at +142)

ACTAGTGATTGACCACCGGTATCGATGCGACTTTTTTTTTTTTTTTCA
ATACACCAAAATTTTTTAGCACAAATAATTAAAC
GATCACCGGTTTGTGIGIGIGIGIGIGIGICIGICG
GATTATTTCGTCAAAGGGGATCTCGATTATCGAATTC

P90C (pJPR1) sequenced sample 6 (5’ mRNA start from ‘T’ at +123)

ACTAGTGATTGACCACCGGTATCGATGCGACTTTTTTTTTTTTTTTAT
TTTTGTTTCCCTTTTGCGCTGTAATATCC
AGCGACTAAATTTTTTGCTGACGACTGACACGACATTTCGATTATCGAATTC

P90C (pJPR1) sequenced sample 7 (5’ mRNA start from ‘A’ at +1)

ACTAGTGATTGACCACCGGTATCGATGCGACTTTTTTTTTTTTTTTAT
TTTTGTTTCCCTTTTGCGCTGTAATATCC
AGCGACTAAATTTTTTGCTGACGACTGACACGACATTTCGATTATCGAATTC

P90C (pJPR1) sequenced sample 8 (5’ mRNA start from ‘A’ at +1)

ACTAGTGATTGACCACCGGTATCGATGCGACTTTTTTTTTTTTTTTAT
TTTTGTTTCCCTTTTGCGCTGTAATATCC
AGCGACTAAATTTTTTGCTGACGACTGACACGACATTTCGATTATCGAATTC

P90C (pJPR1) sequenced sample 9 (5’ mRNA start from ‘A’ at +1)
P90C (pJJ1) sequenced sample 10 (5’ mRNA start from ‘A’ at +133)

P90C (pJJ1) sequenced sample 11 (5’ mRNA start from ‘A’ at +133)

P90C (pJJ1) sequenced sample 12 (5’ mRNA start from ‘T’ at +134)

P90C (pJJ1) sequenced sample 13 (5’ mRNA start from ‘T’ at +140)

P90C (pJJ1) sequenced sample 14 (5’ mRNA start from ‘C’ at +142)

P90C (pJJ1) sequenced sample 15 (5’ mRNA start from ‘G’ at +182)

P90C (pJJ1) sequenced sample 16 (5’ mRNA start from ‘T’ at +6)

P90C (pJJ1) sequenced sample 17 have no similarity in UTR region

2. DNA sequencing data of 5’ RACE in strain P90C (pJJ1)

P90C (pJJ1) sequenced 5’ RACE sample 1 (5’ mRNA start from ‘C’ at +142)

P90C (pJJ1) sequenced 5’ RACE sample 2 (5’ mRNA start from ‘G’ at +182)

P90C (pJJ1) sequenced 5’ RACE sample 3 (5’ mRNA start from ‘T’ at +171)
P90C (pJJ1) sequenced 5' RACE sample 4 (5' mRNA start from ‘A’ at +159)

P90C (pJJ1) sequenced 5' RACE sample 5 (5' mRNA start from ‘C’ at +142)

P90C (pJJ1) sequenced 5' RACE sample 6 (5' mRNA start from ‘C’ at +142)

P90C (pJJ1) sequenced 5' RACE sample 7 & 8 have no similarity in UTR region

3. DNA sequencing data of 5' RACE in strain MS101

MS101 sequenced 5' RACE sample 1 (5' mRNA start from ‘C’ at +142)

MS101 sequenced 5' RACE sample 2 (5' mRNA start from ‘T’ at +139)

MS101 sequenced 5' RACE sample 3 (5' mRNA start from ‘G’ at +138)

MS101 sequenced 5' RACE sample 4 (5' mRNA start from ‘A’ at +1)

MS101 sequenced 5' RACE sample 5 (5' mRNA start from ‘T’ at +134)

MS101 sequenced 5' RACE sample 6 (5' mRNA start from ‘T’ at +138)

MS101 sequenced 5' RACE sample 7 (5' mRNA start from ‘T’ at +134)

MS101 sequenced 5' RACE sample 8 (5' mRNA start from ‘T’ at +138)
4. DNA sequencing data of 5' RACE in strain P90C (pHA2)

P90C (pHA2) sequenced 5' RACE sample 1 (5’ mRNA start from ‘A’ at +1)

ACTAGTGATT GACCACCGTATCGATGTCGACTTTTTTTTTTTTTC
AGGCTTTCGCCACCCCGACATGCGACTTAATAAAAGGGTATAGTGCACATGACCAAGAGCAAGATGACCCACAT
ATTATTATATTAAAAATTGTACACCCCATGATTAGCATAAAATAAATTATAGGGGTCTGGGTGTGGTGTGAC
GTG GCCATTATTTCCGTGCAAAGGGGATCCTGATTAAATCGAATTCC

P90C (pHA2) sequenced 5' RACE sample 2 (5’ mRNA start from ‘A’ at +167)

ACTAGTGATT GACCACCGTATCGATGTCGACTTTTTTTTTTTTTC
AAATTTGGTTCCCTTTCTCGCCTGTAATATCGC
AGCGTCTTCCCACCTCCATGAAACATTGCGACTTAATAAAAAGGTGATAAGTCCTGCATATAAGCATGGACTGACCATGG
TTTTATATTCATTAAAAATTTTGTTACAACCCATTGATTTAGCATAAATAAATTATAGTGGGTTCGGGTTTGTTGTGAC
GTG GCCATTATTTCCGTGCAAAGGGGATCCTGATTAAATCGAATTCC

P90C (pHA2) sequenced 5' RACE sample 3 (5’ mRNA start from ‘T’ at +121)

ACTAGTGATT GACCACCGTATCGATGTCGACTTTTTTTTTTTTTC
ATATTTTTTTTATATAGTGGGTTCGGGTTTGTTGTGACTGTG
GCATTATTTCCGTGCAAAGGGGATCCTGATTAAATCGAATTCC

P90C (pHA2) sequenced 5' RACE sample 4 (5’ mRNA start from ‘A’ at +76)

ACTAGTGATT GACCACCGTATCGATGTCGACTTTTTTTTTTTTTC
AGGCTTTCGCCACCCCGACATGCGACTTAATAAAAGGGTATAGTGCACATGACCAAGAGCAAGATGACCCACAT
ATTATTATATTAAAAATTGTACACCCCATGATTAGCATAAAATAAATTATAGGGGTCTGGGTGTGGTGTGAC
GTG GCCATTATTTCCGTGCAAAGGGGATCCTGATTAAATCGAATTCC

P90C (pHA2) sequenced 5' RACE sample 5 (5’ mRNA start from ‘A’ at +1)

ACTAGTGATT GACCACCGTATCGATGTCGACTTTTTTTTTTTTTC
AGGCTTTCGCCACCCCGACATGCGACTTAATAAAAGGGTATAGTGCACATGACCAAGAGCAAGATGACCCACAT
ATTATTATATTAAAAATTGTACACCCCATGATTAGCATAAAATAAATTATAGGGGTCTGGGTGTGGTGTGAC
GTG GCCATTATTTCCGTGCAAAGGGGATCCTGATTAAATCGAATTCC

P90C (pHA2) sequenced 5' RACE sample 6 (5’ mRNA start from ‘G’ at +183)

ACTAGTGATT GACCACCGTATCGATGTCGACTTTTTTTTTTTTTC
AGGCTTTCGCCACCCCGACATGCGACTTAATAAAAGGGTATAGTGCACATGACCAAGAGCAAGATGACCCACAT
ATTATTATATTAAAAATTGTACACCCCATGATTAGCATAAAATAAATTATAGGGGTCTGGGTGTGGTGTGAC
GTG GCCATTATTTCCGTGCAAAGGGGATCCTGATTAAATCGAATTCC

P90C (pHA2) sequenced 5' RACE sample 7 (5’ mRNA start from ‘G’ at +182)

ACTAGTGATT GACCACCGTATCGATGTCGACTTTTTTTTTTTTTC
AGGCTTTCGCCACCCCGACATGCGACTTAATAAAAGGGTATAGTGCACATGACCAAGAGCAAGATGACCCACAT
ATTATTATATTAAAAATTGTACACCCCATGATTAGCATAAAATAAATTATAGGGGTCTGGGTGTGGTGTGAC
GTG GCCATTATTTCCGTGCAAAGGGGATCCTGATTAAATCGAATTCC