



12-2008

## Interaction of Bacteriophage Mu Middle Transcription Activator Protein Mor with Promoter DNA

Kartik Iyer  
*University of Tennessee Health Science Center*

Follow this and additional works at: <https://dc.uthsc.edu/dissertations>



Part of the [Medical Immunology Commons](#)

---

### Recommended Citation

Iyer, Kartik , "Interaction of Bacteriophage Mu Middle Transcription Activator Protein Mor with Promoter DNA" (2008). *Theses and Dissertations (ETD)*. Paper 382. <http://dx.doi.org/10.21007/etd.cghs.2008.0147>.

This Thesis is brought to you for free and open access by the College of Graduate Health Sciences at UTHSC Digital Commons. It has been accepted for inclusion in Theses and Dissertations (ETD) by an authorized administrator of UTHSC Digital Commons. For more information, please contact [jwelch30@uthsc.edu](mailto:jwelch30@uthsc.edu).

---

# Interaction of Bacteriophage Mu Middle Transcription Activator Protein Mor with Promoter DNA

## Abstract

Gene expression during lytic development of bacteriophage Mu is regulated by a

transcriptional cascade in three phases: early, middle and late. Transcription from the middle promoter Pm requires the 129-amino acid transcriptional activator Mor, a product of early transcription, and the *Escherichia coli* RNA polymerase. The Pm promoter has a recognizable -10 hexamer but lacks a -35 hexamer. Mor binds as a dimer to an imperfect dyad-symmetrical element containing two 6-bp inverted repeats and centered at -43.5 in Pm. The goals of this study were:

1. To test the prediction from the crystal structure of Mor that residues Y70 and Q68 of the  $\beta$ -strand whose side chains extend away from the protein, make base specific interactions in the DNA minor groove.
2. To identify the bases between -30 and -57 of the promoter Pm that are important for its function, in terms of binding to His-Mor and transactivation with interactions with RNAP subunits, thus optimizing the Mor-binding sequence of Pm for crystal structure determination for a Mor-DNA duplex.
3. To identify the critical number of bases of Pm required for the best binding of Mor and to test the stability of Mor binding to DNA probe containing Pm.

To test the prediction of base specific interactions of the side chains of Y70 and

Q68, mutagenesis of the 4-bp spacer region in the minor groove of the Mor-binding site using degenerate oligonucleotides were done to introduce all possible substitutions. Plate phenotyping on MacConkey agar plates was used to select Pm-lacZ clones that gave a defective phenotype indicated by white color. All the white ones had other mutations elsewhere in the plasmid and all the identified substitutions gave functional phenotypes as indicated by red color of colonies. This experiment revealed that the specific bases in the minor groove are not extremely important for interaction with the side chains of Y70 and Q68 as they can tolerate mutations. But, gel shift and  $\beta$ -galactosidase expression data with a subset of these mutations indicated that the bases of the minor groove spacer region do play a modest role in His-Mor binding and activation of Pm as visible from their variations in the binding and transactivation assay.

Specific mutations were introduced in the Pm sequence from positions -30 through

-57 upstream of the transcription start site to identify the critical bases for Mor-Pm

interactions. Since Mor binds as a dimer to Pm, the mutations would indicate whether symmetry of the positions with respect to -43.5 or the specificity of the bases is what determines the importance of the bases at the respective positions. It would also help identify mutations that could increase Mor binding and positions that could contribute to interaction with RNAP subunits. Plate phenotyping, in vitro binding assay and in vivo  $\beta$ -galactosidase assays were done for all the mutations. The 6-bp imperfect dyad-symmetrical sequences flanking the minor groove spacer were found to be the most critical for Mor binding. Within the dyad-symmetry element, bases at positions -38 to -40 and -47 to -49 are the most

---

important as they do not tolerate any base changes. The region flanking the Mor-binding sites on either side does not seem to be critical for Mor binding, but the results indicate their function in transactivation, probably by influencing interactions of Mor with the RNAP subunits or conformational changes in the interactions at Mor-DNA-RNAP interfaces.

Two mutations, -46C alone and in combination with -50T, were specifically

interesting as they bound to wild-type His-Mor more effectively than wild-type promoter, but displayed reduced in vivo activity. This observation led to the prediction that Mor also functions in promoter clearance and that higher binding of Mor to the promoter somehow negatively affects release of the core RNAP for transcription. This could mean that Mor has dual functions at the middle promoter: recruitment of RNAP and release of core RNAP during transcription initiation. Oligonucleotides with these specific mutations can be used for crystal structure determination for a Mor-DNA duplex as they stabilize the complex.

Different length oligonucleotides were used in gel shift assays with wild-type His-

Mor to identify the critical number of bases needed for efficient Mor binding. This

experiment revealed that at least 20-bp, centered at -43.5, is needed for detectable binding to His-Mor.

### **Document Type**

Thesis

### **Degree Name**

Master of Science (MS)

### **Program**

Biomedical Sciences

### **Research Advisor**

Martha M Howe, Ph.D.

### **Keywords**

Bacteriophage Mu, Mu middle promoter Pm, transcription activator protein Mor, promoter-activator interactions, Mor-DNA interactions, promoter mutations

### **Subject Categories**

Medical Immunology | Medical Sciences | Medicine and Health Sciences

# **Interaction of Bacteriophage Mu Middle Transcription Activator Protein Mor with Promoter DNA**

A Thesis  
Presented for  
The Graduate Studies Council  
The University of Tennessee  
Health Science Center

In Partial Fulfillment  
Of the Requirements for the Degree  
Master of Science  
From The University of Tennessee

By  
Kartik Iyer  
December 2008

Copyright © 2008 by Kartik Iyer  
All rights reserved

## **Dedication**

To  
My Spirtual Masters  
Bhagawan Sri Sathya Sai Baba and  
Mata Amritanandamayi Devi.  
My Parents Sri S. Rajagopalan and Rama Rajagopalan.

## **Acknowledgements**

I am extremely grateful to my mentor, Dr. Martha M. Howe for accepting me into her lab and providing an ideal environment for scientific research. She has helped me gain proficiency in Molecular Biology techniques and presentation skills. I also wish to acknowledge her for accepting my work for a Master's thesis when I decided to discontinue my graduate school work through Ph.D. I also wish to thank the members of my committee, Dr. Susan Senogles and Dr. James Bina for their guidance.

I am indebted to my parents Sri Rajagopalan and Rama Rajagopalan for their unfailing support, love and sacrifice throughout my life and accepting me through all the sudden decisions in life. My special thanks to my brother, Deepak for his support. I am thankful to my aunt Meera, uncle Krishnan and cousin Jyothi who made me feel at home in Memphis. I am also indebted to India Cultural Center and Temple for making my days in Memphis memorable in many a ways by giving me an opportunity to serve as a volunteer throughout my stay in Memphis.

I am also thankful to Roopa AvaniGadda and Dr. Rai for making the environment in the lab homely to work and especially for their encouragement and guidance.

Last but not the least, I want to express my deepest sense of gratitude to Bhagawan Sri Sathya Sai Baba and Mata Amritanandamayi Devi who have been constantly guiding me through this journey of life and to whom I dedicate this life. I know I can never fully express my thankfulness to them and that the word "gratitude" does not completely justify my feelings for them. But I am humbled by their guidance and their presence throughout my life.

## Abstract

Gene expression during lytic development of bacteriophage Mu is regulated by a transcriptional cascade in three phases: early, middle and late. Transcription from the middle promoter  $P_m$  requires the 129-amino acid transcriptional activator Mor, a product of early transcription, and the *Escherichia coli* RNA polymerase. The  $P_m$  promoter has a recognizable -10 hexamer but lacks a -35 hexamer. Mor binds as a dimer to an imperfect dyad-symmetrical element containing two 6-bp inverted repeats and centered at -43.5 in  $P_m$ . The goals of this study were:

1. To test the prediction from the crystal structure of Mor that residues Y70 and Q68 of the  $\beta$ -strand whose side chains extend away from the protein, make base specific interactions in the DNA minor groove.
2. To identify the bases between -30 and -57 of the promoter  $P_m$  that are important for its function, in terms of binding to His-Mor and transactivation with interactions with RNAP subunits, thus optimizing the Mor-binding sequence of  $P_m$  for crystal structure determination for a Mor-DNA duplex.
3. To identify the critical number of bases of  $P_m$  required for the best binding of Mor and to test the stability of Mor binding to DNA probe containing  $P_m$ .

To test the prediction of base specific interactions of the side chains of Y70 and Q68, mutagenesis of the 4-bp spacer region in the minor groove of the Mor-binding site using degenerate oligonucleotides were done to introduce all possible substitutions. Plate phenotyping on MacConkey agar plates was used to select  $P_m$ -*lacZ* clones that gave a defective phenotype indicated by white color. All the white ones had other mutations elsewhere in the plasmid and all the identified substitutions gave functional phenotypes as indicated by red color of colonies. This experiment revealed that the specific bases in the minor groove are not extremely important for interaction with the side chains of Y70 and Q68 as they can tolerate mutations. But, gel shift and  $\beta$ -galactosidase expression data with a subset of these mutations indicated that the bases of the minor groove spacer region do play a modest role in His-Mor binding and activation of  $P_m$  as visible from their variations in the binding and transactivation assay.

Specific mutations were introduced in the  $P_m$  sequence from positions -30 through -57 upstream of the transcription start site to identify the critical bases for Mor- $P_m$  interactions. Since Mor binds as a dimer to  $P_m$ , the mutations would indicate whether symmetry of the positions with respect to -43.5 or the specificity of the bases is what determines the importance of the bases at the respective positions. It would also help identify mutations that could increase Mor binding and positions that could contribute to interaction with RNAP subunits. Plate phenotyping, *in vitro* binding assay and *in vivo*  $\beta$ -galactosidase assays were done for all the mutations. The 6-bp imperfect dyad-symmetrical sequences flanking the minor groove spacer were found to be the most critical for Mor binding. Within the dyad-symmetry element, bases at positions -38 to -40 and -47 to -49 are the most important as they do not tolerate any base changes. The region flanking the Mor-binding sites on either side does not seem to be critical for Mor



binding, but the results indicate their function in transactivation, probably by influencing interactions of Mor with the RNAP subunits or conformational changes in the interactions at Mor-DNA-RNAP interfaces.

Two mutations, -46C alone and in combination with -50T, were specifically interesting as they bound to wild-type His-Mor more effectively than wild-type promoter, but displayed reduced *in vivo* activity. This observation led to the prediction that Mor also functions in promoter clearance and that higher binding of Mor to the promoter somehow negatively affects release of the core RNAP for transcription. This could mean that Mor has dual functions at the middle promoter: recruitment of RNAP and release of core RNAP during transcription initiation. Oligonucleotides with these specific mutations can be used for crystal structure determination for a Mor-DNA duplex as they stabilize the complex.

Different length oligonucleotides were used in gel shift assays with wild-type His-Mor to identify the critical number of bases needed for efficient Mor binding. This experiment revealed that at least 20-bp, centered at -43.5, is needed for detectable binding to His-Mor.

# Table of Contents

|                                                                                                                       |    |
|-----------------------------------------------------------------------------------------------------------------------|----|
| <b>CHAPTER 1: INTRODUCTION</b>                                                                                        | 1  |
| Transcription machinery                                                                                               | 1  |
| $\sigma$ subunit                                                                                                      | 4  |
| $\beta$ and $\beta'$ subunits                                                                                         | 6  |
| $\alpha$ subunit                                                                                                      | 6  |
| $\omega$ subunit                                                                                                      | 6  |
| Promoter architecture                                                                                                 | 7  |
| Holoenzyme assembly                                                                                                   | 10 |
| Formation of closed promoter complex                                                                                  | 10 |
| The transcription process                                                                                             | 11 |
| Promoter melting and formation of open complex                                                                        | 11 |
| Abortive initiation                                                                                                   | 13 |
| Promoter clearance                                                                                                    | 13 |
| Transcription elongation                                                                                              | 13 |
| Transcription termination                                                                                             | 15 |
| Transcription regulation                                                                                              | 15 |
| Regulation by $\sigma$ factors                                                                                        | 16 |
| Regulation by small ligands                                                                                           | 16 |
| Regulation by small RNAs                                                                                              | 16 |
| Regulation by activators and repressors                                                                               | 16 |
| Regulation by small molecules                                                                                         | 17 |
| Bacteriophage Mu                                                                                                      | 18 |
| The middle promoter of bacteriophage Mu, $P_m$                                                                        | 22 |
| The middle operon regulator Mor                                                                                       | 23 |
| <br>                                                                                                                  |    |
| <b>CHAPTER 2: MUTATIONAL ANALYSIS OF <math>P_m</math> AND THEIR POSSIBLE EFFECTS ON BINDING AND ACTIVATION BY Mor</b> | 28 |
| Introduction                                                                                                          | 28 |
| The current work                                                                                                      | 30 |
| Materials and methods                                                                                                 | 40 |
| Media, chemicals and enzymes                                                                                          | 40 |
| Bacterial strains and plasmids                                                                                        | 40 |
| Oligonucleotide synthesis                                                                                             | 41 |
| Targeted mutagenesis                                                                                                  | 41 |
| Determination of plate phenotypes                                                                                     | 45 |
| Preparation of competent cells                                                                                        | 46 |
| Transformation protocol                                                                                               | 48 |
| Plasmid purification protocol                                                                                         | 48 |
| Sequencing of the plasmids                                                                                            | 49 |
| Small scale purification of His-Mor                                                                                   | 49 |
| <i>In vivo</i> transactivation assay                                                                                  | 50 |
| Gel mobility shift assay                                                                                              | 50 |
| Results                                                                                                               | 57 |

|                                                                                    |     |
|------------------------------------------------------------------------------------|-----|
| Spacer region mutagenesis and phenotypic frequency                                 | 57  |
| Extensive mutagenesis of P <sub>m</sub>                                            | 58  |
| Observation of phenotypes                                                          | 61  |
| <i>In vivo</i> tranactivation assay                                                | 61  |
| Gel retardation assay results                                                      | 61  |
| The length of the double-stranded oligonucleotides optimum for Mor binding         | 88  |
| Test whether Mor binding stabilizes the duplex oligonucleotide                     | 90  |
| <b>CHAPTER 3: SUMMARIZED RESULTS AND DISCUSSION</b>                                | 92  |
| Results of the spacer region mutations                                             | 92  |
| Results of the mutations in the Mor-binding segments of the promoter               | 92  |
| Results of the mutations in the region beyond Mor-binding segments of the promoter | 97  |
| Major findings from the current work                                               | 101 |
| Future directions                                                                  | 101 |
| <b>LIST OF REFERENCES</b>                                                          | 103 |
| <b>APPENDIX: DETERMINING THE EFFECT OF N- AND C-TERMINAL DELETIONS OF Mor</b>      | 117 |
| <b>VITA</b>                                                                        | 127 |

## List of Tables

|         |                                                                                         |     |
|---------|-----------------------------------------------------------------------------------------|-----|
| Table 1 | Specific mutations of the promoter                                                      | 34  |
| Table 2 | Mutations of P <sub>m</sub> and the rationale behind each of them                       | 37  |
| Table 3 | Oligonucleotides used for promoter mutagenesis                                          | 42  |
| Table 4 | List of oligonucleotides for gel shifts                                                 | 51  |
| Table 5 | Mutations of the spacer region which gave red colonies                                  | 59  |
| Table 6 | Phenotypes of strains with specific promoter mutations                                  | 62  |
| Table 7 | Summarized results of the mutants: phenotypes, $\beta$ -gal values and gel shift assays | 93  |
| Table 8 | Oligonucleotides used for N- and C-terminal deletions of His-Mor                        | 119 |
| Table 9 | Resulting phenotypes of the N- and C-terminal deletion proteins                         | 121 |

## List of Figures

|           |                                                                                                                 |    |
|-----------|-----------------------------------------------------------------------------------------------------------------|----|
| Figure 1  | The central dogma of biology                                                                                    | 2  |
| Figure 2  | Representation of the holoenzyme of bacterial RNAP                                                              | 3  |
| Figure 3  | Representation of the different regions of the $\sigma$ subunit of the RNAP with their functions                | 5  |
| Figure 4  | Eubacterial $\sigma 70$ promoter architecture and promoter elements                                             | 8  |
| Figure 5  | RNA polymerase and its interactions at promoters                                                                | 9  |
| Figure 6  | Structural transitions during the steps of transcription initiation                                             | 12 |
| Figure 7  | Structure-function model of the transcription elongation complex                                                | 14 |
| Figure 8  | Transcriptional organization of the Mu genome                                                                   | 19 |
| Figure 9  | Mu middle and late promoters                                                                                    | 21 |
| Figure 10 | Diagrammatic representation of protein-protein and protein-DNA interactions in the middle promoter open complex | 25 |
| Figure 11 | Ribbon representation of Mor dimer                                                                              | 26 |
| Figure 12 | Proposed model for DNA-binding by Mor                                                                           | 27 |
| Figure 13 | Representation of Q68 and Y70 and possible Mor-DNA contacts                                                     | 31 |
| Figure 14 | The two plasmid system used to study interactions of $P_m$ with His-Mor                                         | 47 |
| Figure 15 | Gel shift for mutations 1, 2 and 3                                                                              | 64 |
| Figure 16 | Gel shift for mutations 4, 5 and 6                                                                              | 66 |
| Figure 17 | Gel shift for mutations 7, 8 and 9                                                                              | 68 |
| Figure 18 | Gel shift for mutations 10, 11 and 12                                                                           | 70 |
| Figure 19 | Gel shift for mutations 13, 14 and 15                                                                           | 72 |
| Figure 20 | Gel shift for mutations 16 and 17                                                                               | 74 |

|           |                                                                                                |     |
|-----------|------------------------------------------------------------------------------------------------|-----|
| Figure 21 | Gel shift for mutations 18 and 19                                                              | 76  |
| Figure 22 | Gel shift for mutations 20 and 21                                                              | 78  |
| Figure 23 | Gel shift for mutations 22 and 23                                                              | 80  |
| Figure 24 | Gel shift for mutations 24 and 25                                                              | 82  |
| Figure 25 | Gel shift for mutation 26                                                                      | 84  |
| Figure 26 | Gel shift for mutations 27, 28 and 29                                                          | 86  |
| Figure 27 | Gel shifts with different size oligonucleotides                                                | 89  |
| Figure 28 | Gel shift to test Mor-oligo duplex stability                                                   | 91  |
| Figure 29 | Summarized results of the spacer region mutations                                              | 95  |
| Figure 30 | Summarized results of the mutations in the Mor-binding segments of the promoter                | 95  |
| Figure 31 | Summarized results of mutations at positions just flanking the Mor-binding segments            | 98  |
| Figure 32 | Summarized results of mutations at positions beyond the Mor-binding segments                   | 98  |
| Figure 33 | Summarized results of additional mutations                                                     | 99  |
| Figure 34 | General conclusions from mutagenesis of the middle promoter                                    | 100 |
| Figure 35 | Amino acid and secondary structure alignment for Mor and C proteins of bacteriophage Mu        | 118 |
| Figure 36 | Schematic diagram of the strategy used for deletion of amino acids from the N- terminus of Mor | 120 |
| Figure 37 | Gel shift for Mor proteins carrying C-terminal deletions                                       | 123 |
| Figure 38 | Gel shift for $\Delta 6C$ and $\Delta 9C$ Mor proteins                                         | 123 |
| Figure 39 | Gel shift for Mor proteins carrying N-terminal deletions                                       | 125 |
| Figure 40 | Gel shift for $\Delta 10N$ and $\Delta 15N$ Mor proteins                                       | 125 |
| Figure 41 | Gel shift with increasing amounts of supernatant                                               | 126 |

## List of Abbreviations

|               |                                                             |
|---------------|-------------------------------------------------------------|
| $\alpha$ -CTD | C-terminal domain of the $\alpha$ subunit of RNA polymerase |
| $\alpha$ -NTD | N-terminal domain of the $\alpha$ subunit of RNA polymerase |
| $\beta$ -ME   | $\beta$ -mercaptoethanol                                    |
| Amp           | ampicillin                                                  |
| bp            | base pair(s)                                                |
| BSA           | bovine serum albumin                                        |
| CAP           | catabolite gene activator protein                           |
| Cm            | chloramphenicol                                             |
| DNA           | deoxy ribo nucleic acid                                     |
| DNase I       | deoxyribonuclease I                                         |
| (d)NTP        | (deoxy) ribonucleoside triphosphate                         |
| DTT           | dithiothreitol                                              |
| HTH           | helix-turn-helix                                            |
| IPTG          | isopropyl- $\beta$ -D-thiogalactopyranoside                 |
| mRNA          | messenger RNA                                               |
| nt            | nucleotide(s)                                               |
| ONPG          | o-nitrophenyl- $\beta$ -galactopyranoside                   |
| PCR           | polymerase chain reaction                                   |
| RNA           | ribo-nucleic acid                                           |
| RNAP          | <i>Escherichia coli</i> RNA polymerase                      |
| SDS           | sodium dodecyl sulfate                                      |
| WT            | wild-type                                                   |

# Chapter 1: Introduction

The central dogma of molecular biology is: DNA is transcribed to RNA which is translated to protein. Transcription is the process by which messenger RNA is made from the DNA using the anti-sense strand as the template. The sense strand of DNA has the codons for amino acid sequence of the protein that it codes for. In the subsequent step, mRNA is translated into amino acid sequence of the protein. Thus, transcription is the first important process by which the genetic information stored in the DNA is expressed (Figure 1).

## Transcription machinery

DNA-dependent RNA polymerase is the key enzyme responsible for transcription in living cells. It catalyzes the formation of 5' - 3' phosphodiester bonds between ribonucleoside triphosphates.

RNA polymerase can be divided into 2 evolutionary conserved groups based on the number of subunits constituting the enzyme; single-subunit RNA polymerases and multi-subunit RNA polymerases. Single subunit RNA polymerases are the simplest form of the enzyme in mitochondria and phages SP6 and T7 (McAllister and Raskin, 1993; Tunitskaya and Kochetkov, 2002). Prokaryotic and eukaryotic RNA polymerases are multi-subunit enzymes. Prokaryotic cells are characterized by one RNA polymerase, while eukaryotic cells have 3 different RNA polymerases: Pol I for rRNA synthesis, Pol II for mRNA synthesis and Pol III for tRNA synthesis (Cramer, 2002).

Structural studies between the core subunits of eubacterial, archeal and eukaryotic RNA polymerases revealed that the homology of their amino acid sequences is extended to the structure and function of these proteins. They also show similarities in the relative position of the subunits, relative positions of the functional determinants, structural folding topologies of the subunits and in their mechanisms of interaction with DNA (Cramer, 2002). The most highly conserved regions are around the active center of the enzyme. This similarity decreases towards the outer part; this region is the target for transcriptional regulators (Ebright, 2000). The involvement of multiple subunits of Pol II and the larger size of pre-initiation complexes in eukaryotes pose a formidable challenge to dissect the key processes of eukaryotic transcription. Instead, the simple and highly similar bacterial RNA polymerase serves as an excellent model system in understanding the transcriptional machinery in general.

The prokaryotic core RNAP consists of 5 subunits:  $\alpha$  (2 copies),  $\beta$ ,  $\beta'$  and  $\omega$  (Figure 2). In association with the  $\sigma$  factor in a multistep and cooperative process, the core RNAP forms the holoenzyme (Borukhov and Nudler, 2003). The assembly pathway of the holoenzyme is represented as follows:

$\alpha \rightarrow \alpha_2 \rightarrow \alpha_2\beta \rightarrow \alpha_2\beta\beta' \rightarrow \alpha_2\beta\beta'\omega$  (CORE)  $\rightarrow \alpha_2\beta\beta'\omega\sigma$  (HOLOENZYME)  
(Ishihama, 1981; Ishihama, 1993).



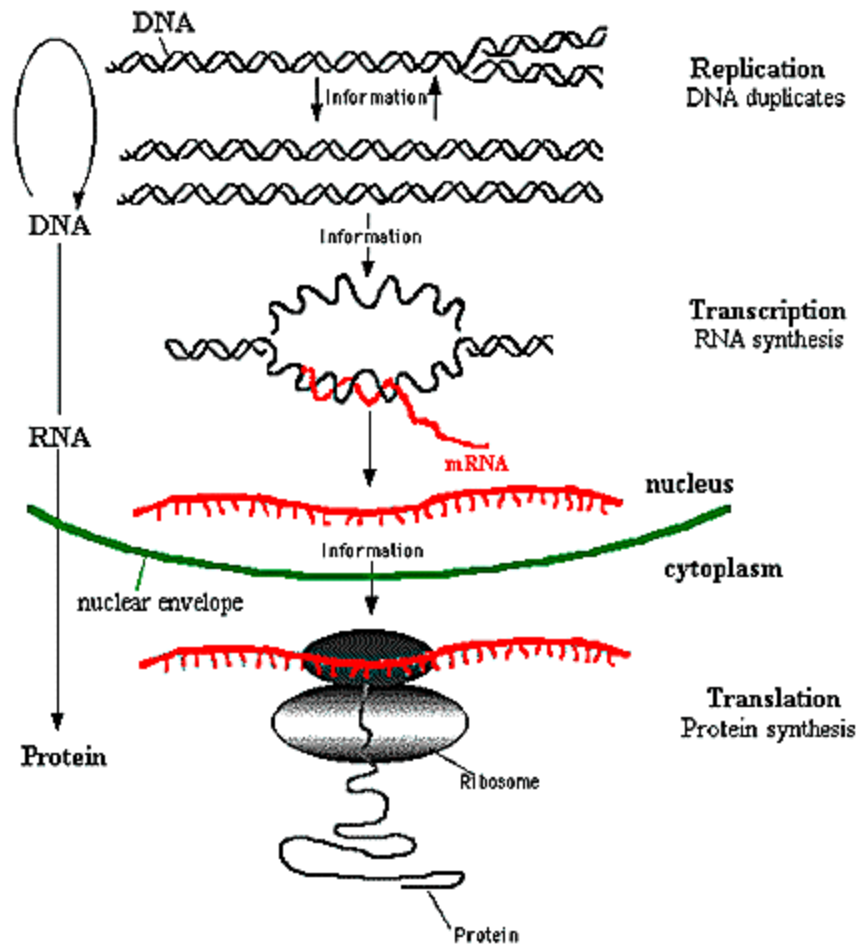


Figure 1: The central dogma of biology.

Transcription of DNA to RNA to protein.

This dogma forms the backbone of molecular biology and is represented by three stages:

1. Replication of DNA: The DNA replicates its information in a process that involves many enzymes.
2. Transcription of the information from DNA to RNA: The DNA codes for the production of messenger RNA (mRNA) during transcription. In eucaryotic cells, the mRNA is processed (essentially by splicing) and migrates from the nucleus to the cytoplasm.
3. Translation of the information from messenger RNA to protein sequence: Messenger RNA carries coded information to ribosomes. The ribosomes "read" this information and use it for protein synthesis. This process is called translation.

Reprinted with permission from Access Excellence @ the National Health Museum  
<http://www.accessexcellence.org/RC/VL/GG/central.php>, Accessed June 18, 2008.

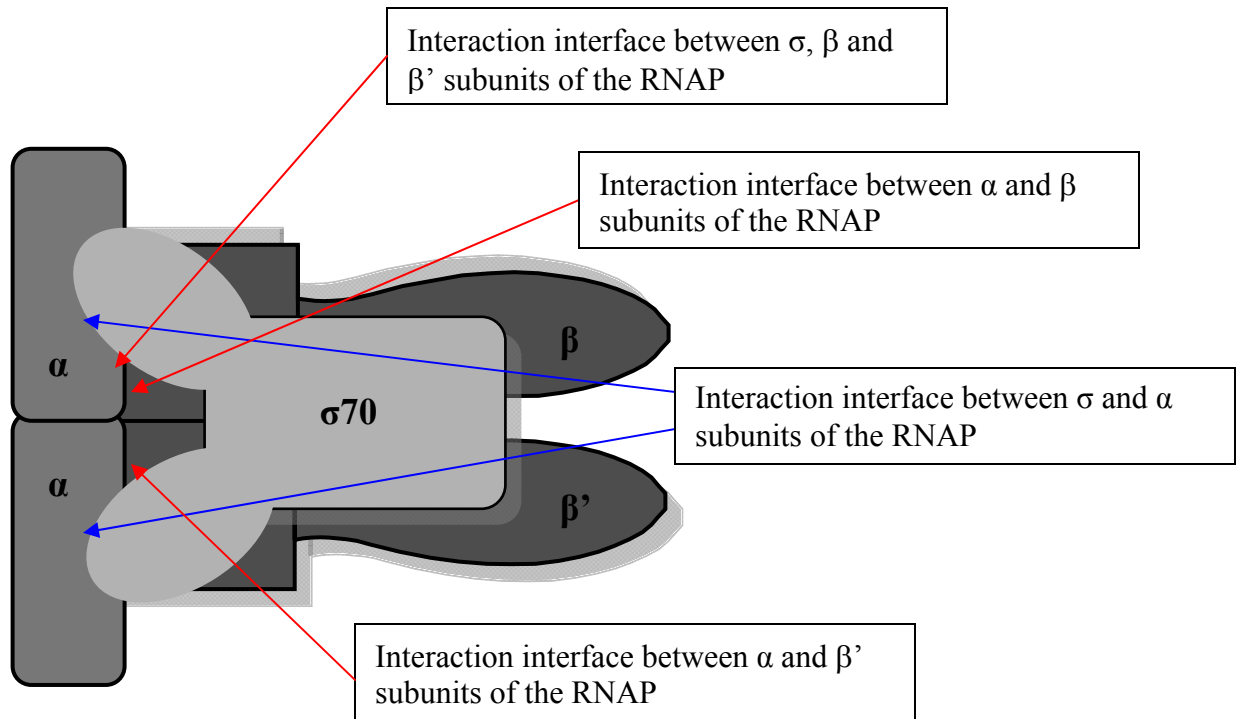


Figure 2: Representation of the holoenzyme of bacterial RNAP.

The holoenzyme, with its subunits  $\sigma 70$ ,  $\alpha$  (2 copies),  $\beta$  and  $\beta'$  are shown in the figure. The smallest subunit  $\omega$  is not shown in the figure. The interaction interfaces between the various subunits are shown by arrows.

## **$\sigma$ subunit**

The  $\sigma$  factor, conferring promoter specificity, is directly responsible for promoter recognition and initiation of transcription. It is involved in the crucial process of conversion of the core enzyme into the holoenzyme during transcription initiation. It also functions in promoter melting, stabilization of the resulting single stranded DNA, binding of initiating nucleotide triphosphates and as a target for multiple transcriptional regulators (Li et al, 1994; Gross et al, 1998; Lonetto et al, 1998; Nickels et al, 2002; Dove et al, 2003; Mitchell et al, 2007). In *E.coli*, there are different  $\sigma$  factors which recognize different promoters. Promoters of most of the genes involved in house-keeping function are recognized by  $\sigma 70$ . Promoters of the genes involved in nitrogen metabolism are recognized by  $\sigma 54$ , also called  $\sigma N$ . Recognition of the promoters of the heat shock response genes is by  $\sigma 32$ . Promoters of the genes for motility and chemotaxis are turned on by  $\sigma 28$ . Promoter recognition of the stationary phase and stress response genes is controlled by  $\sigma 38$  (Ishihama, 2000; Grouse et al, 2006). Based on amino acid sequence comparisons of members of the  $\sigma 70$  family, it has 4 highly conserved regions, regions 1.0-4.0 (Lonetto et al, 1992) which are the functional domains of  $\sigma$  subunit. These are further divided into sub-regions to which specific functions are assigned based on biochemical and genetic studies. These highly conserved regions of the  $\sigma$  subunit have specific functions that contribute to the subunit's structure and function (Figure 3).

Region 1.0: The self-inhibitory domain 1.1 of  $\sigma$ , is responsible for masking the DNA-binding regions of  $\sigma$  before it binds the core complex to form the holoenzyme (Dombroski et al, 1993). It was later shown that this region may bind the main channel of the core enzyme by widening the channel and enhancing the binding of DNA duplex (Bowers and Dombroski, 1999; Borukhov and Severinov, 2002). Region 1.2 along with 2.1 and 2.2 is believed to create the hydrophobic interface for interaction with core enzyme. Recently, along with region 2.4, region 1.2 was also shown to be involved in promoter recognition of the -10 element (Zenkin et al, 2007).

Region 2.0: Region 2.1 and 2.2 help form the holoenzyme by specifically interacting with the  $\beta'$  subunit (Sharp et al, 1999). Region 2.3-2.4 bind single stranded DNA in the formation of the open complex by promoter melting (Juang and Helmann, 1994) and interacts specifically with the -10 element of the promoter (Malhotra et al, 1996; Sevostyanova et al, 2007).

Region 3.0: The subregions, 3.0-3.2 are implicated in promoter recognition, proper positioning of DNA-binding domains of  $\sigma$ , binding of nucleotides, abortive initiation and promoter clearance. Region 3.0 is involved in the recognition of the extended -10 promoter element (TGn -10), which compensates for the absence of a -35 element in these promoters (Malhotra et al, 1996; Barne et al, 1997). Regions 3.1 and 3.2 form a loop called the  $\sigma 3.2$  loop as it is predominantly formed by region 3.2. The presence of this loop causes steric hindrance to the growing nascent transcript and is thought to be responsible for abortive initiation. The transcript needs to displace this loop in order to initiate elongation (Borukhov and Nudler, 2003; Murakami and Darst, 2003).

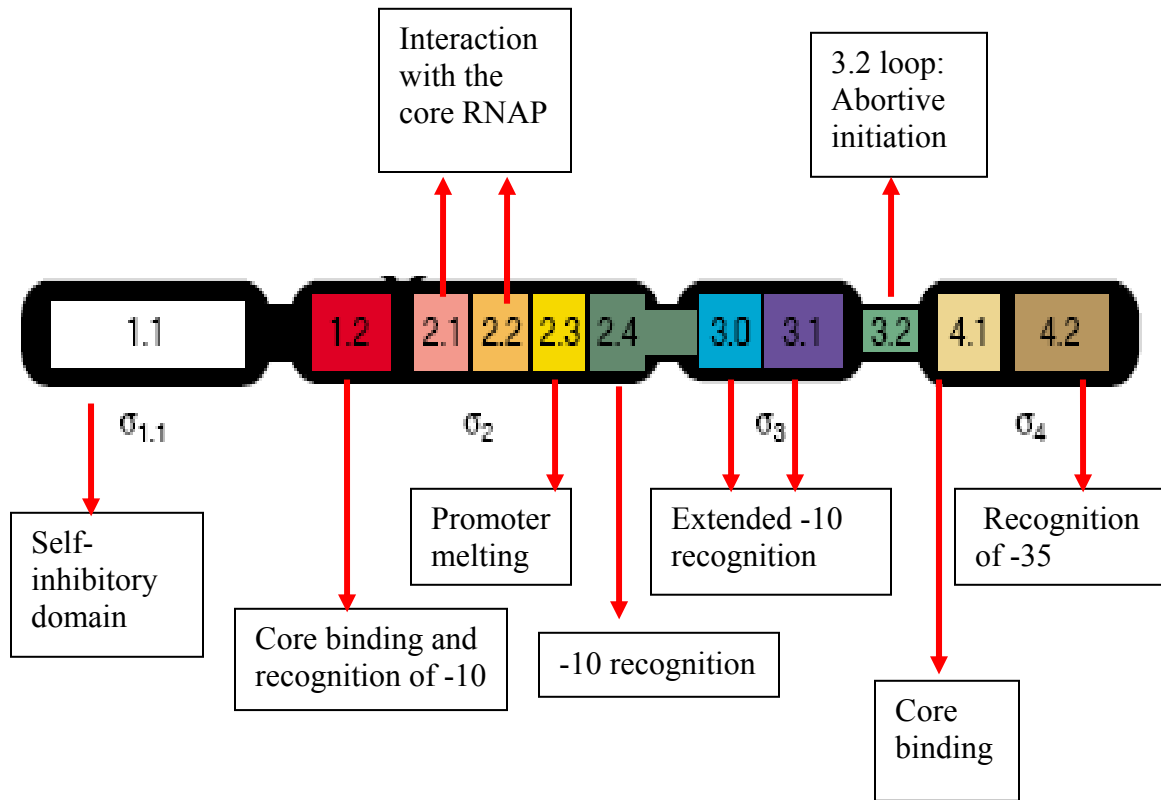


Figure 3: Representation of the different regions of the  $\sigma$  subunit of the RNAP with their functions.

Modified with permission from Elsevier Limited. Murakami KS, Darst SA. Bacterial RNA polymerases: the whole story. *Curr Opin Struc Biol.* 2003, 13(1): 31-9.

Region 4.0: Region 4.0 which is also called the  $\sigma$ -CTD plays a significant role in recognition of the -35 hexamer, interaction with the core enzyme, repositioning of DNA-binding domains of  $\sigma$  and activation of Class II activator dependent promoters (Campbell et al, 2002; Dove et al, 2003; Geszvain et al, 2004). Region 4.2 recognizes the -35 element of the promoters (Campbell et al, 2002). Conformational changes induced by  $\sigma$  4.1 -  $\beta$  flap interactions reorient  $\sigma$ 4.2 and  $\sigma$ 2.4 such that  $\sigma$ 4.2 is in a position to interact with -35 element of the promoters (Callaci et al, 1999; Kuznedelov et al, 2002).

## **$\beta$ and $\beta'$ subunits**

The  $\beta$  and  $\beta'$  subunits form the catalytic center of the RNAP, forming the roof and the floor of the DNA tunnel respectively. Various regions of these subunits are specifically involved in interaction with  $\alpha$  subunit, promoter melting, separation of the DNA:nascent RNA hybrid during elongation, closure of the main channel during elongation and formation of the secondary channel for entry of incoming nucleotides (Zhang et al, 1999; Vassylyev et al, 2002).

## **$\alpha$ subunit**

The  $\alpha$  subunit is a 329-amino acid protein which is folded into two domains that are connected by a 13-amino acid flexible linker. The two domains are the N-terminal domain ( $\alpha$ -NTD; amino acid residues 1-235) and the C-terminal domain ( $\alpha$ -CTD; amino acids 249-329). The dimerization of the 2  $\alpha$ -NTDs generates a platform for the assembly of the core complex. The N-terminal domain of the  $\alpha$  subunit helps in the assembly of the  $\beta$  and  $\beta'$  subunits. It also plays a minor role as a target for catabolite activator protein (CAP) in some CAP-dependent Class II promoters (Ebright and Busby, 1995). The C-terminal domain of the  $\alpha$  subunit can contact the upstream promoter region (UP-element) and can also interact with activator proteins (Ross et al, 1993; Busby and Ebright, 1994; 1995). The flexible linker can be 13 to 36 amino acids long and is thought to provide the flexibility for  $\alpha$ -CTD to reach the UP-elements. The natural length of the linker is the most optimal for  $\alpha$ -CTD-DNA interactions (Meng et al, 2001).

## **$\omega$ subunit**

This is the smallest subunit of RNAP and consists of 90 amino acids. This subunit was in the past thought to be an impurity which tightly binds to RNAP. Deletion of the *rpoZ* gene encoding the  $\omega$  subunit does not affect cell growth or activity of RNAP (Vassylyev et al, 2002). It is now clear that the  $\omega$  subunit predominantly acts as a chaperone in folding of the  $\beta'$  subunit and its subsequent assembly into the  $\alpha_2\beta\beta'$  complex. It also participates in physical protection of the  $\beta'$  subunit, RNAP assembly and  $\sigma$  factor recruitment. It is also thought to play roles in mediation of the stringent response, transcriptional regulation, stress adaptation of RNAP and survival in stationary phase (Mathew and Chatterji, 2006).

## Promoter architecture

Promoters are the minimal DNA sequence elements located upstream of a transcription start site and provide a control point for regulated gene transcription. The promoter acts as a specific recognition site for RNAP and transcription factor proteins. Analysis of many *E.coli* promoters has revealed that there are 3 conserved promoter elements, characteristic of a typical eubacterial  $\sigma 70$  promoter (Hawley and McClure, 1983) as represented in Figure 4:

1. The -10 sequence: Also called Pribnow Box, it is centered approximately 10 base pairs upstream of the transcription start site, with a consensus sequence 5' TATAAT 3'.
2. The -35 sequence: Centered approximately 35 base pairs upstream of the transcription start site, with a consensus sequence 5' TTGACA 3'.
3. The spacer: It is typically located  $17 \pm 1$  base pairs between the two important promoter elements, -10 and -35 (Harley and Reynolds, 1987; deHaseth et al, 1998).

Thus, the consensus promoter architecture can be represented as follows:

5' TTGACA ----  $17 \pm 1$  ---- TATAAT 3'.

The strength of a promoter can be correlated to the degree of similarity of these promoter sequence elements to the consensus; the higher the similarity, stronger the promoter.

Promoter recognition by the RNAP holoenzyme (Figure 5) is primarily achieved through interactions of  $\sigma 2.4$  with -10 hexamer and  $\sigma 4.2$  with -35 hexamer, though recent research has indicated the involvement of other regions of RNAP either directly or indirectly making these interactions more complex. It has been shown that  $\sigma 1.2$  can also play a significant role in the recognition of the -10 hexamer (Malhotra et al, 1996; Campbell et al, 2002; Murakami et al, 2002; Vassylyev et al, 2002; Zenkin et al, 2007).

An additional class of promoters, called extended -10 promoters, is characterized by the presence of a TGN trinucleotide sequence immediately upstream of the -10 hexamer and the absence of a recognizable -35 element (Keilty and Rosenberg, 1987). Recent research raises the possibility that a few more bases may play a role in these extended -10 promoters (Shultzaberger et al, 2007; Hook-Barnard et al, 2006). In the absence of a recognizable -35 element, the promoter activity is independent of interactions of  $\sigma 4.2$  with the -35 hexamer (Kumar et al, 1993), but is dependent on contacts between  $\sigma 3.0$  and the extended -10 sequence (Barne et al, 1997).

A strong correlation between the strength of rRNA P1 promoters and a 20-bp long AT-rich region located upstream of the -35 hexamer identified a new promoter element called the Upstream Promoter element (UP-element). This was defined as the promoter recognition element for the  $\alpha$ -CTD of RNAP (Ross et al, 1993; Ross, 1993; Ross et al, 1994; Ebright and Busby, 1995; Husnain and Thomas, 2008). The additional interaction of the  $\alpha$ -CTD with the UP-element results in efficient recruitment of RNA polymerase to the promoter and contributes to the strength of the UP-element containing promoters

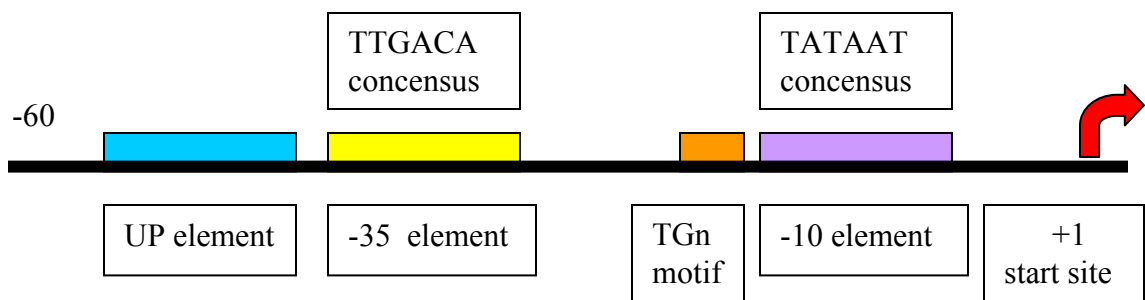


Figure 4: Eubacterial  $\sigma_{70}$  promoter architecture and promoter elements.

The DNA sequences from -60 to the +1 transcription start site are shown as a thick black line. The red bent arrow represents the +1 transcription start site and the rectangle boxes denote the specific elements of a bacterial promoter. Individual promoter elements are as written below their respective positions and the TGn motif represents the extended -10 element. The consensus sequences for the -10 and -35 are given above their respective positions.

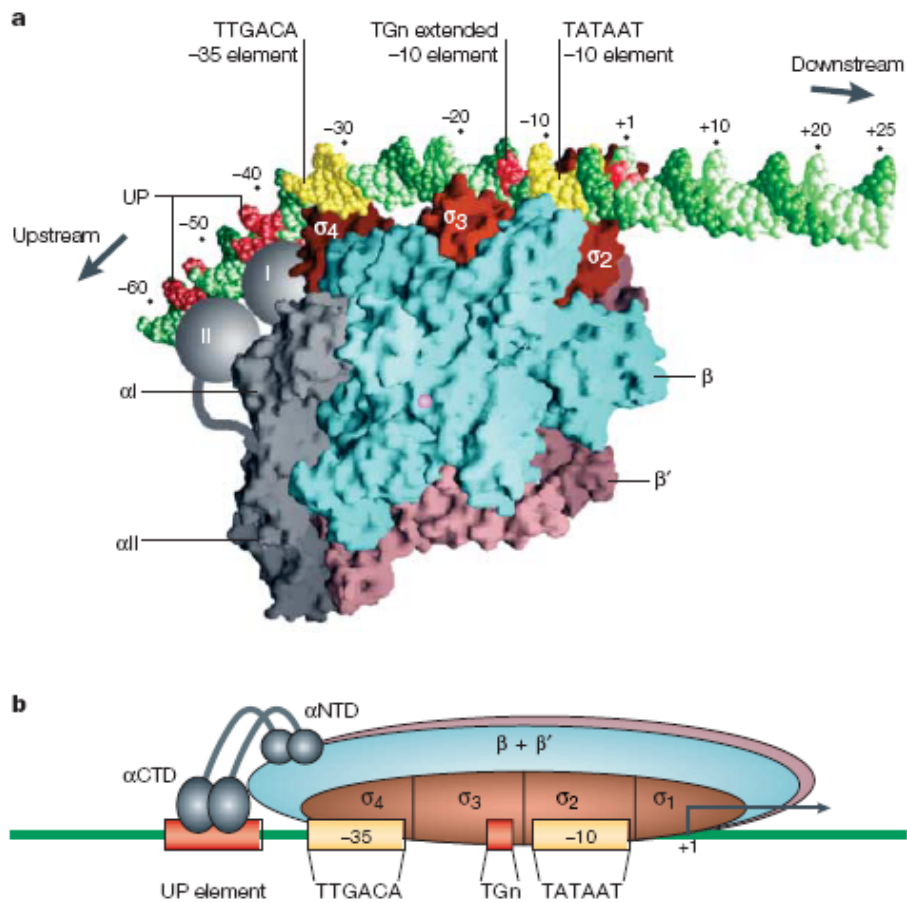


Figure 5: RNA polymerase and its interactions at promoters.

a) A model based on crystallographic studies of the initial docking of the RNA polymerase holoenzyme to a promoter. The DNA strands are shown in green, with the -10 and -35 elements highlighted in yellow and the TGn extended -10 and the UP elements highlighted in red. RNA polymerase is shown with the  $\beta$  and  $\beta'$  subunits coloured light blue and pink, respectively,  $\alpha$ -NTDs are coloured grey and the different domains of  $\sigma$  are coloured red. Grey spheres labelled I and II, represent the domains of  $\alpha$ -CTD that bind to the promoter. The RNA polymerase active site is denoted by the  $Mg^{2+}$  ion, (magenta).

b) A cartoon representation of the model of RNAP interaction at a bacterial promoter, illustrating the different interactions between promoter elements and the RNA polymerase. The consensus sequences for the -35 (TTGACA), extended -10 (TGn) and -10 (TATAAT) elements are shown.

Reprinted with permission from Nature Publishing Group. Browning DF, Busby SJ. The regulation of bacterial transcription initiation. *Nat Rev Microbiol.* 2004, 2(1): 57-65.



(Aiyar et al, 1998; Gourse et al, 2000). This element has two subsites, proximal and distal, which function as binding sites for the two  $\alpha$ -CTD's (Estrem et al, 1999).

### **Holoenzyme assembly**

Dimerization of  $\alpha$  subunit is the first step in the assembly pathway, which provides a platform for the other subunits of the RNAP ( $\beta$ ,  $\beta'$ ,  $\omega$  and  $\sigma$ ) to assemble. Thus,  $\alpha$  plays a critical role in the assembly of the core enzyme (reviewed by Ishihama, 1981; Ishihama et al, 1987).

The assembly pathway of the holoenzyme is represented as follows:  
 $\alpha \rightarrow \alpha_2 \rightarrow \alpha_2\beta \rightarrow \alpha_2\beta\beta' \rightarrow \alpha_2\beta\beta'\omega$  (CORE)  $\rightarrow \alpha_2\beta\beta'\omega\sigma$  (HOLOENZYME).

Association of  $\sigma$  subunit to the core enzyme forms the holoenzyme complex. The type of  $\sigma$  factor associated with the core enzyme directs it to activate transcription from a specific promoter. Under normal growth conditions,  $\sigma^{70}$  associates with the core enzyme to direct transcription from the housekeeping genes. The type of  $\sigma$  factor to associate with the core enzyme depends on the cellular needs at that time. Alternative  $\sigma$  factors come to play during stressful or nutrient-depriving conditions. Occlusion of the core-binding determinants of  $\sigma$  by anti-sigma factors depending on environmental needs dictates the availability of the appropriate  $\sigma$  type for interactions with core RNAP (Mathee and Hughes, 1998; Campbell et al, 2003). The interactions between most of the  $\sigma$  factors and the core enzyme are similar and comparable. Interactions between  $\sigma^{1.2}$  and  $\sigma^{2.2}$  with the coiled coil element of  $\beta'$  and of  $\sigma^{4.0}$  with the  $\beta$ -flap domain form the primary and secondary contacts respectively between  $\sigma$  and the core RNA polymerase (Borukhov and Severinov, 2002; Vassylyev et al, 2002; Murakami et al, 2002).

### **Formation of closed promoter complex**

Binding of the core enzyme to the  $\sigma$  induces considerable conformational changes in the  $\sigma$  subunit that have important functional aspects. These conformational changes place the  $\sigma^{4.0}$  and  $\sigma^{2.0}$  in optimal places in the complex conferring on them the necessary specificity in terms of interactions with -35 and -10 elements of the promoter respectively (Callaci, 1999; Kuznedelov et al, 2002; Geszvain et al, 2004). These conformational changes in the RNAP holoenzyme also have functional implications in promoters with the extended -10 or the UP-element. Specific interactions between the  $\sigma^{2.4}$  and  $\beta$  facilitates opening of the main channel allowing entry of duplex DNA into it as shown in Figure 6 (Polyakov et al, 1995; Murakami and Darst, 2003). Transcriptional activators play a major important role in the recruitment of RNAP to the promoter through specific interactions with the RNAP subunits, primarily the more accessible  $\alpha$ -CTD (Ebright and Busby, 1995).

The role of DNA bending at this step of association of RNAP with the DNA in presence of activator is stoichiometrically important. DNA bending in some cases is a

consequence of the activator-DNA interactions while in others DNA bending facilitates protein-DNA interactions. This distortion of DNA is important for the complex protein-protein and protein-DNA interactions at the promoter. The activators involved in DNA bending are called architectural activators as they influence the promoter architecture. Regulators of transcription through DNA bending sustain or inhibit the active promoter configuration to the effect that they can either activate or repress transcription depending on their effect (Perez-Martin and Espinosa, 1993; Perez-Martin and de Lorenzo, 1997).

The regions 2.4 and 4.2 of  $\sigma$  subunit recognize the -10 and -35 promoter elements respectively relative to the transcription start site (+1). This forms the closed promoter complex which is in equilibrium with free RNAP and promoter DNA. The formation of this closed promoter complex is the first step in transcription initiation. It also involves other interactions (like the one between  $\alpha$ -CTD and UP element) and active conformational changes at the promoter making it optimum for transcription.

### **The transcription process**

There are three stages in the transcription process: initiation, elongation and termination. Biochemically distinct steps characterize the basic transcription cycle of *E.coli* RNAP. Transcription initiation can be further sub-divided into three stages: open promoter complex formation, abortive initiation and promoter escape (Borukhov and Severinov 2002). Promoter complex formation itself is a multi-step process involving binding of RNAP to the promoter forming the closed complex (RPc), transition from closed complex to transcriptionally competent open complex (RPO) with additional intermediate stages involving active conformational changes in the complex. The open complex is transcriptionally and enzymatically active. After a few abortive transcripts, promoter clearance allows the transition into transcription elongation during which RNA is made by adding nucleotide tri phosphates to the growing chain till the RNAP encounters the transcription stop site at which the transcription process is terminated. Thus, the process of RNA synthesis by RNAP requires that this protein has multiple functions either by itself or in concert with other proteins; it acts as a sequence-specific DNA-binding protein, then as a DNA strand-separation protein and finally as RNA synthetase.

### **Promoter melting and formation of open complex**

The exposure of the template strand for transcription requires strand separation of the duplex DNA. The DNA duplex in the -10 region is typically melted from -11 to +3 and includes the transcription start site +1 (Young et al, 2002). Region 2.3 of  $\sigma$  subunit stabilizes the initial transcription bubble extending downstream past the transcription start site to form the open promoter complex (Malhotra et al, 1996). In this position, and associated with significant conformational changes, the single stranded template strand is enclosed in an active site tunnel formed by the main channel (Figure 6). The downstream double stranded DNA from +5 to +12 from the transcription start site is enclosed by the  $\beta$

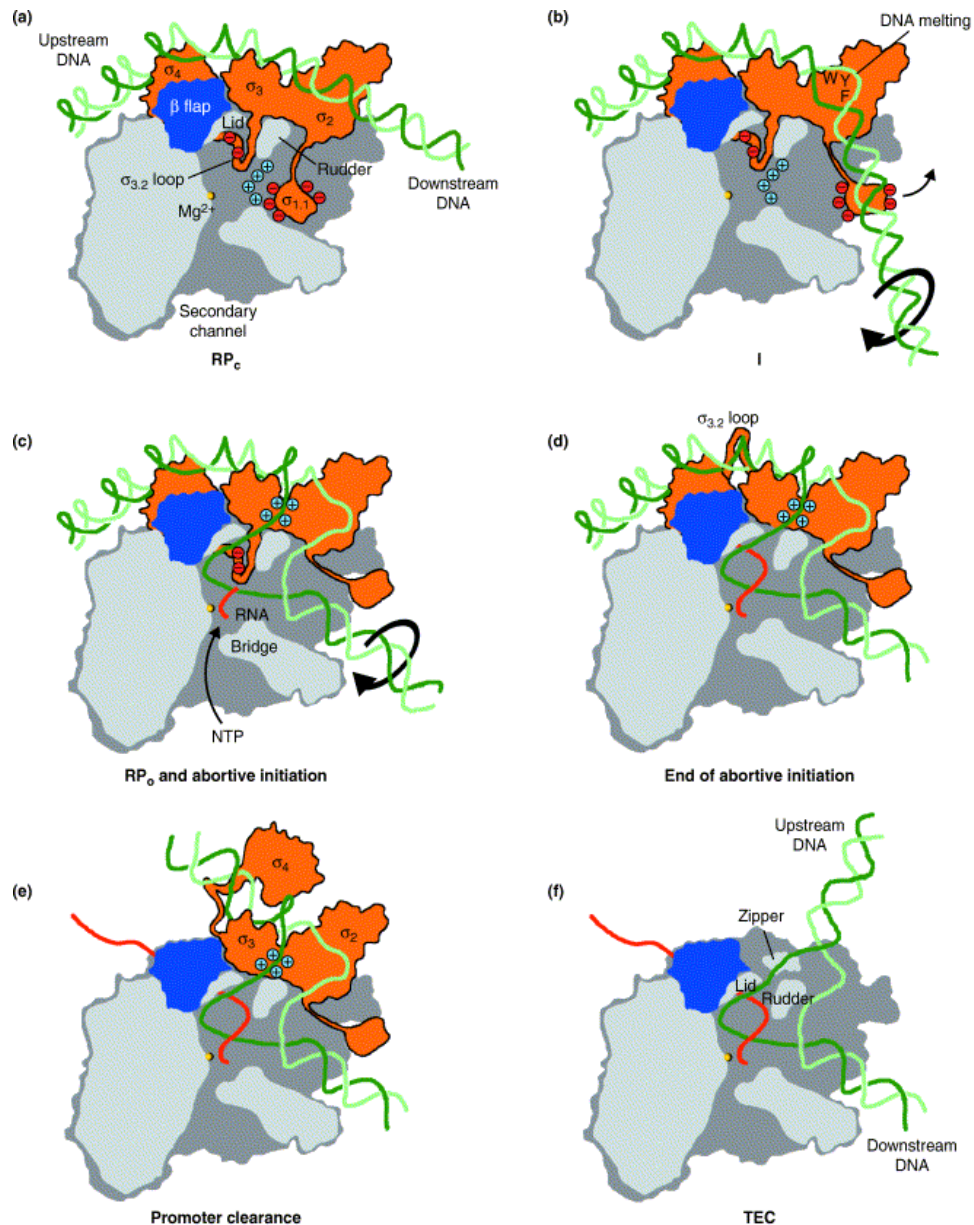


Figure 6: Structural transitions during the steps of transcription initiation.

Shown are cross-sectional views of the RNAP holoenzyme ( $\beta$ -flap, blue;  $\sigma$ , orange; rest of RNAP, gray; catalytic  $Mg^{2+}$ , yellow sphere), promoter DNA (template strand, dark green; nontemplate strand, light green;  $-10$  and  $-35$  elements, yellow) and the RNA transcript (red) at the  $RP_c$  (a), intermediate (I) (b),  $RP_o$  and abortive initiation (c), end of abortive initiation (d), promoter clearance (e) and TEC (f) stages of transcription initiation. The view is looking down on top of the  $\beta$  subunit, but with most of  $\beta$  removed, revealing the inside of the RNAP active site channel.

Reprinted with permission from Elsevier Limited. Murakami KS, Darst SA. Bacterial RNA polymerases: the whole story. *Curr Opin Struc Biol.* 2003, 13(1): 31-9.

and  $\beta'$  subunits, stabilizing the complex. In the presence of nucleotides that enter the complex through a secondary channel and in association with  $\sigma 3.0$ , transcription is initiated from the +1 transcription start site in the RNAP active site (Severinov et al, 1994; Murakami et al, 2002).

### **Abortive initiation**

The process of abortive initiation involves synthesis of a few short aborted transcripts, about 2-12 nucleotides long and their release from the complex in an unsuccessful attempt to continue transcription. The  $\sigma 3.2$  loop plays a major role in this process of abortive initiation. Two structural elements of  $\sigma 3.2$ , a short helix and a long linker were identified to be the root cause for abortive initiation. The elongating RNA chain of 2-3 nucleotides encounters the short helix and this might cause the dissociation of transcript from the complex. Similarly, the  $\sigma 3.2$  loop is encountered by longer transcripts of 2-8 nucleotides on their way out of the complex (Vassylyev et al, 2002). At this juncture, if the RNA chain fails to displace the  $\sigma 3.2$  loop out of its way, this nascent transcript is cut by the RNAP and it dissociates from the complex. When the RNA chain elongates to about 12 nucleotides long, it is sufficient to fill the RNA-DNA hybrid and the upstream RNA channel completely under the  $\beta$ -flap, and it is now capable of displacing the  $\sigma 3.2$  loop. This ends the process of abortive transcription (Murakami and Darst, 2003).

### **Promoter clearance**

This biochemical step drives the complex to initial stages of transcription elongation. During promoter escape, interactions between  $\sigma 4.2$  and the -35 element are destabilized. The RNAP can now leave the promoter and translocate downstream as RNA chain is elongated. The initial stage of transition into elongation does not require the complete dissociation of  $\sigma$  subunit, though eventually the  $\sigma$  factor is completely released from the complex, forming the ternary elongation complex with the core RNAP (Murakami and Darst, 2003). The formation of this elongation complex requires significant conformational changes in the complex. This ternary elongation complex has high processing ability to elongate the transcript.

### **Transcription elongation**

During promoter clearance, the core enzyme leaves the promoter and enters the highly productive elongation phase of transcription. As the productive initial transcribing complex enters the elongation phase,  $\sigma 70$  is released from the promoter while the core enzyme continues transcription (Carpousis and Gralla, 1985; Krummel and Chamberlin, 1989). This is diagrammatically represented in Figure 7. Translocation is proposed to involve scrunching of downstream DNA into the enzyme interior, generating strain that ultimately leads to the displacement of  $\sigma$  contacts on promoter DNA (Vo et al, 2003).

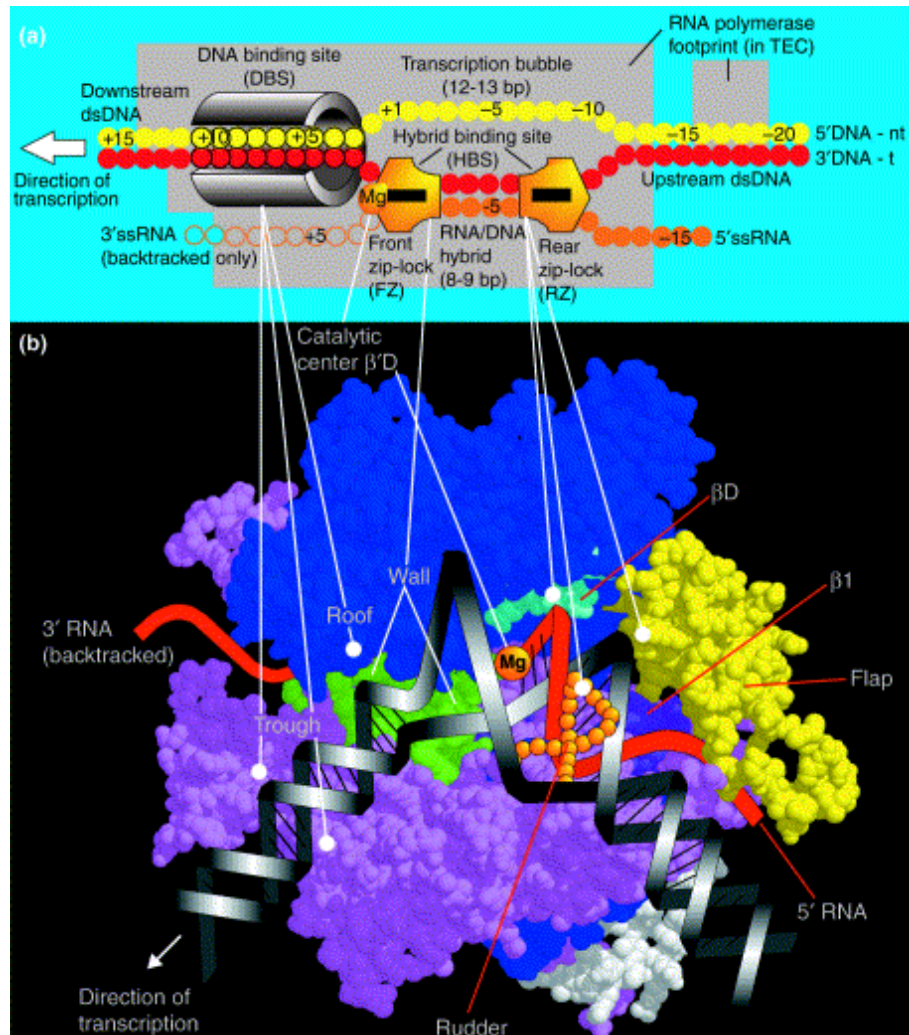


Figure 7: Structure-function model of the transcription elongation complex.

(a) Structure of the nucleic acid scaffold of TEC. Five parts are distinguished: the upstream DNA duplex; the downstream DNA duplex; the 8-9 bp RNA-DNA hybrid; the single-stranded region of DNA in the transcription bubble; and the emergent ssRNA upstream of the hybrid. Filled circles represent DNA (template strand in red; nontemplate strand in yellow). The DNA-binding site is represented by a sliding clamp that encloses 9 bp of the downstream DNA duplex. The hybrid binding site that accommodates the RNA-DNA heteroduplex is represented by two zip-locks that hold onto the edges of the heteroduplex and either zip or unzip the hybrid, maintaining its constant size during lateral movement of RNAP. The gray area represents the RNAP footprint.

(b) Schematic model of TEC showing its main features. White lines show the correspondence between the functional and structural features of TEC.

Reprinted with permission from Elsevier Limited. Korzheva N, Mustaev A. Transcription elongation complex: structure and function. *Curr Opin Microbiol.* 2001, 4(2): 119-25.

This results in the formation of a stable elongation complex in which the transcript is tightly bound in the RNA exit channel. The minimal elongation complex, a ternary complex, consists of the core RNAP bound to the DNA and nascent RNA, with many cooperative contacts increasing the complex stability (Borukhov et al, 2005). The process of RNA-DNA hybrid formation and subsequent RNA displacement from the template, and the mechanism of DNA translocation is a highly complex process involving tremendous structural and stoichiometric changes in the individual components of the machinery (Vassylyev et al, 2007). This process of transcription elongation is extremely complex as the RNAP goes through various conformational states for function and regulation (reviewed by Nudler, 1999; Erie, 2002). The earliest model proposed for elongation of transcript is called the inchworm model, according to which two distinct RNA and DNA-binding sites exist at the leading and lagging edge of the active site of the polymerase (Uptain et al, 1997). The catalytic site for phosphodiester bond formation is at the lagging edge of the RNA binding site. RNA chain elongation proceeds in two phases; nucleotide addition and translocation. As nucleotide addition to the 3'-OH terminus continues, the leading edge is filled with the RNA transcript, conformational changes in the complex releases the template and transcript leading to translocation of the elongation complex (Krummel and Chamberlin, 1992; Uptain et al, 1997). The complexity of this process is enhanced by the fact that this could involve alternating phases of discontinuous (or inchworm-like) as well as monotonic movements of the polymerase.

### **Transcription termination**

Transcription in *E.coli* is terminated by two distinct mechanisms: Rho-dependent and Rho-independent. Rho-independent terminators are intrinsic terminators of inverted repeats coded in the DNA sequence followed by a stretch of A's. The IR forms a hairpin in the RNA which pulls the stem bases away from the DNA template, leaving only the weak U-A pairs to maintain the DNA-RNA hybrid, thus releasing the RNA and making it fall off the DNA due to conformational changes in the RNAP. In Rho-dependent termination, the hexameric protein Rho binds to the nascent RNA of the transcription complex and releases it at defined Rho-dependent terminators along the template. Thus, at termination the RNA transcript is released from the elongation complex and conformational changes in the RNAP core release it from the DNA (Richardson, 1993; Ciampi et al, 2006; Banerjee et al, 2006).

### **Transcription regulation**

Potentially, regulation of transcription can occur at any stage in the process, starting at initiation and continuing upto termination of transcription. Regulatory factors can function in enhancing or disrupting the various stages during transcription. Research in this field have identified and characterized the following regulatory elements in the process of transcription.

## Regulation by $\sigma$ factors

The type of  $\sigma$  factor associated with the core enzyme directs it to activate transcription from a specific promoter. Under normal growth conditions,  $\sigma^{70}$  associates with the core enzyme to direct transcription from the housekeeping genes. The type of  $\sigma$  factor to associate with the core enzyme depends on the cellular needs at that time. Alternative  $\sigma$  factors come into play during stressful or nutrient-depriving conditions. Occlusion of the core-binding determinants of  $\sigma$  by anti-sigma factors depending on environmental needs dictates the availability of the appropriate  $\sigma$  type for interactions with core RNAP (Mathee and Hughes, 1998; Campbell et al, 2003).

## Regulation by small ligands

Small ligands provide an alternative mechanism by which RNA polymerase can respond quickly and efficiently to the environment. The best example is guanosine 3', 5' bisphosphate (ppGpp), which is synthesized when amino-acid availability is restricted to the extent that translation is also limited. The ppGpp ligand works by destabilizing open complexes at promoters. In fact, although the interaction of ppGpp with RNA polymerase is not promoter-specific, ppGpp-dependent inhibition occurs at promoters that control many of the genes that encode the protein products that are needed for translation. Many of these promoters recruit RNA polymerase very effectively and so, potentially, can initiate transcription at the maximum possible rate. However, to achieve these rates, the open complex must be stabilized and this requires higher ATP and lower ppGpp concentrations (Browning and Busby, 2004).

## Regulation by small RNAs

The bacterial 6S RNA was the first sRNA shown to inhibit transcription by binding directly to the housekeeping holoenzyme form of RNA polymerase ( $\sigma^{70}$ -RNA polymerase in *E. coli*) (reviewed by Wassarman, 2007). It resides within the active site of RNA polymerase and blocks access to promoter DNA. This 6S RNA regulation of transcription leads to altered cell survival, perhaps by redirecting resource utilization under nutrient-limiting conditions.

## Regulation by activators and repressors

Based on the nature of regulation, positive regulators (those that turn the genes ON) are called activators and negative regulators (those that turn the genes OFF) are called repressors. Small molecules that cooperatively interact with activators are called inducers and those that interact with repressors are called co-repressors. Most of the positively controlled bacterial promoters lack one or more promoter elements due to which RNA polymerase is unable to bind to these promoters by itself. Activators that function at this step are sequence-specific DNA-binding proteins that bind to upstream

sequences in the promoter. Protein-protein interactions between activator proteins and RNAP subunits demands conformational changes to stereotypically accommodate the interactions (reviewed by Browning and Busby, 2004). Activators that act at the transcription initiation stage target the recruitment of RNAP to the promoter through interactions with various subunits of RNAP and could also mediate transition from closed complex to open complex by stabilizing the intermediate steps of isomerization (Dove et al, 2000). Based on the subunit of RNA polymerase it interacts with and the location of the binding site on the promoter, many activators are grouped into two classes, Class I and Class II. Class I activators have their binding site several bases upstream of the -35 element even as far as -60 from the transcription start site and recruit the RNA polymerase through its interaction with  $\alpha$ -CTD of RNAP. In contrast, Class II activators bind to promoters immediately upstream of or overlapping the -35 element and can affect multiple steps of initiation by its interaction with the  $\alpha$ -CTD,  $\alpha$ -NTD and/or C-terminal domain of the  $\sigma$  subunit of RNA polymerase. Activators can also function by altering the conformation of the target promoter to enable the interaction of RNAP with -10 and/or -35 promoter elements (reviewed by Browning and Busby, 2004).

Most of the repressors act by binding to the promoter thereby preventing RNAP binding by steric hindrance, but they can also affect open complex formation and promoter clearance. Repressors can also function by looping of the promoter elements thereby preventing their interaction with RNAP. They can also function by modulation of functional domains of activator proteins thereby preventing their interaction with RNAP (reviewed by Browning and Busby, 2004). Initial stages of the RNAP binding to the promoters is the primary target for most repressors (Rojo, 1999; Rojo, 2001). All the classical repressors, LacI, PurR, TetR and TrpR are examples of repressors that act by binding to sequences within the promoter and block RNA polymerase binding to the promoter by steric hindrance. Repressors can also function by destabilization of open complexes leading to release of RNAP from the promoter or by stimulation of isomerisation. Most of the repressors that act at the promoter clearance step generally increase the affinity of RNA polymerase for the promoter, thereby preventing the release of the enzyme from the promoter (Rojo et al, 1998). Any stabilization of  $\sigma^{4.0}$ /flap interaction might also result in stronger  $\sigma^{4.0}$ /-35 contacts that prevent promoter clearance.

### **Regulation by small molecules**

The secondary channel of RNA polymerase that transports NTPs to the active site is a target for a new class of regulators (Nickels and Hochschild, 2004). Microcin J25, a 21-residue long peptide with antibiotic properties blocks the NTP entry by binding within the secondary channel (Adelman et al, 2004; Mukhopadhyay et al, 2004). DksA, a regulator of rRNA transcription also mediates regulation by binding within the secondary channel (Perederina et al, 2004). Structural studies of the multi-subunit RNAPs from bacterial and eukaryotic cells have illuminated the presence of large, solvent-accessible surfaces with numerous functionally crucial cavities and channels, the blocking of which would likely inhibit transcription. Among transcription factors that bind to bacterial RNA polymerase (RNAP) and modulate its activity, a number of small molecules irreversibly



inhibit RNAP thereby causing cell death (reviewed by Artsimovitch and Vassylyev, 2006). Small molecule inhibitors such as rifampicin and sorangicin bind to the  $\beta$  subunit within the channel and act by directly blocking the path of the growing RNA chain 2-3 nucleotides in length (Campbell et al, 2001; Campbell, 2005). Recent studies of the RNAP inhibitors indicate that although most of them possess unique mechanisms of action, they are unable to stop RNAP directly through steric hindrance, but rather utilize an indirect approach by limiting the access of the  $Mg^{2+}$  ions and substrates to the RNAP active site. Such indirect modulation of the RNAP catalytic center may in fact represent a common theme in transcription regulation (Artsimovitch et al, 2005; reviewed by Artsimovitch and Vassylyev, 2006; reviewed by Chopra, 2007).

### **Bacteriophage Mu**

Bacteriophage Mu is a temperate bacteriophage of *E. coli* K-12 and also infects several other bacterial species like *Citrobacter freundii*, *Klebsiella*, *Pseudomonas*, *Rhizobium* and *Alcaligenes eutrophus* (reviewed by Symonds et al, 1987; Pato, 1990). It is a generalized transducing phage as well as the largest and the most efficient transposon (Harshey and Jayaram, 2006). Being a temperate phage, Mu has a regulatory circuit that functions as a switch between the lysogenic and lytic modes during its life cycle. Upon infection, a complex of phage proteins and the linear double-stranded 37-kb DNA genome, flanked by heterogenous host sequences, is injected into the host cell (Au et al, 2006). The genome is then converted into a non-covalently closed form through the action of phage-encoded transposition proteins which binds to the ends of Mu DNA forming a protein-DNA complex that has been proposed to serve as an intermediate to integration into the host chromosome (Nakai et al, 2001). Through a random non-replicative transposition mechanism, the phage genome is then integrated into the bacterial host chromosome, leading to duplication of the 5-bp target site which flanks the integrated copy of the Mu genome (Liebart et al, 1982; Chaconas et al, 1983; Sivan et al, 1988).

While a majority of the infected cells enter the lytic cycle, a fraction of the population becomes stable lysogens. Lysogeny is maintained by the repressor  $c$ , by binding to the early promoter,  $P_e$  and is stabilized by three host proteins: IHF (Integration host factor), H-NS (Histone-like nucleoid structuring protein) and FIS (Factor for inversion stimulation) (Betermier et al, 1993; Ranquet et al, 2005). Entry into the lytic cycle is marked by multiple rounds of replicative transposition and concomitant packaging into about 100 viral particles (Chaconas et al, 1981). Mu uses the host RNA polymerase with the primary sigma factor  $\sigma 70$  to transcribe its genes (Toussaint and Lecocq, 1974). Gene expression during the lytic cycle occurs in three phases: early, middle and late and is characterized by a highly controlled regulatory cascade (Marrs and Howe, 1990). The early genes are transcribed from the early promoter  $P_e$ , the middle genes from middle promoter  $P_m$  and the late genes from the late promoters  $P_{lys}$ ,  $P_I$ ,  $P_P$  and  $P_{mom}$  (Margolin et al, 1989) (Figure 8). Thus, the lytic development of phage Mu presents an excellent model system to study the structural, mechanistic and functional aspects of both transposition and transcription activation.

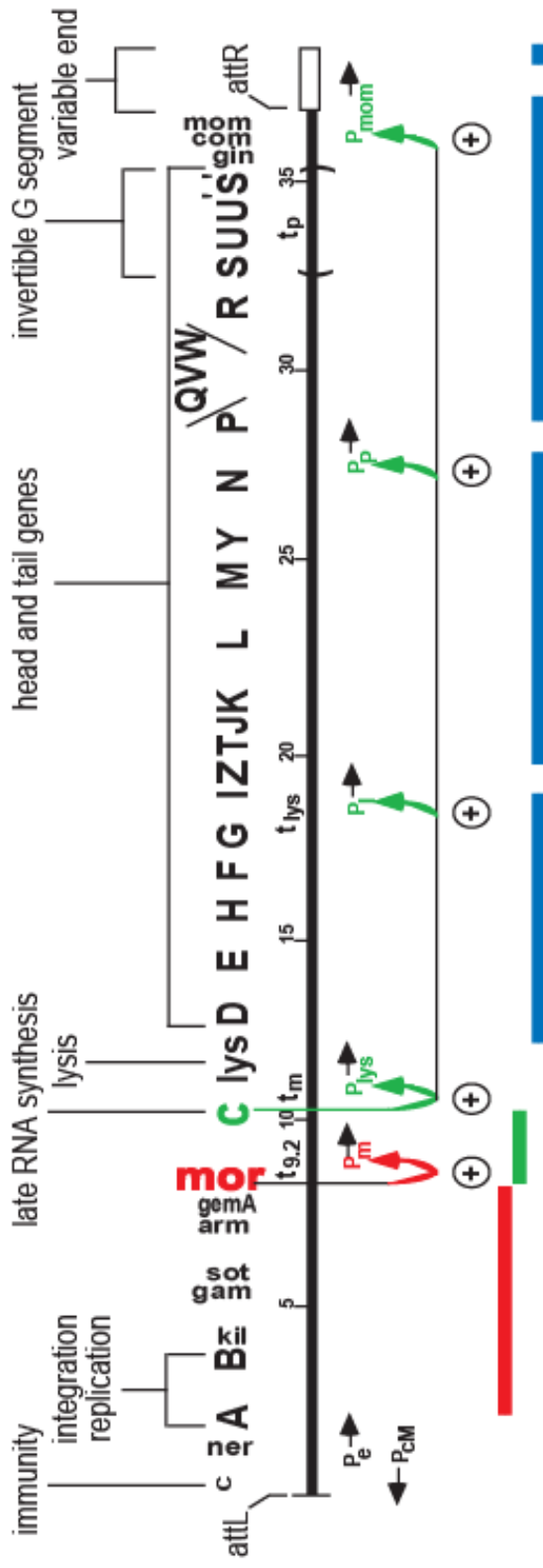


Figure 8: Transcriptional organization of the Mu genome.

The solid line represents the 36.7-kb Mu DNA with marks at 5-kb intervals. The ends of the genome are defined by the attachment site, attL and attR. The rectangle at the right represents attached host DNA present in phage particles. The location of Mu genes and their functions are given above the genome line. The locations of the promoters, the direction of transcription are shown below the line. The curved arrows and the plus signs indicate the transcriptional control points of the actiavtors Mor and C at their respective promoters.

Reprinted with permission from American Society for Biochemistry and Molecular biology. Kumaraswami M, Howe MM, Park HW. Crystal structure of the Mor protein of bacteriophage Mu, a member of the Mor/C family of transcription activators. J Biol Chem. 2004, 279(16): 16581-90.

Expression of the early operon from  $P_e$  is independent of *de novo* protein synthesis and phage replication and is stimulated by IHF (Wijffelman et al, 1974; Krause and Higgins, 1986; Marrs and Howe, 1990). The early promoter  $P_e$  has a good match to a typical *E.coli* promoter with recognizable -10 and -35 elements for recognition and binding by the bacterial RNAP (Krause and Higgins, 1986); it does not require the presence or function of an activator. Transcription from the early promoter can be detected within a few minutes of infection and induction (Wijffelman and van de Putte, 1974) and results in the expression of early genes encoding proteins functioning in the lysis/lysogeny decision (Giphart-Gassler et al, 1981; Van Leerdam et al, 1982; Goosen and van de Putte, 1984; Krause and Higgins, 1984), the integration and replication functions A and B (Wijffelman and Lotterman, 1977; Toussaint et al, 1983), the Cro-like repressor Ner (van Leerdam et al, 1982) and a number of non-essential and growth-enhancing functions. This transcript also encodes the activator for  $P_m$ , the middle operon regulator Mor (Mathee and Howe, 1990).

The middle and four late promoters have recognizable -10 hexamers, but lack sequences similar to the canonical -35 promoter element. Thus, they are subject to positive regulation by activator proteins (Figure 9). Transcription from these promoters depends on phage-encoded transcription activator proteins: Mor for  $P_m$  and C for the late promoters (Margolin et al, 1989; Mathee and Howe, 1990; Stoddard and Howe, 1990).

Middle operon transcription initiating at  $P_m$  requires phage replication and the early gene product Mor. The middle transcript is detectable about 8 minutes into the lytic cycle and increases steadily thereafter (Stoddard and Howe, 1989; Marrs and Howe, 1990; Stoddard and Howe, 1990; Mathee and Howe, 1990). Sequence analysis of the middle operon led to the prediction of five open reading frames with the most distal one encoding C, the activator of late transcription (Hattman et al, 1985; Heisig and Kahmann, 1986; Margolin and Howe, 1986; Stoddard and Howe, 1987). With the exception of C, the other proteins encoded by the open reading frames of the middle operon do not seem to be important for phage development (Mathee and Howe, 1993).

The genes of the late phase, expressed from the C-regulated late promoters ( $P_{lys}$ ,  $P_I$ ,  $P_P$  and  $P_{mom}$ ) encode functional components for phage morphogenesis, cell lysis and DNA modification. They are expressed about 20 minutes after induction and continue till lysis (Margolin et al, 1989; Stoddard and Howe, 1989). These promoters share an imperfectly conserved sequence overlapping the position of the absent -35 consensus hexamer (Stoddard and Howe, 1989; Margolin et al, 1989; Marrs and Howe, 1990).

The roles of specific Mu late genes are as follows:

1. The first gene of the late phase, *lys*, promotes cell lysis at the end of the lytic cycle (Faalen and Toussaint, 1973).
2. Genes D-J as per their organizational positions on the Mu genome are required for phage head formation (Shore and Howe, 1982; Grundy and Howe, 1985).
3. Genes K-W as per their positions on the genome, encode the structural and functional components of the tail assembly process (Grundy and Howe, 1985).

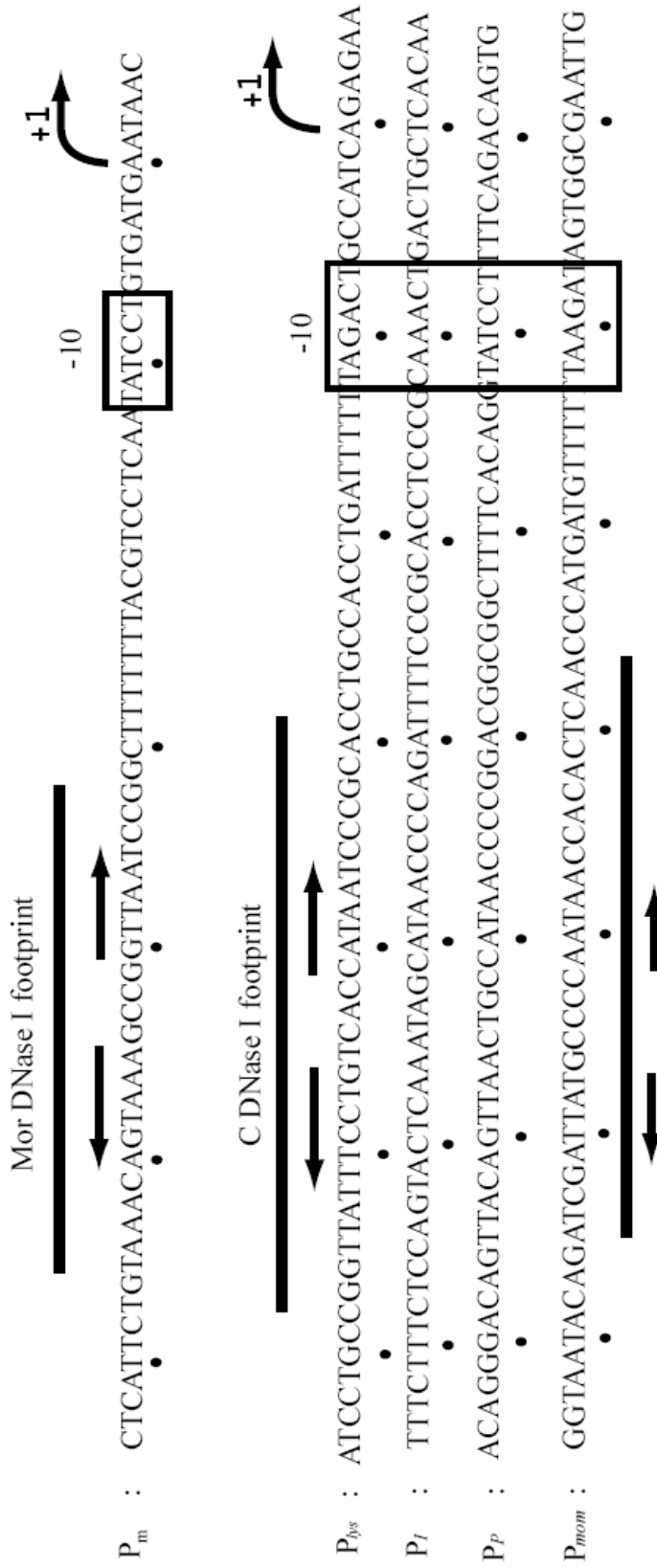


Figure 9: Mu middle and late promoters.

The middle promoter and the four late promoters are aligned based on Mor/C binding sites, -10 hexamers and the transcription start sites. The solid bars denote the region protected by Mor and C in their respective promoters. The imperfect dyad symmetry element within the protein footprints is indicated by inverted arrows. The -10 hexamer is marked by a box around it. The curved arrow represents the transcription start site +1. The dots denote 10-bp intervals.

Reprinted with permission from Kumaraswami M. Structural and functional characterization of Phage Mu Mor gene. 2005, University of Tennessee Health Science Center, Memphis.

4. The host range for infection is specified by the products of the genes S, S', U and U' which are located within and overlapping the invertible G segment (Grundy and Howe, 1984). Different tail fiber proteins result from the two alternative arrangements of these genes, thus specifying adsorption to different lipopolysaccharide receptors (Sandulache et al, 1984) on the cell surface of either *E.coli* K12 or a heterogeneous group of enteric bacteria (Kamp et al, 1978; van de Putte et al, 1980).
5. The inversion of the G segment is catalyzed by the site-specific recombinase encoded by the *gin* gene (Kahmann et al, 1985).
6. The *mom* gene encodes a DNA modification function which converts about 15% of all adenines to  $\alpha$ -N-(9- $\beta$ -D-2'-deoxyribofuranosylpurin-6-yl) glycinamide, which confers resistance of the infecting Mu DNA to cleavage by a wide range of restriction enzymes (reviewed by Kahmann, 1983; reviewed by Hattman, 1999).
7. The *com* gene encodes the small zinc-coordinating translational activator of *mom* gene expression, Com (Wulczyn et al, 1989; Witkowski et al, 1995; Wulczyn and Kahmann, 1991).

### **The middle promoter of bacteriophage Mu, P<sub>m</sub>**

A strong basal promoter should ideally be close to the consensus sequence elements for efficient RNAP binding to initiate transcription. The region downstream of the -10 hexamer including the transcription start site should be prone to melting, and the upstream elements should contribute to RNAP binding. The importance of the spacer region is underscored by the fact that it should be of permissible length and show functional significance for productive RNAP binding, melting of the promoter region and conformational changes in RNAP required for efficient initiation. While some promoters in *E. coli* are active without the help of any accessory factors, many promoters need additional factors for function (Collado-Vides et al, 1991). The common feature of these promoters is the presence of a proximal site, usually between -20 and -70, facilitating the binding of activators specific for the promoters. Thus, it seems that the entire promoter region from about -70 to the transcription start site is functionally important contributing to various factors for productive transcription. Deletion and footprinting analyses in the promoter region revealed that sequences upstream of -62 and downstream of +10 are dispensable for promoter activity in terms of binding of Mor and RNAP although binding of the  $\alpha$ -CTD to an UP-like element leads to increased levels of transcription (Ma and Howe, 2004).

The initiation of transcription at P<sub>m</sub> represents an important regulatory step in the production of viable phage progeny in the lytic. Transcription from P<sub>m</sub> is dependent on phage replication, thus coordinating Mu DNA synthesis, the production of proteins for cell lysis and Mu genome packaging. The product of the middle operon, C protein triggers the expression of late genes which result in packaging of the Mu DNA into phage particles and cell lysis. The middle promoter P<sub>m</sub> is a model promoter for positive regulation of transcription. It has a recognizable -10 hexamer, but lacks similarity to the consensus -35 hexamer. Initiation of transcription from this promoter depends on the

presence of RNAP holoenzyme and the phage-encoded activator Mor, both *in vivo* and *in vitro* (Mathee and Howe, 1990; 1993). Mutational analysis revealed that there are three regions important for promoter activity:

1. The -10 region: involved in binding of the  $\sigma 70$  subunit of RNAP.
2. An imperfect dyad symmetry element: DNase I footprinting analysis showed that Mor protects  $P_m$  sequences from -33 to -56 (Kahmeyer-Gabbe and Howe, 1996). An imperfect dyad symmetry element with a 6-bp spacer was identified between -36 and -51 (AGTAAagccggTTAAT).
3. Region -29 to -31: between the Mor footprint and -10 region, which usually overlaps -35 consensus, has no similarity to the consensus. Both up and down mutations in this region had no detectable effect on Mor binding. DNase I footprinting with Mor resulted in hypersensitivity to cleavage between -29 and -32, indicating an alteration of DNA structure. Hypersensitivity to the single-strand specific agent potassium permanganate was also detected at -32 and -33 during simultaneous binding of Mor and RNAP holoenzyme. Thus, this region was hypothesized to undergo a conformational change during protein binding and promoter activation causing a distortion or unwinding necessary for precise alignment of Mor and RNAP (Artsimovitch and Howe, 1996).

### **The middle operon regulator Mor**

The most distal ORF of the early operon encodes a small protein of molecular mass 14.7 kDa, which activates transcription from the middle promoter  $P_m$ ; hence, it is called the Middle operon regulator protein Mor. It is the only Mu-encoded protein required for transcription of the middle operon both *in vivo* and *in vitro* (Mathee and Howe, 1990; 1993). It is a 129 amino acid, sequence-specific DNA-binding protein (Mathee and Howe 1990) with an isoelectric point of 6.3. The N-terminus of Mor is composed of a mixture of acidic and hydrophobic residues, whereas the C-terminal region is rich in basic residues, characteristic of a DNA-binding region. Mor is closely related and highly homologous to the C protein of bacteriophage Mu (about 31% sequence similarity) (Mathee and Howe, 1990) and they carry out similar functions. They are the founding members of a new family of transcription factors called the Mor/C family; they exhibit little or no homology to any other proteins except for Mor/C proteins in Mu-like prophages identified in bacterial genome sequences. Both proteins bind to DNA as dimers to a site located immediately upstream of the -35 region in their respective promoters. Consistent with their dimeric nature, an imperfect dyad symmetry element is present in the binding sites of both Mor and C. Like Mor, C protein also has an acidic and hydrophobic N-terminal region and a basic C-terminal region.

The presence of an imperfect dyad symmetry element within the Mor-binding site (Artsimovitch and Howe, 1996) and the observation of dimeric species in glutaraldehyde cross-linking experiments suggested a dimeric nature for the protein. Mor binding introduces a distortion in the promoter DNA in the form of a bending angle of  $\sim 40^\circ$  upon binding (Mo, 2004). Footprinting studies and mutational analysis of the middle promoter

identified the region from -36 to -51 of  $P_m$  as Mor binding site. The Mor binding site is characterized by the presence of 6-bp long AT-rich imperfect dyad symmetry element that is separated by a 4-bp long GC-rich spacer (Artsimovitch and Howe, 1996). Middle promoter activation by Mor requires the C-terminal domains of both alpha and sigma subunits of RNA polymerase (Artsimovitch et al, 1996). Experiments with purified Mor and  $\alpha$ -CTD proved that Mor binding to  $P_m$  is strengthened by the addition of  $\alpha$ -CTD and Mor might use its interaction with the  $\alpha$ -CTD to recruit the RNA polymerase to the promoter through a synergistic stabilization of various protein-DNA interactions. Consistent with this, an UP-like element located 7-8 bp upstream of Mor binding site was identified as a binding site for the  $\alpha$ -CTD of RNAP (Ma and Howe, 2004). Based on the above findings, a working model was proposed in which Mor binds to the promoter in a region centered at -43.5 and carries out transcription activation by recruiting RNA polymerase through its interaction with the C-terminal domains of the  $\alpha$  and  $\sigma$  subunits of RNA polymerase (Figure 10) (Artsimovitch and Howe, 1996).

Solving the crystal structure of Mor at a resolution of 2.2 Å was a landmark event, giving great insights into our understanding of the structural and functional significance of different regions of Mor with respect to  $P_m$  and RNAP interactions (Kumaraswami et al. 2004). The ribbon representation of the Mor dimer is shown in Figure 11. The crystal structure of Mor revealed that the Mor monomer is folded into two independent domains. The N-terminal domain is composed of two  $\alpha$ -helices,  $\alpha 1$  and  $\alpha 2$ , running in opposite directions with an angle of 120° to each other. The C-terminal domain contains three helices,  $\alpha 3$ ,  $\alpha 4$ , and  $\alpha 5$  folded into the Helix-turn-Helix (HTH) DNA-binding motif. The N- and C-terminal domains are connected by a  $\beta$ -strand linker (Figure 11). Three glycine residues (Gly<sup>65</sup>, Gly<sup>66</sup>, and Gly<sup>67</sup>) N-terminal to the  $\beta$ -strand and one glycine (Gly<sup>74</sup>) C-terminal to it potentially create two flexible junctions between the N- and C-terminal domains. Helices  $\alpha 1$  and  $\alpha 2$  of the N-terminal domain and the  $\beta$ -strand, form the structural elements for Mor dimerization and are arranged with respect to the 2-fold symmetry-related axis to form an intertwined four-helix bundle with a pair of antiparallel  $\beta$ -strands capping one end of the bundle (Kumaraswami et al, 2004). At the Mor dimer interface, many hydrophobic residues, including leucines and isoleucines, interact with their symmetry-related equivalents in four layers of hydrophobic interactions forming an extensive hydrophobic core, thus stabilizing the dimer. The three helices,  $\alpha 3$ ,  $\alpha 4$ , and  $\alpha 5$ , in the Mor C-terminus form a three-helix bundle with a classical helix-turn-helix DNA-binding motif. In Mor, helix  $\alpha 3$  serves the scaffolding role anchoring and stabilizing the other helices. Mor helices  $\alpha 4$ ,  $\alpha 5$  and the turn between them contain conserved residues characteristic of the DNA-binding HTH motif.

Comparison of the structural elements with other HTH proteins provided evidence for an "ends-on" base recognition, with helix  $\alpha 5$  serving as the Mor recognition helix (Kumaraswami et al, 2004). Modeling of the Mor- $P_m$  interaction led to the prediction that Mor would bend the  $P_m$  sequence away from Mor. This can place the two adjacent major grooves of  $P_m$  in an optimum position to interact with the HTH domains of the Mor dimer at the two ends (Figure 12). Thus, conformational changes in both DNA and Mor are predicted to take place to achieve optimum structural and functional interactions (Kumaraswami et al, 2004).

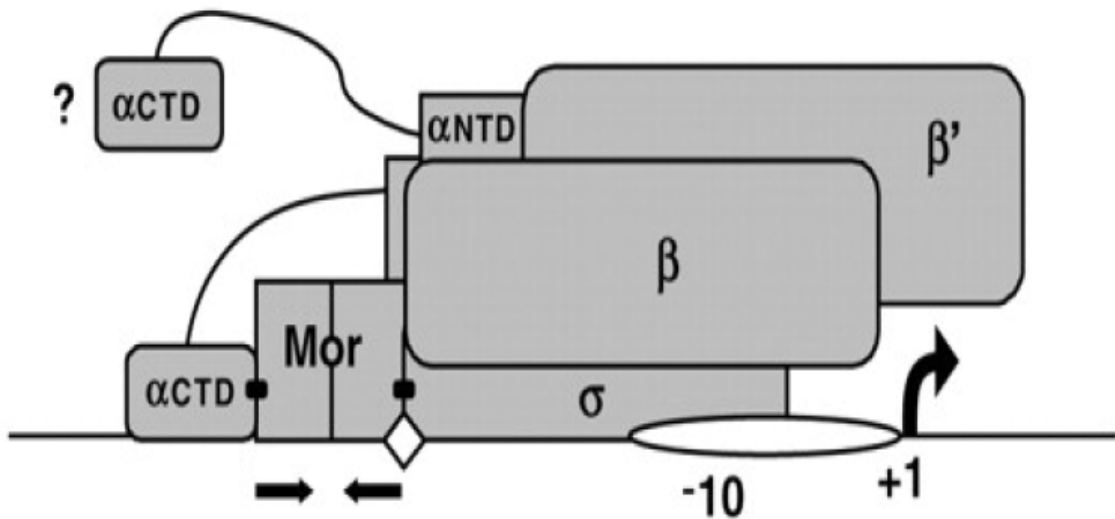


Figure 10: Diagrammatic representation of protein-protein and protein-DNA interactions in the middle promoter open complex.

Middle promoter is represented by the horizontal line. The dyad symmetry element for Mor binding is shown as inverted arrows; the spacer distortion from -34 to -32 by open diamond; the strand separation in the open complex from -12 to +1 by an open oval. The black rectangles represent the interfaces of Mor-RNAP interactions.

Reprinted with permission from American Society for Biochemistry and Molecular biology. Kumaraswami M, Howe MM, Park HW. Crystal structure of the Mor protein of bacteriophage Mu, a member of the Mor/C family of transcription activators. *J Biol Chem.* 2004, 279(16): 16581-90.



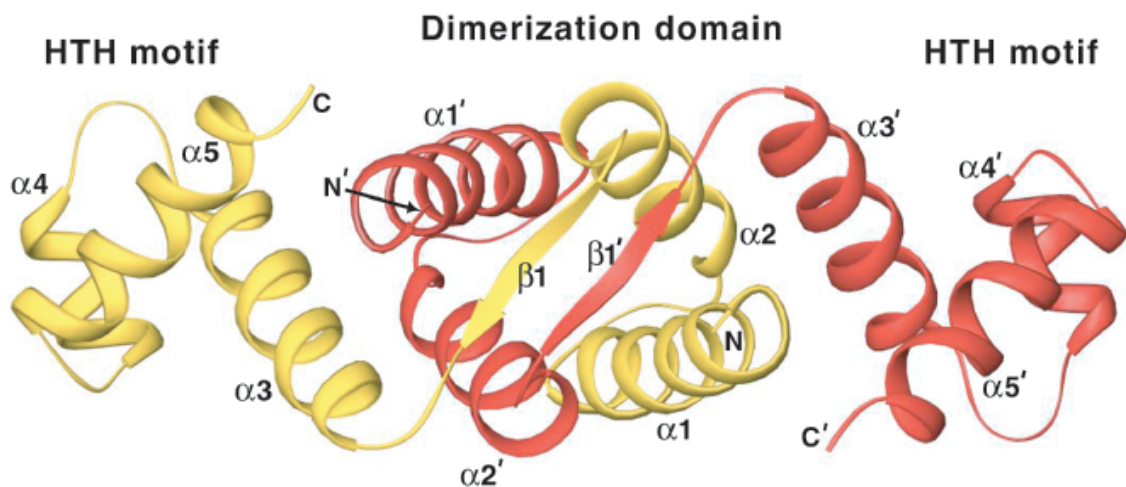


Figure 11: Ribbon representation of Mor dimer.

It is viewed parallel to the crystallographic two-fold axis. One monomer is in yellow and the other in red. The secondary structure elements of the second monomer are identified with a prime.

Reprinted with permission from American Society for Biochemistry and Molecular biology. Kumaraswami M, Howe MM, Park HW. Crystal structure of the Mor protein of bacteriophage Mu, a member of the Mor/C family of transcription activators. *J Biol Chem.* 2004, 279(16): 16581-90.

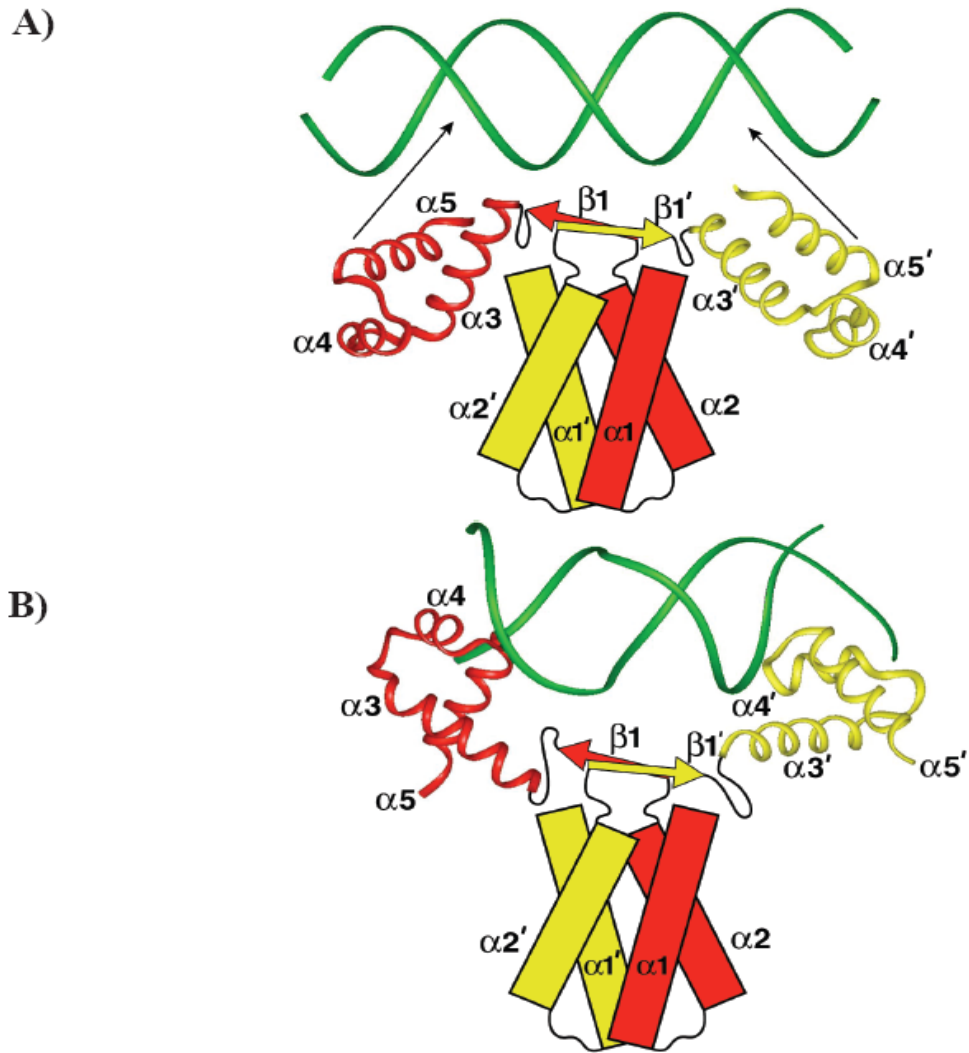


Figure 12: Proposed model for DNA-binding by Mor.

The dimerization domains are represented as cylinders for  $\alpha$ -helices and arrows for  $\beta$ -strands. Mor HTH motifs are shown as ribbon representations. The arrows show the proposed interactions of the HTH motifs with the two adjacent major grooves. Panel B shows the DNA with a  $\sim 40^\circ$  bend as the HTH domains dock into the major grooves (Kumaraswami, 2005).

Reprinted with permission from American Society for Biochemistry and Molecular biology. Kumaraswami M, Howe MM, Park HW. Crystal structure of the Mor protein of bacteriophage Mu, a member of the Mor/C family of transcription activators. *J Biol Chem.* 2004, 279(16): 16581-90.

## Chapter 2: Mutational analysis of $P_m$ and their possible effects on binding and activation by Mor

### Introduction

Bacteriophage Mu is a temperate phage with its life cycle divided into two phases: lytic and lysogenic phases (Howe, 1973; Hattman, 1999). Transcription during the lytic cycle has a regulatory cascade composed of three phases of gene expression: early, middle and late (Marrs and Howe, 1990). Thus, the life cycle of phage Mu presents a valuable model system for studying transcriptional regulation (Howe, 1973; Hattman, 1999). The early promoter  $P_e$ , has both -10 and -35 hexamers (Krause and Higgins, 1986), a typical promoter architecture of basal promoters. In contrast, the middle promoter  $P_m$ , and the four late promoters are model promoters for positive regulation; they have recognizable -10 hexamers but lack -35 hexamers (Margolin et al, 1989; Stoddard and Howe, 1990). Transcription from these promoters requires phage-encoded activators. Transcription from  $P_m$  requires a product of the early transcript, Mor and transcription from the late promoters requires a product of the middle transcript, C (Margolin and Howe, 1990; Mathee and Howe, 1993).

The middle operon regulator (Mor) protein is the product of the last gene in the early transcript and it is the only Mu protein required for transcription from the middle promoter  $P_m$ . Mor is a 129-amino acid long, dimeric, sequence-specific DNA-binding protein (Mathee and Howe, 1990; Kahmeyer-gabbe and Howe, 1996; Artsimovitch and Howe, 1996). Mor does not share amino acid sequence similarity with other families of transcription factors and, together with the closely related phage-encoded activator C, defines a new family of transcription factors called the Mor/C family of transcription activators (Kumaraswami et al, 2004).

The dimeric nature of Mor is supported by the following evidence:

- a) Extensive mutagenesis of the Mor-binding site in  $P_m$  revealed an imperfect dyad-symmetry element with a central 4-bp spacer, suggesting the requirement of a dimeric form of Mor to recognize the symmetrical sequences (Artsimovitch and Howe, 1996).
- b) Glutaraldehyde cross-linking experiments identified a cross-linked species corresponding to the size of a Mor dimer (Artsimovitch and Howe, 1996).
- c) Mutational analysis of the Mu late promoter,  $P_{lys}$  and SMCC cross-linking experiments with the Mu late gene activator C, a closely-related homologue of Mor, provided evidence for a dimeric form of C bound to a dyad-symmetrical binding site (Chiang and Howe, 1993; Ramesh and Nagaraja, 1996; Jiang, 1999; Zhao, 1999).
- d) The crystal structure of Mor captured the dimeric form of Mor through the two-fold crystallographic symmetry axis (Figure 11) (Kumaraswami et al, 2004).

Mor recognizes and binds to a 16-bp region in  $P_m$  located at -36 to -51 bases upstream of the start site +1 (Artsimovitch and Howe, 1996). Mor activates transcription from  $P_m$  by recruiting the host RNA polymerase to the middle promoter (Mo, 2004). The basis for Mor-dependent  $P_m$  activation include protein-protein interactions between Mor and the C-terminal domains of the  $\alpha$  and  $\sigma$  subunits of RNA polymerase which are required for efficient recruitment of RNA polymerase to  $P_m$  (Artsimovitch et al, 1996; Ma and Howe, 2004; Mo, 2004).

Mutational analysis of  $P_m$  revealed three important regions for its function:

1. The -10 region which is important for  $P_m$ - $\sigma$  interactions (Artsimovitch and Howe, 1996).
2. The 16-bp Mor-binding site located immediately upstream of the missing -35 hexamer from -36 to -51. Extensive mutagenesis of  $P_m$  defined two AT-rich 5-bp long repeats in the imperfect dyad symmetry element that are separated by a GC-rich 5-bp spacer (Artsimovitch and Howe, 1996).
3. An UP-like element situated upstream of the Mor-binding site from -54 to -62 comprising the binding element for the  $\alpha$ -CTD of RNA polymerase (Ma and Howe, 2004).

The crystal structure of His-tagged Mor at a resolution of 2.2 Å revealed a dimeric Mor protein with the two subunits being related to each other by a crystallographic two-fold symmetry axis. The first 26 amino acids of the N-terminus and the last 9 amino acids from the C-terminus were not visible in the structure. The Mor monomer is folded into two recognizable independent domains: an N-terminal dimerization domain and a C-terminal helix-turn-helix DNA-binding domain (Kumaraswami et al, 2004).

The N-terminal domain is composed of two  $\alpha$ -helices,  $\alpha_1$  and  $\alpha_2$ . These two helices run in opposite directions with an angle of 120° relative to each other. The interlocking of N-terminal helices of both monomers against each other along with the anti-parallel interactions between the  $\beta$ -strands constitutes the dimerization domain. At the Mor dimer interface, many hydrophobic residues interact with their symmetry-related equivalents. Key residues involved in dimerization include several leucines of helix  $\alpha_1$  and a mixture of isoleucines and leucines of helix  $\alpha_2$  (Kumaraswami et al, 2004).

The C-terminal domain of Mor contains three helices,  $\alpha_3$ ,  $\alpha_4$  and  $\alpha_5$ , which are folded into the classic DNA-binding HTH motif. The specific positioning of the two HTH motifs at the opposite ends of the Mor dimer would structurally complement their interaction with the bases in the symmetry element. In Mor, helix  $\alpha_3$  functions as the structural scaffold anchoring  $\alpha_4$  and  $\alpha_5$ . The recognition helices  $\alpha_5$ , and the turn between  $\alpha_4$  and  $\alpha_5$  contain conserved residues characteristic of the DNA-binding HTH motif and structural coordinates similar to those of HTH proteins with an “ends-on” base recognition.

The N- and C-terminal domains are connected by a  $\beta$ -strand linker. Three glycine residues (G65, G66 and G67) are located in the loop N-terminal to the  $\beta$ -strand and one glycine (G74) is present in the C-terminal loop immediately after the  $\beta$ -strand, potentially creating two flexible junctions between the N- and C-terminal domains. Docking studies with a 16-bp B-DNA structure indicated that the two HTH motifs are located too far apart to interact with two adjacent major grooves in linear DNA. In order to overcome the calculated 9Å discrepancy, it was proposed that both protein and DNA might need to undergo conformational changes to allow Mor binding to DNA (Kumaraswami et al, 2004). Consistent with this, a structural distortion with a 45° bending angle was observed in  $P_m$  upon Mor binding (Mo, 2004). The highly conserved glycines flanking the  $\beta$ -strand are hypothesized to be the pivot points for the conformational changes in Mor.

Another interesting feature observed in His-Mor structure is the location and direction of the side chains of Y70 and Q68 in the  $\beta$ -strand region. These two side chains are protruding away from the molecule in upward direction and not involved in any intra or inter molecular interactions. In our current working model, in the presence of DNA, conformational changes originating from glycines (G65, G66, G67 and G74) place the N-terminus of the recognition helix  $\alpha 5$  in position to interact with the bases in the adjacent major grooves. This arrangement of protein and DNA would place the  $\beta$ -strand region under the minor groove of the central spacer between the 6-bp inverted repeats, such that the two residues Y70 and Q68 might interact with the minor groove of the spacer (Figure 12).

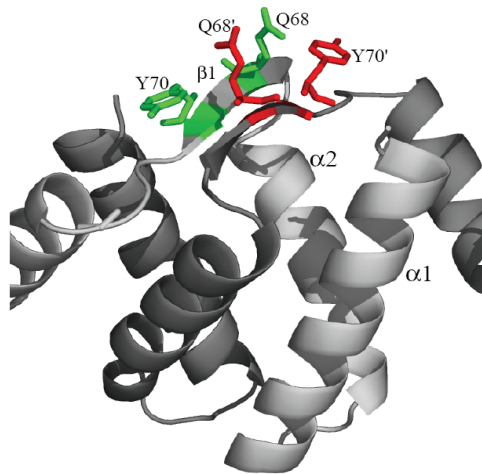
### **The current work**

Sequence-specific DNA-binding proteins read the information in DNA through the two structural features of a B-form DNA: major groove and minor groove. In most cases, the specific recognition of DNA takes place principally in the major groove due to its better accessibility and greater information content (Bewley et al, 1998; Jones et al, 1999). The minor groove has a minor role in such interactions because of its narrower and deeper nature (Travers, 1995). Proteins with such dual binding mode generally have a HTH DNA-binding motif, which is used in major groove interactions (Jones et al, 1999).

The crystal structure of Mor identified two regions as possible candidates to make DNA contacts (Kumaraswami et al, 2004). The HTH motifs located in the C-terminal domain of Mor dimer are predicted to make interactions in the major groove of dyad symmetry element and the mutational analysis results were in agreement with the structural predictions. The HTH motif-major groove interactions place the  $\beta$ -strand linker region in the proximity of the minor groove of the spacer region located on the same face of the two major grooves.

One remarkable feature of the  $\beta$ -strand linker region is the side chains of Q68 and Y70 that are protruding away Mor, but towards the possible direction of DNA indicating possibility of minor groove interactions (Figure 13). The primary sequence comparisons

A



B

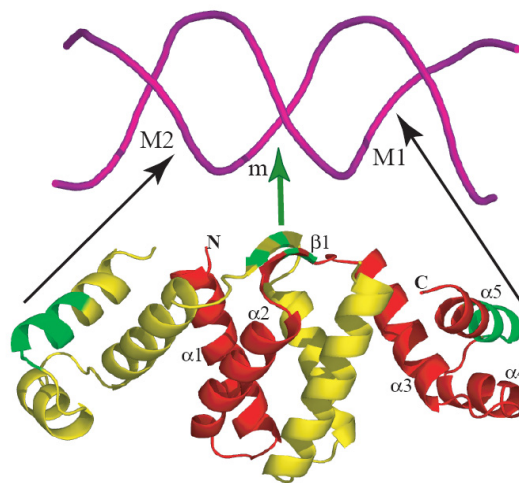


Figure 13: Representation of Q68 and Y70 and possible Mor-DNA contacts.

A) Representation of Q68 and Y70. The side chains of Q68 and Y70 from different monomers are color-coded (red from the dark gray monomer and green from the gray monomer).

B) Possible Mor-DNA contacts. His-Mor dimer structure is as ribbons with the monomers in red and yellow; DNA in magenta. Green arrow represents the predicted Mor-minor groove interactions as a result of the Mor-major groove interactions which are shown as black arrows.

Reprinted with permission from Kumaraswami M. Structural and functional characterization of Phage Mu Mor gene. 2005, University of Tennessee Health Science Center, Memphis.

of the amino acids in the  $\beta$ -strand linker region of the members of Mor/C family revealed a high degree of sequence homology at these two positions indicating the importance of those to Mor function and suggesting a common function in all the members.

Comprehensive mutational analysis of Q68 and Y70 and inhibition assay with a GC-specific minor groove binding drug, chromomycin elucidated the important contribution of the residues (Kumaraswami, 2005). As in case of most of the transcription regulators, Mor needs to bind DNA in order to carry out transcription activation from  $P_m$ . The mutant proteins failed to activate transcription underlining the importance of the side chains of Q68 and Y70 for Mor function, possibly by participating in the DNA-binding function (Kumaraswami, 2005). Chromomycin inhibited Mor binding to promoter sequences at concentration as low as 2  $\mu$ M indicating the importance of minor groove to Mor binding. Chromomycin was unable to completely disrupt the preformed Mor-DNA complex even at concentrations as high as 50  $\mu$ M, thus, suggestive of Mor-minor groove interactions. The inhibition might be due to the disruption of Mor-minor groove contacts by chromomycin binding to the GC-rich spacer sequences or the structural effects extending into the flanking major grooves as a result of chromomycin binding to the minor groove.

Protein-minor groove contacts have been classified into three major categories: intercalation, specific interactions and indirect readout. To gain access to the information content of narrower minor groove, protein pry open the minor groove by inserting the side chains of amino acids between adjacent bases by interdigitations which unstack the adjacent bases and introduce a kink or bend in the direction of the duplex DNA. In support of intercalation, a bending angle about 45° was observed in  $P_m$  upon Mor binding. Though the deficiency of chemical features presented by the minor groove is generally considered to be insufficient for specific recognition (Travers, 1995), some structures revealed specific contacts between different side chains and base edges exposed in the minor groove through hydrogen bonding, van der waals interactions, and water-mediated contacts (Moravek, 2002). In most cases it occurs in tandem with intercalation as it facilitates the specific interactions by wedging open the minor groove. Two pieces of evidence discounted the possibility of such specific interactions between Mor and the minor groove bases of the spacer region. First, single base substitutions identified in the spacer region of  $P_m$  appeared to be less detrimental to Mor binding compared to the mutations in the symmetry element. The tolerance to various mutations in the spacer region indicates the lesser role played by this region. Second, the conserved nature of Q68 and Y70 disfavors the specific contacts between identical side chains and chemically dissimilar spacer sequences of middle and four different late promoters. It is observed that amino acids that make specific contacts with different sequences are generally not conserved. It might be useful to test the effect of symmetrical substitutions in the spacer region on Mor binding. In indirect read-out, the conformational flexibility of the non-contacted bases in a protein binding site influences the affinity of a protein-DNA complex (Perez Martin J and de Lorenzo V, 1997; Mauro et al, 2003). Structural alterations in the DNA are crucial for the proper spatial relationship between the protein and DNA. Considering the high flexibility predicted for the GC-rich spacer region (Gabrielian and Pongor, 1996), indirect readout between Mor and the minor groove of the

spacer region is a possibility. It is possible that DNA oscillates between different conformations due to the flexibility conferred upon by the GC-rich spacer region and interactions between Mor and the minor groove sequences might trap the DNA in the bent conformation.

Deletion mapping and DNase I footprinting revealed that Mor protects  $P_m$  sequences -33 to -56, thus indicating that Mor-binding sequence in  $P_m$  overlaps with this region (Kahmeyer-Gabbe and Howe, 1996). Thus, Mor binds to the promoter in a region centered at -43.5 and carries out transcription activation by recruiting RNA polymerase through its interaction with the C-terminal domains of the  $\alpha$  and  $\sigma$  subunits of RNA polymerase (Artsimovitch and Howe, 1996; Ma, 2004). Also, previous mutational analysis of  $P_m$  identified two mutations upstream of the Mor-binding site, G-57T with twice the activity of the wild-type and A-54T with half the wild-type activity, suggesting that sequences upstream might play a role in  $P_m$  function (Artschimovitch, 1996).

The goal of this study is to study the effect of mutations in  $P_m$  from -30 to -57 spanning the region previously shown to be important for Mor-  $P_m$  interactions. This would also help us to optimize the Mor-binding region of  $P_m$  useful for crystallographic studies of Mor-DNA duplex. Extensive mutagenesis studies were also conducted to study the effect of mutations in the spacer region of -42 to -45 because this is the spacer region in which the side chains of Y70 and Q68 from the  $\beta$ -strand protrude into. This was done to test if Mor-DNA interactions in the minor groove occurred through specific base contacts. In addition to the extensive mutagenesis of the spacer region, targeted mutagenesis was carried out for the entire region with specific mutations chosen for each position as given in Table 1. The rationale behind each of the specific mutation is discussed in the Table 2.

For the spacer region, mutagenesis was done with degenerate oligonucleotides for the 4 positions in pair and as a group. Plate phenotyping of the mutants, gel shift assays for the mutants with wild-type His-Mor and  $\beta$ -galactosidase assays were used to differentiate the effect of the mutations. Gel mobility shift assays were done with oligonucleotide probes. The length of the oligonucleotide probe needed for efficient Mor binding was first standardized and various mutations were then introduced into the specific length probe.

The center of the Mor-binding site in  $P_m$  is -43.5 position between -43 and -44 from the transcription start site of +1. The mutations that we made either introduced symmetrical changes in positions that were not symmetrical centered at -43.5 or disrupted the symmetry by introducing a different base. The symmetrical changes refer to positions the same number of bases from the center of -43.5. As Mor binds  $P_m$  as a dimer, these mutations would help us study the effect of symmetry element for binding of the dimer. They would also help us study the contribution of the specific bases for Mor binding. The contribution of the bases to the stability of Mor-Promoter duplex can be studied by mutagenesis of this region of  $P_m$ .



Table 1: Specific mutations of the promoter.

A comprehensive table which gives all the mutations that were done for  $P_m$  spanning the region from -57 to -30 with respect to the transcriptional start site of +1. The second row gives the bases in the wild-type sequence of  $P_m$  and their positions are numbered above them. Following that, the 29 mutations are given sequentially. The mutation numbers are in the first column and the letters in red are the mutations that were introduced at that position. The unfilled boxes indicate that the bases in these positions are same as that in the wild-type promoter. The blue underlined bases indicate the dyad-symmetry element predicted to interact with Mor in the major groove. The purple letters indicate the position of the bases of the minor groove spacer region.

|    | -57 | -56 | -55 | -54 | -53 | -52 | -51 | -50 | -49 | -48 | -47 | -46 | -45 | -44 | -43 | -42 | -41 | -40 | -39 | -38 | -37 | -36 | -35 | -34 | -33 | -32 | -31 | -30 |   |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---|
|    | G   | T   | A   | A   | A   | C   | A   | G   | T   | A   | A   | A   | G   | C   | C   | G   | T   | T   | T   | T   | A   | T   | C   | C   | G   | G   | C   | T   |   |
| 1  |     |     |     |     |     |     |     |     |     |     |     | C   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
| 2  |     |     |     |     |     |     |     |     | T   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
| 3  |     |     |     |     |     |     |     |     |     | C   |     |     |     |     |     |     |     |     | G   |     |     |     |     |     |     |     |     |     |   |
| 4  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | C   |     |     |     |     |     |     |     |   |
| 5  |     |     |     |     |     | G   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
| 6  |     |     |     |     |     | A   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | T   |     |     |     |     |     |   |
| 7  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | T   |     |     |     |     |   |
| 8  |     |     |     |     |     |     | G   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
| 9  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | T   |     |     |     |   |
| 10 |     |     |     |     |     |     |     | C   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
| 11 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | T   |     |     |   |
| 12 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | A   |     |   |
| 13 |     |     |     |     |     |     | A   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | T   |   |
| 14 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |   |
| 15 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | A |

|    | -57 | -56 | -55 | -54 | -53 | -52 | -51 | -50 | -49 | -48 | -47 | -46 | -45 | -44 | -43 | -42 | -41 | -40 | -39 | -38 | -37 | -36 | -35 | -34 | -33 | -32 | -31 | -30 |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | G   | T   | A   | A   | A   | C   | A   | A   | T   | A   | A   | A   | G   | C   | C   | G   | G   | T   | T   | A   | A   | T   | C   | C   | G   | G   | C   | T   |
| 16 |     |     |     |     |     |     |     | T   |     |     |     |     |     |     |     | T   |     |     |     |     |     |     |     |     |     |     |     |     |
| 17 |     |     |     |     |     |     |     | T   |     |     |     | C   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 18 |     |     |     |     |     |     |     |     |     |     |     |     | A   | A   | A   | A   |     |     |     |     |     |     |     |     |     |     |     |     |
| 19 |     |     |     |     |     |     |     |     |     |     |     |     | T   | A   | A   | T   |     |     |     |     |     |     |     |     |     |     |     |     |
| 20 |     |     |     |     |     |     |     |     |     |     |     |     | C   |     | G   |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 21 |     |     |     |     |     |     |     |     |     |     |     |     |     | G   | G   |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 22 | T   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | T   |     |
| 23 | T   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 24 | T   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | T   |     |     |
| 25 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | T   |     |
| 26 |     |     |     |     |     |     |     | T   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 27 |     |     |     |     |     |     |     |     | T   |     |     |     |     |     |     |     |     |     |     | A   |     |     |     |     |     |     |     |     |
| 28 |     |     |     |     |     |     |     |     | A   | T   | T   |     |     |     |     |     |     |     | A   | A   | T   |     |     |     |     |     |     |     |
| 29 |     |     |     |     |     |     |     | T   | A   | T   | T   | C   |     |     |     |     |     | A   | A   | T   |     |     |     |     |     |     |     |     |

Table 2: Mutations of P<sub>m</sub> and the rationale behind each of them.

| <b>Mutation numbers</b> | <b>Mutation</b>           | <b>Rationale behind the mutation</b>                                                                                                                                                             |
|-------------------------|---------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Mutation 1              | -46 A to C                | To make symmetrical* substitution corresponding to -41G.                                                                                                                                         |
| Mutation 2              | -41 G to T                | To make symmetrical substitution corresponding to -46A.                                                                                                                                          |
| Mutation 3              | -48 A to C and -39 T to G | To maintain symmetry, but with a GC combination at these positions. To test if specific bases are important or is it just a factor of symmetry.                                                  |
| Mutation 4              | -37 A to C                | To make symmetrical substitution corresponding to -50G. Previous data** recorded slightly higher <i>in vivo</i> activation as compared to wild-type.                                             |
| Mutation 5              | -52 C to G                | To make symmetrical substitution corresponding to -35C.                                                                                                                                          |
| Mutation 6              | -52 C to A and -35 C to T | To make these positions symmetrical with respect to each other with an AT combination. Previous data for -35 C to T recorded slightly higher <i>in vivo</i> activation as compared to wild-type. |
| Mutation 7              | -34 C to T                | To make symmetrical substitution corresponding to -53A. Previous data recorded slightly higher <i>in vivo</i> activation as compared to wild-type.                                               |
| Mutation 8              | -53 A to G                | To make symmetrical substitution corresponding to -34C. Previous data recorded slightly higher <i>in vivo</i> activation as compared to wild-type.                                               |
| Mutation 9              | -33 G to T                | To make symmetrical substitution corresponding to -54A. Previous data recorded slightly higher <i>in vivo</i> activation as compared to wild-type.                                               |
| Mutation 10             | -54 A to C                | To make symmetrical substitution corresponding to -33G                                                                                                                                           |
| Mutation 11             | -32 G to T                | To make symmetrical substitution corresponding to -55A. Previous data recorded slightly higher <i>in vivo</i> activation as compared to wild-type.                                               |
| Mutation 12             | -31 C to A                | To make symmetrical substitution corresponding to -56T. Previous data recorded twice the wild-type activity.                                                                                     |

Table 2 (continued).

| <b>Mutation numbers</b> | <b>Mutation</b>                       | <b>Rationale behind the mutation</b>                                                                                                                                                                                 |
|-------------------------|---------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Mutation 13             | -56 T to A and -31 C to T             | To maintain the symmetry of mutation 12 but by swapping positions of A and T. substitution at -31 gave twice the wild-type activity as compared to wild-type.                                                        |
| Mutation 14             | -57 G to T                            | Previous data revealed twice the activity as compared to wild-type sequence.                                                                                                                                         |
| Mutation 15             | -57 G to T and -30 T to A             | Previous data of substitution at -57 revealed twice the activity as compared to wild-type sequence. To make symmetrical substitution corresponding to -57 when changed to T.                                         |
| Mutation 16             | -50 G to T and -41 G to T             | To make symmetrical substitutions at both positions corresponding to -37A and -46A respectively such that the 6 base-pair dyad symmetry is perfect and AT-rich.                                                      |
| Mutation 17             | -50 G to T and -46 A to C             | To make symmetrical substitutions at both positions corresponding to -37A and -41G respectively such that the 6 base-pair dyad symmetry is perfect and AT-rich except at positions -41 and -46, where it is GC-rich. |
| Mutation 18             | -42 through -45 AAAA                  | Making these positions A-rich to test if this mutation makes any change in the minor groove dimensions that could affect Mor binding or is it just the specific bases that are important in the minor groove spacer. |
| Mutation 19             | -42 through -45 TAAT                  | Making these positions AT-rich, but maintaining the wild-type symmetry. To test if this mutation makes any change in the minor groove dimensions that could affect Mor binding.                                      |
| Mutation 20             | -42 through -45 GGCC                  | Making the spacer region symmetrical centered at -43.5.                                                                                                                                                              |
| Mutation 21             | -42 through -45 GGGG                  | Making it G-rich to test if specific bases are important in the minor groove or is it just the minor groove dimensions.                                                                                              |
| Mutation 22             | -57 G to T, -53 A to G and -31 C to T | Previous data with the specific substitutions at the 3 positions gave twice the wild-type activity. To test if these substitutions together are cumulative in their effect or otherwise.                             |
| Mutation 23             | -57 G to T and -53 A to G             | A subset of mutation 22 with the same rationale.                                                                                                                                                                     |

Table 2 (continued).

| <b>Mutation numbers</b> | <b>Mutation</b>                        | <b>Rationale behind the mutation</b>                                                                                                                                                                                           |
|-------------------------|----------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Mutation 24             | -57 G to T and -31 C to T              | A subset of mutation 22 with the same rationale.                                                                                                                                                                               |
| Mutation 25             | -53 A to G and -31 C to T              | A subset of mutation 22 with the same rationale.                                                                                                                                                                               |
| Mutation 26             | -50 G to T                             | To test the individual contribution by -50T from mutation 17, as mutation 17 gave twice the wild-type binding by His-Mor in a gel shift <i>in vitro</i> binding assay.                                                         |
| Mutation 27             | -48 A to T and -39 T to A              | To maintain symmetry at these positions with respect to each other but by swapping the bases from wild-type sequence to test if binding by wild-type His-Mor is affected by symmetry or the specific bases in these positions. |
| Mutation 28             | -49 to -47 ATT and<br>-40 to -38 AAT   | To maintain symmetry at these positions with respect to each other but by swapping the bases from wild-type sequence to test if binding by wild-type His-Mor is affected by symmetry or the specific bases in these positions. |
| Mutation 29             | -50 to -46 TATTC and<br>-40 to -38 AAT | To test effect of mutation 28 in combination with mutation 17 (gave twice the wild-type promoter binding by His-Mor in the gel shift assay).                                                                                   |

\* Symmetry centered at -43.5.

\*\* Previous data - Artsimovitch, 1996.

## Materials and methods

### Media, chemicals and enzymes

Protein over-expression and routine cell growth were done in LB medium (Howe, 1973), whereas cultures for  $\beta$ -galactosidase assays were grown in minimal medium with casamino acids (M9CA) (Margolin et al, 1989). Bacto agar, Bacto tryptone and Bacto yeast extract used for LB medium were purchased from Difco Laboratories and sodium chloride from Fisher Scientific. MacConkey lactose plates with 25 g/L of MacConkey agar (Difco) and 25 g/L of MacConkey agar base (Difco) was used for plate phenotyping. Chloramphenicol and ampicillin were purchased from Sigma. Isopropyl- $\beta$ -D-thiogalacto-pyranoside (IPTG) was from US Biologicals and o-nitrophenyl- $\beta$ -galactopyranoside (ONPG) was from American Bioorganics. Acrylamide, bisacrylamide, 40% acrylamide/bis-acrylamide (29:1) solution, tetramethylethylenediamine (TEMED) and ammonium persulfate (APS) were purchased from BioRad. Chloroform and glycerol were obtained from Fisher Scientific. Imidazole was purchased from Sigma and sodium dodecyl sulfate (SDS) was from BioRad. Trizma base was from Sigma. Glacial acetic acid, ethylene diamine tetraacetic acid (EDTA) and boric acid were purchased from Fisher Scientific. D-Glucose, M9 salts, thiamine, biotin, potassium chloride, dithiothreitol, IGEAL CA-630 and Ficoll-400 were all purchased from Sigma. Glycine and  $\beta$ -mercaptoethanol were from BioRad. Casamino acids was from Difco and sodium carbonate was from JT Baker. Magnesium sulfate, calcium chloride, monosodium and disodium phosphates were from Sigma. Absolute ethanol and 95% ethanol were from AAPER Alcohol and Chemical Co., Shelbyville, KY. Seakem ME and LE grade agarose were from Lonza, Rockland Inc. The Bradford protein assay reagent and Bio-safe Coomassie stain for protein gels were from BioRad. Dimethyl sulfoxide (DMSO) for freezing the bacterial strains was from Sigma. The restriction enzymes *Bam*H1 and *Eco*R1 were from New England Biolabs. *Taq* polymerase was from Roche Diagnostics and T4 DNA ligase was from New England Biolabs. T4 polynucleotide kinase (PNK) and dNTPs were obtained from Promega Corporation. Shrimp alkaline phosphatase (SAP) was purchased from Roche Diagnostics. [ $\gamma$  P<sup>32</sup>]-ATP was obtained from Perkin Elmer Life Science, Boston, MA. Talon spin columns used for protein purification were from Clontech. Miniprep and midiprep kits for small and large scale plasmid extraction respectively were purchased from Promega Corporation. Kit for purification of PCR products was obtained from Qiagen. Slide-A-Lyzer dialysis cassette (10-kDa cutoff) for protein purification was from Pierce. Columns for purification of oligo probes used in gel shifts, G25 (for oligonucleotides from 10-20 bases in length) and G50 (for oligonucleotides 20-50 bases in length) were from GE Healthcare. BioMax MR films for exposing the gel shift gels were from Kodak.

### Bacterial strains and plasmids

The host strain background for most of the plasmid constructions and *in vivo* assays was *Escherichia coli* K-12 strain MH13312 (*mcrA*  $\Delta$ *pro-lac* *thi* *gyrA* *endA* *hsdR*

*relR supE44 recA / F' pro<sup>+</sup> lacI<sup>Q1</sup> ΔlacZY*), a derivative of JM109 carrying an F' plasmid deleted for both *lacZ* and *lacY* and expressing higher than normal levels of Lac repressor (Artsimovitch and Howe, 1996). It has no antibiotic resistance. Strain MH13435 was made by transforming plasmid pIA14, that has wild-type P<sub>m</sub> sequence from -61 to +10 fused to *lacZ*, into MH13312. Plasmid pIA14 has the ampicillin resistance gene. Strain MH13418 was made by transforming plasmid pIA69, which has the *his-mor* gene under the control of IPTG-inducible P<sub>lacUV5</sub> promoter, into MH13312. This was used as a host for transformation for phenotypic and quantitative assays of the ability of Mor to activate transcription from the various mutations introduced into P<sub>m</sub>. Plasmid pIA69 has the chloramphenicol resistance gene. Strain MH13315 with plasmid pIA12, which does not have a wild-type promoter, P<sub>m</sub> was used as a vector for all clonings to facilitate selection of clones. Plasmid pIA12 has the ampicillin resistance gene. Strain MH13355 is the protein expression strain, containing λDE3 that encodes T7 RNA polymerase used for protein over-expression and detection of wild-type protein His-Mor.

### Oligonucleotide synthesis

Oligonucleotides were synthesized by Integrated DNA Technologies Inc. on a commercial nucleic acid synthesizer, Model ABI394, using the phosphoramidite chemistry method (Caruthers et al, 1983). For Mod-PCR mutagenesis (Chiang and Howe 1993) degeneracy was introduced into the targeted position by simultaneous delivery of equal volumes of solution of wild-type nucleotide and an equimolar mix of all four nucleotides resulting in a mis-incorporation rate of 0.25/nucleotide. To get a higher frequency of specific nucleotide, especially T, degenerate oligo MUT 52 was ordered with a higher frequency of T in all four positions. For site specific mutagenesis, nucleotides indicated in the table were delivered at the specific positions. The sequence of oligonucleotides used for mutagenesis, sequencing and probe preparation in this chapter are shown in Table 3. All the oligonucleotides were diluted to a working concentration of 100 picomoles/ul.

### Targeted mutagenesis

Mutagenesis of the spacer region was specifically targeted at the four bases of the spacer from -42 through -45 positions of P<sub>m</sub> with respect to the transcription start site of +1 and the effect of these mutations were studied for the contribution of the minor groove bases in interaction with His-Mor. Three degenerate primers were designed to introduce all possible substitutions in the spacer:

1. MUT 50: This top strand oligonucleotide was designed with NCCN in the spacer region with degeneracy at the first and the fourth positions.
2. MUT 51: This top strand oligonucleotide was designed with GNNG in the spacer region with middle two positions of the spacer degenerate.
3. MUT 52: This top strand oligonucleotide was designed with NNNN in the spacer region with degeneracy at all four positions.



Table 3: Oligonucleotides used for promoter mutagenesis.

| Primer  | Sequence                                                                                  | Comments                                                                                   |
|---------|-------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| KRI 125 | GGCGAATTCTTCTGTAAACAGTAA <b>C</b><br>GCCGGTTAATC                                          | Top strand mutagenic primer designed to incorporate mutation <b>1</b> in P <sub>m</sub> .  |
| KRI 126 | GGCGAATTCTTCTGTAAAGAGTAA <b>A</b><br>GCCGGTTAATC                                          | Top strand mutagenic primer designed to incorporate mutation <b>5</b> in P <sub>m</sub> .  |
| KRI 127 | GGCGAATTCTTCTGTAA <b>G</b> CAGTAA <b>A</b><br>GCCGGTTAATC                                 | Top strand mutagenic primer designed to incorporate mutation <b>8</b> in P <sub>m</sub> .  |
| KRI 128 | GGCGAATTCTTCTGTAC <b>A</b> CAGTAA <b>A</b><br>GCCGGTTAATC                                 | Top strand mutagenic primer designed to incorporate mutation <b>10</b> in P <sub>m</sub> . |
| KRI 129 | GGCGAATTCTTCT <b>T</b> TAAACAGTAA <b>A</b><br>GCCGGTTAATC                                 | Top strand mutagenic primer designed to incorporate mutation <b>14</b> in P <sub>m</sub> . |
| KRI 130 | GGCGAATTCTTCTGTAAAC <b>A</b> TAA <b>C</b><br>GCCGGTTAATC                                  | Top strand mutagenic primer designed to incorporate mutation <b>17</b> in P <sub>m</sub> . |
| KRI 131 | GGCGAATTCTTCTGTAAACAGTAA <b>A</b><br><b>AAA</b> AGTTAATC                                  | Top strand mutagenic primer designed to incorporate mutation <b>18</b> in P <sub>m</sub> . |
| KRI 132 | GGCGAATTCTTCTGTAAACAGTAA <b>A</b><br><b>TAAT</b> GTTAATC                                  | Top strand mutagenic primer designed to incorporate mutation <b>19</b> in P <sub>m</sub> . |
| KRI 133 | GGCGAATTCTTCTGTAAACAGTAA <b>A</b><br><b>CCG</b> GGTTAATC                                  | Top strand mutagenic primer designed to incorporate mutation <b>20</b> in P <sub>m</sub> . |
| KRI 134 | GGCGAATTCTTCTGTAAACAGTAA <b>A</b><br><b>GGG</b> GGTTAATC                                  | Top strand mutagenic primer designed to incorporate mutation <b>21</b> in P <sub>m</sub> . |
| KRI 135 | GGCGAATTCTTCT <b>T</b> TAA <b>G</b> CAGTAA <b>A</b><br>GCCGGTTAATC                        | Top strand mutagenic primer designed to incorporate mutation <b>23</b> in P <sub>m</sub> . |
| KRI 136 | GGCGAATTCTTCTGTAAAC <b>A</b> TAA <b>A</b><br>GCCGGTTAATC                                  | Top strand mutagenic primer designed to incorporate mutation <b>26</b> in P <sub>m</sub> . |
| KRI 137 | GGCGAATTCTTCTGTAAACAGTAA <b>A</b><br>GCCGGTTAATC <b>C</b> TGCTTTTTT <b>T</b> ACG          | Top strand mutagenic primer designed to incorporate mutation <b>9</b> in P <sub>m</sub> .  |
| KRI 138 | GGCGAATTCTTCTGTAAACAGTAA <b>A</b><br>GCCGGTTAATC <b>C</b> G <b>T</b> CTTTTTT <b>T</b> ACG | Top strand mutagenic primer designed to incorporate mutation <b>11</b> in P <sub>m</sub> . |

Table 3 (continued).

| Primer  | Sequence                                                        | Comments                                                                                   |
|---------|-----------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| KRI 139 | GGCGAATTCTTCTGTAAACAGTAAA<br>GCCGGTTAATCCGGATT TTT TTT TACG     | Top strand mutagenic primer designed to incorporate mutation <b>12</b> in P <sub>m</sub> . |
| KRI 140 | GGCGAATTCTTCTGAAAACAGTAAA<br>GCCGGTTAATCCGGT TTT TTT TTT TACG   | Top strand mutagenic primer designed to incorporate mutation <b>13</b> in P <sub>m</sub> . |
| KRI 141 | GGCGAATTCTTCTTTAAACAGTAAA<br>GCCGGTTAATCCGGCAT TTT TTT TTT TACG | Top strand mutagenic primer designed to incorporate mutation <b>15</b> in P <sub>m</sub> . |
| KRI 142 | GGCGAATTCTTCTGTAAACAGTAAA<br>GCCGTTTAATCCGGCT TTT TTT TTT TACG  | Top strand mutagenic primer designed to incorporate mutation <b>2</b> in P <sub>m</sub> .  |
| KRI 143 | GGCGAATTCTTCTGTAAACAGTCAA<br>GCCGGTGAATCCGGCT TTT TTT TTT TACG  | Top strand mutagenic primer designed to incorporate mutation <b>3</b> in P <sub>m</sub> .  |
| KRI 144 | GGCGAATTCTTCTGTAAACAGTAAA<br>GCCGGTTACTCCGGCT TTT TTT TTT TACG  | Top strand mutagenic primer designed to incorporate mutation <b>4</b> in P <sub>m</sub> .  |
| KRI 145 | GGCGAATTCTTCTGTAAA AAGTAAA<br>GCCGGTTAATTCGGCT TTT TTT TTT TACG | Top strand mutagenic primer designed to incorporate mutation <b>6</b> in P <sub>m</sub> .  |
| KRI 146 | GGCGAATTCTTCTGTAAACAGTAAA<br>GCCGGTTAATCTGGCT TTT TTT TTT TACG  | Top strand mutagenic primer designed to incorporate mutation <b>7</b> in P <sub>m</sub> .  |
| KRI 147 | GGCGAATTCTTCTGTAAACAT TAAA<br>GCCGTTTAATCCGGCT TTT TTT TTT TACG | Top strand mutagenic primer designed to incorporate mutation <b>16</b> in P <sub>m</sub> . |
| KRI 148 | GGCGAATTCTTCTTTAAGCAGTAAA<br>GCCGGTTAATCCGGT TTT TTT TTT TACG   | Top strand mutagenic primer designed to incorporate mutation <b>22</b> in P <sub>m</sub> . |
| KRI 149 | GGCGAATTCTTCTTTAAACAGTAAA<br>GCCGGTTAATCCGGT TTT TTT TTT TACG   | Top strand mutagenic primer designed to incorporate mutation <b>24</b> in P <sub>m</sub> . |
| KRI 150 | GGCGAATTCTTCTGTAAAGCAGTAAA<br>GCCGGTTAATCCGGT TTT TTT TTT TACG  | Top strand mutagenic primer designed to incorporate mutation <b>25</b> in P <sub>m</sub> . |
| KRI 151 | GGCGAATTCTTCTGTAAACAGTTAA<br>GCCGGTAAATCCGGCT TTT TTT TTT TACG  | Top strand mutagenic primer designed to incorporate mutation <b>27</b> in P <sub>m</sub> . |
| KRI 152 | GGCGAATTCTTCTGTAAACAGATTA<br>GCCGGAATATCCGGCT TTT TTT TTT TACG  | Top strand mutagenic primer designed to incorporate mutation <b>28</b> in P <sub>m</sub> . |

Table 3 (continued).

| Primer  | Sequence                                                                 | Comments                                                                                    |
|---------|--------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
| KRI 153 | GGCGAATTCTTCTGTAAACAT <b>TATTCG</b><br>CCGG <b>AAT</b> ATCCGGCTTTTTTTACG | Top strand mutagenic primer designed to incorporate mutation <b>29</b> in P <sub>m</sub> .  |
| MUT 50  | GGCGAATTCTTCTGTAAACAGTAAA<br>NCCNGT <b>T</b> AATC                        | Top strand degenerate primer to mutagenize the last two positions of the spacer.            |
| MUT 51  | GGCGAATTCTTCTGTAAACAGTAAA<br>G <b>NN</b> GGT <b>T</b> AATC               | Top strand degenerate primer to mutagenize the middle two positions of the spacer.          |
| MUT 52  | GGCGAATTCTTCTGTAAACAGTAAA<br>N <b>NNN</b> GT <b>T</b> AATC               | Top strand degenerate primer to mutagenize all four positions of the spacer.                |
| MUT 528 | GGCGAATTCTTCTGTAAACAGTAAA<br>GGACG <b>T</b> T <b>AATC</b>                | Top strand mutagenic primer with GGAC in the spacer region.                                 |
| MUT 529 | GGCGAATTCTTCTGTAAACAGTAAA<br>GACG <b>G</b> T <b>T</b> AATC               | Top strand mutagenic primer with GAGC in the spacer region.                                 |
| MUT 530 | GGCGAATTCTTCTGTAAACAGTAAA<br>A <b>CCC</b> G <b>T</b> T <b>AATC</b>       | Top strand mutagenic primer with ACCC in the spacer region.                                 |
| MLK 7   | CCTGGATCCGTACGGTTATTCATCAC<br>AG                                         | Bottom strand wild-type primer to be used with all the above primers for PCR amplification. |

The oligonucleotide sequences are written from the 5' to 3' direction. The letters in red indicate the mutations introduced. The numbers in bold red are the mutations as they were numbered in this study.

The rationale behind the various degenerate oligonucleotides was that if Q68 and Y70 contacted the bases of the minor groove spacer:

1. If it made contacts with the G's, then we would get more frequency of mutant phenotypes with MUT 50 and MUT 52.
2. If it made contacts with the C's then we would get more frequency of mutant phenotypes with MUT 51 and MUT 52.
3. If all four positions were equally important for the contacts of the side chains of the amino acids then we would get more or less equal distribution of mutant phenotypes with all the three degenerate oligonucleotides.

A single step PCR mutagenesis was carried out with degenerate top strand primers MUT 50, MUT 51 and MUT 52 and the bottom strand wild-type primer MLK 7 in separate reactions. This introduced a library of substitutions in the spacer region corresponding to the various bases in the degenerate oligonucleotides. For introducing all other mutations in  $P_m$ , primers with the specific mutations were designed as given in the Table 3 and a single step PCR mutagenesis was carried out with these top strand primers and the bottom strand wild-type primer MLK 7 in separate reactions. Thus, the PCR products had incorporated the specific mutations that were designed for mutagenesis of the various positions of  $P_m$ . The program used for amplification was as follows: 1 min at 94°C, 1 min at 50°C and 1 min at 72°C for 30 cycles followed by 5 min at 72°C in a PTC-100 thermal cycler (MJ Research Inc., Watertown, MA). All the mutagenic oligonucleotides had the *EcoR*I site at its 5' end preceded by 3 extra nucleotides and the primer MLK 7 had the *Bam*H1 site followed by additional nucleotides to facilitate restriction digest of the insert PCR products.

The rest of the experimental protocol was similar for both the spacer region mutagenesis and all other  $P_m$  mutations. PCR products were cleaned up using Qiaquick PCR purification kit and then subjected to double restriction digest with *EcoR*I and *Bam*H1 at 37°C for 4 hours. The vector pIA14 that was used for spacer region mutagenesis and vector pIA12 that was used for all other  $P_m$  mutagenesis were also subjected to similar restriction digests and also phosphatased with shrimp alkaline phosphatase to prevent re-ligation of vectors cut by just one enzyme. Ligation reactions with the digested PCR products as inserts and digested vector were incubated at 16°C overnight for about 18-20 hours. The ligation mixture was transformed into MH13418 competent cells, which has His-Mor containing pIA69 plasmid.

### **Determination of plate phenotypes**

Transformants were plated on MacConkey lactose agar plates with chloramphenicol (25 µg/ml), ampicillin (40 µg/ml) and IPTG (50 µM). The two plasmid system was used throughout this study. The promoter plasmid, pIA14 has ampicillin resistance and the protein plasmid, pIA69 has chloramphenicol resistance. The bacterial cells which have both the plasmids will grow on plates with ampicillin and chloramphenicol. The induction of His-Mor expression is through the action of IPTG

which induces the  $P_{lacUV5}$  promoter in pIA69 plasmid. The His-Mor thus expressed, binds to  $P_m$  on the pIA12 derivative, pIA14 which had the wild-type  $P_m$  or the cloned  $P_m$  with mutations. Gene *lacZ* is immediately downstream of the  $P_m$  promoter and is under its control for activity. Figure 14 illustrates the two plasmid system.

The level of activation of  $P_m$  depends on the interaction of His-Mor with the promoter and this determines the level of expression of the *lacZ* gene. Normal expression of *lacZ* is indicated by a red color on MacConkey plates and any defect in its expression is observed by a different colony phenotype. Red colonies with a white halo indicate lower than wild-type expression of *lacZ*. No expression of *lacZ* is observed as white colonies on the plate. Red centered phenotype indicates an intermediate expression. The expression levels of *lacZ* from the plate phenotyping and *in vivo*  $\beta$ -galactosidase activity can thus be used to study interactions between His-Mor, promoter DNA and RNAP.

The phenotypic frequency of the mutant libraries was calculated based on the distribution of different phenotypes among the candidates. Plate phenotyping was done by observing the color of the colonies on the MacConkey lactose agar plates after incubating the plates at 37°C for 16 hours. The fully functional mutations of  $P_m$  gave red colonies similar to wild-type and the defective ones gave red centered or white phenotypes depending on partially defective or fully defective phenotypes. The colors were scored relative to those of positive control (wild-type pIA14 plasmid in transformation) and the negative control (transformation of pIA12 which lacks  $P_m$ ). A few candidates from each phenotype were selected and purified and their promoter plasmids, pIA14 were sequenced to identify the mutations.

### **Preparation of competent cells**

Competent cells were prepared by  $CaCl_2$  method (Mandel and Higa, 1970):

1. MH13418, which has the protein His-Mor plasmid pIA69, was streaked on LB agar plate with chloramphenicol (25  $\mu$ g/ml) and was incubated overnight at 37°C.
2. Single colony was inoculated into 5 ml plain LB and grown overnight at 37°C shaker.
3. Next morning, the seed culture was inoculated into 100 ml plain LB w/o antibiotics and grown in 37°C shaker till the OD reached 0.4-0.6 (for about 3 hrs).
4. Flasks were kept on ice for 15 min.
5. 25 ml of the culture was taken in each of the 30 ml Corex tubes and centrifuged at 10000 rpm for 15 min (SS34 rotor) and the supernatant was discarded.
6. 12.5 ml (50% of culture volume) sterile 0.1 M  $CaCl_2$  was taken in each tube and left for 15 min. on ice and then resuspended.
7. It was centrifuged at 10000 rpm for 15 min (SS34 rotor) and the supernatant was discarded. The pellet was resuspended in 2.5 ml  $CaCl_2$  and 375  $\mu$ l glycerol.
8. Resuspended cells were left in cold room for 2 hrs to overnight, distributed into 200  $\mu$ l aliquots, flash-frozen and stored at -80°C freezer.

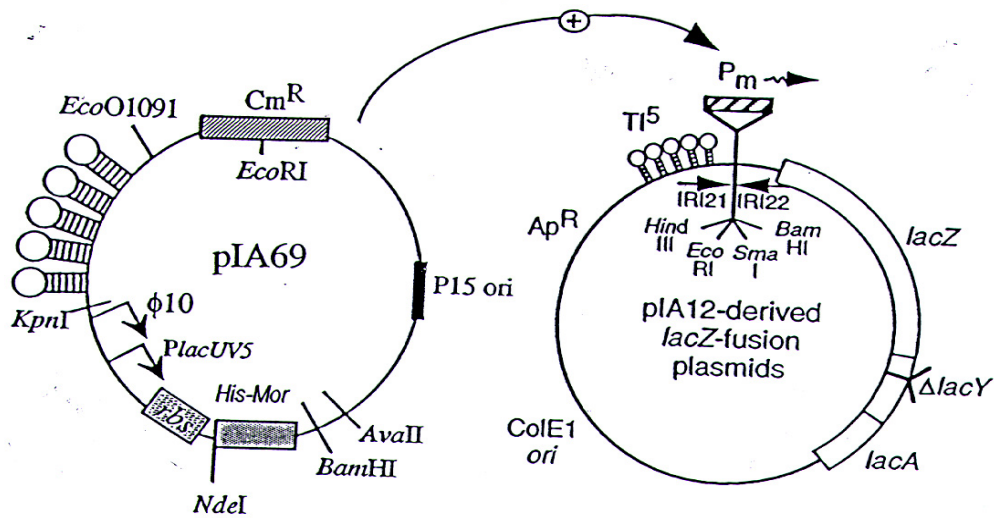


Figure 14: The two plasmid system used to study interactions of  $P_m$  with His-Mor.

Modified with permission from Nucleic Acids Research. Artsimovitch I, Howe M. Transcription activation by the bacteriophage Mu Mor protein: analysis of promoter mutations in  $P_m$  identifies a new region required for promoter function. Nucl. Acids Res. 1996, 24(3): 450-7.

## **Transformation protocol**

The MH13418 competent cells were thawed on ice for 30 min. The ligation mix was then added into the competent cells tube and left standing for 30 min on ice. Heat shock treatment was given for exactly 2 min at 42°C and then the tubes were placed immediately on ice for 10 min. The entire volume of cells was grown on 2 ml LB broth for 2 hrs on 37°C shaker. 150 µl of the outgrowth transformants was plated on MacConkey lactose agar plates with chloramphenicol (25 µg/ml), ampicillin (40 µg/ml) and IPTG (50 µM) for plate phenotyping and selection of the clones.

## **Plasmid purification protocol**

Wizard Plus Midiprep DNA Purification kit from Promega Corporation was used for the extraction of the plasmids pIA14 from MH13435 and pIA12 from MH13315.

1. The strain was streaked on LB agar plates with ampicillin (40 µg/ml) and left overnight in 37°C incubator.
2. A single colony was inoculated into 100 ml LB broth with ampicillin (40 µg/ml) and grown overnight in 37°C shaker.
3. Cells were pelleted at 8000 rpm for 18 min at 4°C.
4. The pellet was resuspended in 3 ml cell resuspension solution provided in the kit.
5. The cells were lysed in 3ml cell lysis solution provided in the kit.
6. The reaction was neutralized using 3 ml neutralization Solution from the kit.
7. It was then centrifuged at 12000 rpm for 15 min at 4°C.
8. The supernatant was mixed with the resuspended resin and allowed to pass through the midiprep column from the kit in a vacuum manifold.
9. After all the liquid passed through the column, it was washed twice with 15 ml column wash solution.
10. The plasmid from the column was extracted with 300 µl preheated milliQ water by centrifugation for 20 sec at 10000 rpm.

The plasmid preparation thus obtained was further purified with ethaonol precipitation for further purity as follows:

1. 30 µl sodium acetate and 750 µl absolute ethanol was added to the 300 µl plasmid preparation and kept standing at -80°C for about 40 min.
2. The DNA was centrifuged at 10000 rpm for 10 min and the supernatant was discarded.
3. The DNA pellet was washed twice in 1ml of 70% ethanol.
4. The pellet was then dried in vacuum for 10 min and then resuspended in 150 µl milliQ water at pH 8.0.

The purified DNA was observed on a 1% agarose gel and quantitated on the Thermo Scientific NanoDrop™ 1000 Spectrophotometer.

## Sequencing of the plasmids

All the plasmids were sequenced using MUT 49 as the bottom strand sequencing primer at the Molecular Resource Center (MRC) facility at the University of Tennessee Health Science Center, Memphis. At the MRC, automated DNA sequencing was accomplished with ABI Model 3130XL Genetic Analyzers that use four-color fluorescence-based sequencing based on the Sanger method. Plasmid or PCR produced DNAs were used as suitable substrates for the sequencing protocol. Samples were submitted as template/primer mixes, with the ratio of DNA/primer dependent on the type of template employed. Following the incorporation of the labeled ddNTPs in the extension products, the reactions were purified by gel filtration, dried down, re-suspended in formamide, and run on the analyzer. DNA sequencing reactions were purified with the BigDye<sup>®</sup> XTerminator<sup>™</sup> Purification Kit result in high signal strength when analyzed on a DNA sequencer. Data were uploaded on the MRC server as either ab.1 files that could be opened with ABI software or programs such as FinchTv or text files. Electropherograms, which are the graphic colored printouts of the sequencing, were used to read the sequences. A BLAST was performed for each of the sequences with “bl2seq” from the NCBI site, using the sequence from wild-type promoter and the two sequences were aligned and the base-pair mismatches were observed to check the incorporation of the specific mutations designed.

## Small scale purification of His-Mor

Overnight cultures (2 ml) of MH13355 protein expression strain with freshly transformed His-Mor plasmid were transferred to 100 ml of LB supplemented with 25 µg/ml of chloramphenicol. The cells were grown at 37°C until the  $A_{600}$  reached 0.4 - 0.6, and then His-Mor over-expression was induced with 1 mM IPTG for 3 hrs. After harvesting by centrifugation at 8000 rpm for 15 min at 4°C, the cell pellets were resuspended in 4 ml of Resuspension Buffer M containing 20 mM Tris-HCl pH 7.9, 200 mM NaCl, 0.5 mM phenylmethylsulfonylfluoride, 10% glycerol, 1 mM 2-mercaptoethanol and lysed by sonication (output 40 cycles 4 for 4 minutes). After centrifugation, the supernatants were loaded into TALON spin columns (now equilibrated thrice with Buffer M) and washed sequentially once with 1 ml of BufferM containing 0.1% IGEPAL CA-630 detergent, once with 1 ml of Buffer M containing 500 mM NaCl and once with 1 ml of Buffer M containing 40 mM imidazole. Proteins were eluted four times, 0.25 ml/time, with Buffer M containing 100 mM imidazole. The final two elutions were pooled together. The second elution and the pooled fractions were dialysed overnight at 4°C in a 10-kDa cutoff Slide-A-lyzer dialysis cassette in the dialysis buffer solution (20 mM Tris-HCl pH 7.9, 50 mM NaCl, 10% glycerol, 1 mM EDTA, and 1 mM DTT). It was then supplemented with glycerol to a final concentration of 15%, aliquoted and stored at -80°C. Purity of the preparations was assayed on a 12.5% SDS polyacrylamide gel stained with silver nitrate, and the protein concentration was determined by a Bradford assay with IgG as the standard (Bradford, 1976).



### ***In vivo* transactivation assay**

Cells were grown overnight in 2 ml of M9CA medium [M9 media (20% 5X M9 salts, 20 mM glucose, 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, 10 µg/ml each of thiamine and Biotin) + 0.2% casamino acid] supplemented with 25-30 µg/ml chloramphenicol and 40 µg/ml ampicillin. The overnight culture was diluted 1:50 into 10 ml of fresh M9CA medium with antibiotics and grown at 37°C until the A<sub>600</sub> reached 0.4 - 0.6. A 1-ml sample was removed to serve as an uninduced control, and the remaining culture was induced with 2 mM IPTG for 60 min. Based on the plate phenotype of individual mutants, dilutions of the cells were made using M9CA medium, and the cells were permeabilized by mixing with 10 µl of chloroform and 25 µl of 0.1% SDS. After incubation for 20 min on ice, 0.5 ml of ONPG (4 mg/ml to a final concentration of 0.8 µg/µl) in Buffer Z (60 mM Na<sub>2</sub>HPO<sub>4</sub>, 40 mM NaH<sub>2</sub>PO<sub>4</sub>, 10 mM KCl, 1 mM MgSO<sub>4</sub> and 50 mM 2-mercaptoethanol) was added, and the mixture was incubated at 28°C for 20 min. The reactions were stopped by adding 250 µl of 1 M Na<sub>2</sub>CO<sub>3</sub> and spectrophotometer readings were taken at 420 nm for the enzymatic reaction and 600 nm for cell density. The β-galactosidase activities were calculated according to Miller's formula (Miller, 1972) and normalized relative to that of a wild-type culture assayed in parallel and set to 1000.

### **Gel mobility shift assay**

The list of oligonucleotides used for gel shifts is given in Table 4. All the bottom strand oligonucleotides were radioactively labeled. Labeled probes were prepared using γ-P<sup>32</sup> labeled bottom strand primer in a 10 µl labeling reaction with 1 µl T4 polynucleotide kinase, 4 µl milliQ water, 1 µl kinase buffer and 3 µl γ-P<sup>32</sup> ATP (500 µCurie stock from Perkin Elmer) and incubated at 37°C for 60 min and then at 95°C for 2 min to inactivate the enzyme. Then, 1.1 µl of 0.5 M NaCl (final concentration of 50 mM), and 5 µl of the top strand oligo was added to the reaction mix, kept on a 100°C heat block and the block was disconnected for the oligonucleotides to anneal gradually as the temperature drops from 100°C to room temperature. After it drops to room temperature, 35 µl binding buffer (20 mM Tris-HCl pH 7.9, 50 mM NaCl, 5% glycerol and 1 mM DTT) was added and the double stranded radioactively labeled oligonucleotides were purified using G25 (for oligonucleotides from 10-20 bases in length) and G50 (for oligonucleotides 20-50 bases in length) columns from GE Healthcare.

A 20 µl reaction volume containing probe (about 30 radioactive counts per second as measured with a Geiger counter), 50 ng of calf-thymus DNA, and different concentrations of His-Mor protein in binding buffer (20 mM Tris-HCl pH 7.9, 50 mM NaCl, 5% glycerol and 1 mM DTT) was incubated at room temperature for 20 minutes. Two different amounts (400 ng and 800 ng) of wild-type His-Mor were used with each of the labeled probes in most cases while in some cases three amounts (200 ng, 400 ng and 800 ng) of protein was used. There was one tube with a negative control without His-Mor and with the probe alone. One positive control was used for every gel which had the wild-type promoter P<sub>m</sub> labeled probe with wild-type His-Mor.

Table 4: List of oligonucleotides for gel shifts.

| <b>Oligo numbers</b> | <b>Sequence</b>                              | <b>Comments</b>                                                            |
|----------------------|----------------------------------------------|----------------------------------------------------------------------------|
| KRI 57               | CAGTAAAGCCGGTTAATC                           | Top strand 18-mer oligo probe.                                             |
| KRI 58               | GATTAACCGGCTTTACTG                           | Bottom strand 18-mer oligo probe.                                          |
| KRI 59               | ACAGTAAAGCCGGTTAATCC                         | Top strand 20-mer oligo probe.                                             |
| KRI 60               | GGATTAACCGGCTTTACTGT                         | Bottom strand 20-mer oligo probe.                                          |
| MUT 18               | AACAGTAAAGCCGGTTAATCCG                       | Top strand 22-mer oligo probe.                                             |
| MUT 19               | CGGATTAACCGGCTTTACTGTT                       | Bottom strand 22-mer oligo probe                                           |
| KRI 49               | AAACAGTAAAGCCGGTTAATCCGG                     | Top strand 24-mer oligo probe.                                             |
| KRI 50               | CCGGATTAACCGGCTTTACTGTTT                     | Bottom strand 24-mer oligo probe.                                          |
| KRI 51               | TAAACAGTAAAGCCGGTTAATCCGGC                   | Top strand 26-mer oligo probe.                                             |
| KRI 52               | GCCGGATTAACCGGCTTTACTGTTTA                   | Bottom strand 26-mer oligo probe.                                          |
| KRI 53               | GTAAACAGTAAAGCCGGTTAATCCGGCT                 | Top strand 28-mer oligo probe.                                             |
| KRI 54               | AGCCGGATTAACCGGCTTTACTGTTTAC                 | Bottom strand 28-mer oligo probe.                                          |
| KRI 55               | TGTAAACAGTAAAGCCGGTTAATCCGGCTT               | Top strand 30-mer oligo probe.                                             |
| KRI 56               | AAGCCGGATTAACCGGCTTTACTGTTTACA               | Bottom strand 30-mer oligo probe.                                          |
| KRI 61               | TGTAAACAGTAAAGCCGGTTAATCCGGCTT               | Top strand 30-mer oligo probe with mutation 1 in P <sub>m</sub> .          |
| KRI 62               | AAGCCGGATTAACCGGC <sup>G</sup> TTACTGTTTACA  | Bottom strand 30-mer oligo probe with mutation 1 in P <sub>m</sub> .       |
| KRI 63               | TGTAAACAGTAAAGCCG <sup>T</sup> TTAATCCGGCTT  | Top strand 30-mer oligo probe with mutation 2 in P <sub>m</sub> .          |
| KRI 64               | AAGCCGGATTAAC <sup>A</sup> CGGCTTTACTGTTTACA | Bottom strand 30-mer oligo probe containing mutation 2 in P <sub>m</sub> . |

Table 4 (continued).

| Oligo numbers | Sequence                        | Comments                                                             |
|---------------|---------------------------------|----------------------------------------------------------------------|
| KRI 65        | TGTAAACAGTCAAGCCGGTGAATCCGGCTT  | Top strand 30-mer oligo probe with mutation 3 in P <sub>m</sub> .    |
| KRI 66        | AAGCCGGATTCAACCGGCTTGACTGTTTACA | Bottom strand 30-mer oligo probe with mutation 3 in P <sub>m</sub> . |
| KRI 67        | TGTAAACAGTAAAGCCGGTTACTCCGGCTT  | Top strand 30-mer oligo probe with mutation 4 in P <sub>m</sub> .    |
| KRI 68        | AAGCCGGAGTAACCGGCTTTACTGTTTACA  | Bottom strand 30-mer oligo probe with mutation 4 in P <sub>m</sub> . |
| KRI 69        | TGTAAAGAGTAAAGCCGGTTAATCCGGCTT  | Top strand 30-mer oligo probe with mutation 5 in P <sub>m</sub> .    |
| KRI 70        | AAGCCGGATTAACCGGCTTTACTCTTTACA  | Bottom strand 30-mer oligo probe with mutation 5 in P <sub>m</sub> . |
| KRI 71        | TGTAAAAGTAAAGCCGGTTAATTCGGCTT   | Top strand 30-mer oligo probe with mutation 6 in P <sub>m</sub> .    |
| KRI 72        | AAGCCGATTAACCGGCTTTACTTTTTACA   | Bottom strand 30-mer oligo probe with mutation 6 in P <sub>m</sub> . |
| KRI 73        | TGTAAACAGTAAAGCCGGTTAATCTGGCTT  | Top strand 30-mer oligo probe with mutation 7 in P <sub>m</sub> .    |
| KRI 74        | AAGCCAGATTAACCGGCTTTACTGTTTACA  | Bottom strand 30-mer oligo probe with mutation 7 in P <sub>m</sub> . |
| KRI 75        | TGTAAGCAGTAAAGCCGGTTAATCCGGCTT  | Top strand 30-mer oligo probe with mutation 8 in P <sub>m</sub> .    |
| KRI 76        | AAGCCGGATTAACCGGCTTTACTGCTTACA  | Bottom strand 30-mer oligo probe with mutation 8 in P <sub>m</sub> . |
| KRI 77        | TGTAAACAGTAAAGCCGGTTAATCCTGCTT  | Top strand 30-mer oligo probe with mutation 9 in P <sub>m</sub> .    |

Table 4 (continued).

| Oligo numbers | Sequence                                         | Comments                                                              |
|---------------|--------------------------------------------------|-----------------------------------------------------------------------|
| KRI 78        | AAGC <b>A</b> GGATTAACCGGCTTTACTGTTTACA          | Bottom strand 30-mer oligo probe with mutation 9 in P <sub>m</sub> .  |
| KRI 79        | TGTAC <b>C</b> ACAGTAAAGCCGGTTAATCCGGCTT         | Top strand 30-mer oligo probe with mutation 10 in P <sub>m</sub> .    |
| KRI 80        | AAGCCGGATTAACCGGCTTTACTGT <b>G</b> TACA          | Bottom strand 30-mer oligo probe with mutation 10 in P <sub>m</sub> . |
| KRI 81        | TGTAAACAGTAAAGCCGGTTAATCCG <b>T</b> CCTT         | Top strand 30-mer oligo probe with mutation 11 in P <sub>m</sub> .    |
| KRI 82        | AAG <b>A</b> CGGATTAACCGGCTTTACTGTTTACA          | Bottom strand 30-mer oligo probe with mutation 11 in P <sub>m</sub> . |
| KRI 83        | TGTAAACAGTAAAGCCGGTTAATCCGG <b>A</b> TT          | Top strand 30-mer oligo probe with mutation 12 in P <sub>m</sub> .    |
| KRI 84        | AA <b>T</b> CCGGATTAACCGGCTTTACTGTTTACA          | Bottom strand 30-mer oligo probe with mutation 12 in P <sub>m</sub> . |
| KRI 85        | TG <b>A</b> AAACAGTAAAGCCGGTTAATCCGG <b>T</b> TT | Top strand 30-mer oligo probe with mutation 13 in P <sub>m</sub> .    |
| KRI 86        | AA <b>A</b> CCGGATTAACCGGCTTTACTGTTT <b>T</b> CA | Bottom strand 30-mer oligo probe with mutation 13 in P <sub>m</sub> . |
| KRI 87        | <b>T</b> TTAAACAGTAAAGCCGGTTAATCCGGCTT           | Top strand 30-mer oligo probe with mutation 14 in P <sub>m</sub> .    |
| KRI 88        | AAGCCGGATTAACCGGCTTTACTGTTT <b>AA</b>            | Bottom strand 30-mer oligo probe with mutation 14 in P <sub>m</sub> . |
| KRI 89        | <b>T</b> TTAAACAGTAAAGCCGGTTAATCCGG <b>C</b> AT  | Top strand 30-mer oligo probe with mutation 15 in P <sub>m</sub> .    |
| KRI 90        | A <b>T</b> GCCGGATTAACCGGCTTTACTGTTT <b>AA</b>   | Bottom strand 30-mer oligo probe with mutation 15 in P <sub>m</sub> . |
| KRI 91        | TGTAAAC <b>A</b> TTAAAGCCG <b>T</b> TTAATCCGGCTT | Top strand 30-mer oligo probe with mutation 16 in P <sub>m</sub> .    |

Table 4 (continued).

| Oligo numbers | Sequence                         | Comments                                                              |
|---------------|----------------------------------|-----------------------------------------------------------------------|
| KRI 92        | AAGCCGGATTAAACGGCTTTAATGTTTACA   | Bottom strand 30-mer oligo probe with mutation 16 in P <sub>m</sub> . |
| KRI 93        | TGTAACACATTAAACGCCGGTTAATCCGGCTT | Top strand 30-mer oligo probe with mutation 17 in P <sub>m</sub> .    |
| KRI 94        | AAGCCGGATTAAACCGGCCTTAATGTTTACA  | Bottom strand 30-mer oligo probe with mutation 17 in P <sub>m</sub> . |
| KRI 95        | TGTAACACAGTAAAAGGTTAATCCGGCTT    | Top strand 30-mer oligo probe with mutation 18 in P <sub>m</sub> .    |
| KRI 96        | AAGCCGGATTAACTTTTTTTACTGTTTACA   | Bottom strand 30-mer oligo probe with mutation 18 in P <sub>m</sub> . |
| KRI 97        | TGTAACACAGTAAATAATGTTAATCCGGCTT  | Top strand 30-mer oligo probe with mutation 19 in P <sub>m</sub> .    |
| KRI 98        | AAGCCGGATTAAACATTAATTTACTGTTTACA | Bottom strand 30-mer oligo probe with mutation 19 in P <sub>m</sub> . |
| KRI 99        | TGTAACACAGTAAAACGGGTTAATCCGGCTT  | Top strand 30-mer oligo probe with mutation 20 in P <sub>m</sub> .    |
| KRI 100       | AAGCCGGATTAAACCCGGTTTACTGTTTACA  | Bottom strand 30-mer oligo probe with mutation 20 in P <sub>m</sub> . |
| KRI 101       | TGTAACACAGTAAAGGGGTTAATCCGGCTT   | Top strand 30-mer oligo probe with mutation 21 in P <sub>m</sub> .    |
| KRI 102       | AAGCCGGATTAAACCCCCTTTACTGTTTACA  | Bottom strand 30-mer oligo probe with mutation 21 in P <sub>m</sub> . |
| KRI 103       | TTTAAGCAGTAAAGCCGGTTAATCCGGTTT   | Top strand 30-mer oligo probe with mutation 22 in P <sub>m</sub> .    |
| KRI 104       | AAACCGGATTAAACCGGCTTTACTGCTTAA   | Bottom strand 30-mer oligo probe with mutation 22 in P <sub>m</sub> . |
| KRI 105       | TTTAAGCAGTAAAGCCGGTTAATCCGGCTT   | Top strand 30-mer oligo probe with mutation 23 in P <sub>m</sub> .    |

Table 4 (continued).

| Oligo numbers | Sequence                        | Comments                                                              |
|---------------|---------------------------------|-----------------------------------------------------------------------|
| KRI 106       | AAGCCGGATTAACCGGCTTACTGCTTAA    | Bottom strand 30-mer oligo probe with mutation 23 in P <sub>m</sub> . |
| KRI 107       | TTAAACAGTAAAGCCGGTTAATCCGGTTT   | Top strand 30-mer oligo probe with mutation 24 in P <sub>m</sub> .    |
| KRI 108       | AAACCGGATTAACCGGCTTACTGTTTAA    | Bottom strand 30-mer oligo probe with mutation 24 in P <sub>m</sub> . |
| KRI 109       | TGTAAGCAGTAAAGCCGGTTAATCCGGTTT  | Top strand 30-mer oligo probe with mutation 25 in P <sub>m</sub> .    |
| KRI 110       | AAACCGGATTAACCGGCTTACTGCTTACA   | Bottom strand 30-mer oligo probe with mutation 25 in P <sub>m</sub> . |
| KRI 111       | TGTAACAATAAGCCGGTTAATCCGGCTT    | Top strand 30-mer oligo probe with mutation 26 in P <sub>m</sub> .    |
| KRI 112       | AAGCCGGATTAACCGGCTTAAATGTTTACA  | Bottom strand 30-mer oligo probe with mutation 26 in P <sub>m</sub> . |
| KRI 113       | TGTAACAGTTAAGCCGGTAATCCGGCTT    | Top strand 30-mer oligo probe with mutation 27 in P <sub>m</sub> .    |
| KRI 114       | AAGCCGGATTAACCGGCTTAATCTGTTTACA | Bottom strand 30-mer oligo probe with mutation 27 in P <sub>m</sub> . |
| KRI 115       | TGTAACAGATTAGCCGGAATATCCGGCTT   | Top strand 30-mer oligo probe with mutation 28 in P <sub>m</sub> .    |
| KRI 116       | AAGCCGGATTTCCGGCTAATCTGTTTACA   | Bottom strand 30-mer oligo probe with mutation 28 in P <sub>m</sub> . |
| KRI 117       | TGTAACAATTCGCCGGAATATCCGGCTT    | Top strand 30-mer oligo probe with mutation 29 in P <sub>m</sub> .    |
| KRI 118       | AAGCCGGATTTCCGGCGAATATGTTTACA   | Bottom strand 30-mer oligo probe with mutation 29 in P <sub>m</sub> . |
| KRI 119       | ATTAACGCCGGTTAAT                | Top strand 16-mer oligo with mutation 17 in P <sub>m</sub> .          |

Table 4 (continued).

| <b>Oligo numbers</b> | <b>Sequence</b>                       | <b>Comments</b>                                                 |
|----------------------|---------------------------------------|-----------------------------------------------------------------|
| KRI 120              | ATTAACCGGC <b>GTTAAT</b>              | Bottom strand 16-mer oligo with mutation 17 in P <sub>m</sub> . |
| KRI 121              | CAT <b>TAA</b> CGCCGGT <b>TAATC</b>   | Top strand 18-mer oligo with mutation 17 in P <sub>m</sub> .    |
| KRI 122              | GATTAACCGGC <b>GTTAATG</b>            | Bottom strand 18-mer oligo with mutation 17 in P <sub>m</sub> . |
| KRI 123              | ACAT <b>TAA</b> CGCCGGT <b>TAATCC</b> | Top strand 20-mer oligo with mutation 17 in P <sub>m</sub> .    |
| KRI 124              | GGATTAACCGGC <b>GTTAATGT</b>          | Bottom strand 20-mer oligo with mutation 17 in P <sub>m</sub> . |

The reaction mixture subjected to electrophoresis on a 10% non-denaturing, native acrylamide gel (19.5 ml milliQ water, 3 ml of 5X TBE buffer, 7.5 ml of 40% 29:1 acrylamide/bis-acrylamide solution, 150  $\mu$ l of 10% ammonium persulfate and 30  $\mu$ l TEMED solution) and run in 0.5 X TBE buffer at 260 V for 90min at 4°C. Initial exposure of the gels to X-OMAT BioMax-MR film was done without drying for 6-18 hours as indicated at -80°C. The exposed films were then developed and observed for His-Mor binding to the labeled probes. Quantitative assays for the His-Mor binding was done in a phosphoimager using the Image Quant software which gave the percentage of the probe bound by wild-type His-Mor for each of the probes.

## Results

### Spacer region mutagenesis and phenotypic frequency

For the spacer region, degenerate oligonucleotides were used to introduce all possible substitutions in the positions -42 through -45. The 3 primers that were designed for the purpose were:

1. MUT 50 with NCCN in the spacer region with degeneracy at the first and the fourth positions.
2. MUT 51 with GNNG in the spacer region with degeneracy at the middle two positions.
3. MUT 52 with NNNN in the spacer region with degeneracy at all four positions.

The initial analysis showed a very low incorporation of T in the oligonucleotides, so we ordered another primer MUT 520 with NNNN in the spacer region with a higher percentage of incorporation of T in all the four positions. The degenerate primers were designed for Mod-PCR (Chiang and Howe, 1993) in a way to change the targeted region by mutating the targeted bases to a random nucleotide using a 0.25 incorporation rate for each base A/T/G/C so that we got a library of substitutions in the region.

The mutagenesis was performed in PCR reactions using one degenerate primer as the top strand primer and a wild-type primer, MLK 7 as the bottom strand primer. All the mutagenic oligonucleotides were designed in such a way that they had the *Eco*RI site at its 5' end preceded by 3 extra nucleotides and the primer MLK 7 was designed with the *Bam*HI site at its 5' end followed by additional nucleotides to facilitate restriction digest. The resulting PCR products were digested with *Eco*RI and *Bam*HI and cloned into a similarly digested vector, pIA14 such that the P<sub>m</sub> region of pIA14 was removed and replaced with this library of mutations. The ligation mixture was transformed into MH13418 containing the plasmid pIA69 with the wild-type His-Mor gene inducible by IPTG. The vector itself is the reported plasmid with the P<sub>m</sub>-*lacZ* fusion. The activities of mutant promoters were examined by the color development of the transformants on MacConkey lactose plates with chloramphenicol (25  $\mu$ g/ml), ampicillin (40  $\mu$ g/ml) and IPTG (50  $\mu$ M).



The frequency of phenotypes in a mutant library was used as a preliminary indicator of the functional relevance of the bases in the spacer region. Mutant library with MUT 50 gave 13.2% white colonies indicating defective phenotypes. Mutant library with MUT 51 gave 11.2% white colonies, while the library with MUT 52 gave 9.8% whites. This initial analysis of the percentage of defective phenotypes was used as an indication as to the importance of the bases of the minor groove. Since just about 10% of colonies gave mutant phenotypes, it could be concluded that the minor groove spacer region can tolerate mutations and that interactions in the minor groove may not be base-specific. An equal proportion of red and white colonies were chosen for sequencing of the plasmids to determine the mutations incorporated in those positions.

The 62 unique red sequences that were isolated are given in the Table 5 along with their corresponding  $\beta$ -gal values. Three unique sequences (GGAC, GACG and ACCC) were identified which conferred defective phenotypes. All other sequences which gave us defective phenotypes had mutations elsewhere in the plasmid, so they were not used for further analysis. To re-check the three mutations which gave defective phenotypes, these specific primers mutations were introduced in the spacer region. This time, majority of the clones had the wild-type phenotype even though they contained the specific mutations that previously gave defective phenotypes. This could be attributed to mutations elsewhere in the plasmid which gave defective phenotypes in the original cloning.

### **Extensive mutagenesis of $P_m$**

In order to optimize Mor binding at  $P_m$ , extensive targeted mutagenesis was performed for the positions of  $P_m$  from -30 through -57 and specific mutations were introduced into those positions as given in Table 1. In addition to this, four specific mutations were also introduced in the 4-bp spacer region that was used as a subset of the spacer region mutants for analysis of *in vivo* transcription activation assays and gel shifts. Primers KRI 125 through KRI 153 were used for this purpose. The mutagenesis was performed in PCR reactions using one mutagenised primer as the top strand primer and a wild-type primer, MLK 7 as the bottom strand primer. All the mutagenic oligonucleotides were designed in such a way that they had an *EcoRI* site at its 5' end preceded by 3 extra nucleotides and the primer MLK 7 was designed with the *BamHI* site at its 5' end, again followed by additional nucleotides to facilitate restriction digestion of the PCR products with the two oligonucleotides. The resulting PCR products were digested with *EcoRI* and *BamHI* and cloned into a similarly digested vector, pIA12 such that the  $P_m$  region had incorporated the mutations designed for the specific positions. The ligation mixture was transformed into MH13418 containing the plasmid pIA69 with the wild-type *his-mor* gene inducible by IPTG. The vector itself is the reporter plasmid with the  $P_m$ -*lacZ* fusion. The activities of mutant promoters were examined by color development of the transformants on MacConkey lactose plates containing half the amount of lactose and also chloramphenicol (25  $\mu$ g/ml), ampicillin (40  $\mu$ g/ml) and IPTG (50  $\mu$ M).

Table 5: Mutations of the spacer region which gave red colonies.

| <b>-45</b><br><b>G</b> | <b>-44</b><br><b>C</b> | <b>-43</b><br><b>C</b> | <b>-42</b><br><b>G</b> | <b>β-GAL</b><br><b>VALUES</b> |
|------------------------|------------------------|------------------------|------------------------|-------------------------------|
| G                      | C                      | C                      | A                      | 784                           |
| A                      | C                      | C                      | A                      | 1310                          |
| A                      | C                      | C                      | G                      | 1189                          |
| G                      | C                      | C                      | C                      | 1251                          |
| G                      | G                      | G                      | G                      | 1058                          |
| G                      | A                      | C                      | G                      | 689                           |
| G                      | A                      | A                      | G                      | 946                           |
| G                      | G                      | C                      | G                      | 1014                          |
| G                      | G                      | A                      | G                      | 1360                          |
| G                      | A                      | G                      | G                      | 729                           |
| C                      | A                      | A                      | A                      | 1122                          |
| T                      | C                      | G                      | G                      | 290                           |
| A                      | G                      | C                      | G                      | 993                           |
| A                      | A                      | A                      | G                      | 445                           |
| G                      | A                      | G                      | A                      | 588                           |
| G                      | G                      | C                      | C                      | 708                           |
| C                      | G                      | A                      | G                      | 1222                          |
| G                      | C                      | A                      | G                      | 637                           |
| A                      | C                      | G                      | G                      | 1227                          |
| C                      | A                      | A                      | C                      | 640                           |
| C                      | G                      | G                      | C                      | 1223                          |
| G                      | G                      | G                      | A                      | 964                           |
| G                      | A                      | G                      | C                      | 772                           |
| C                      | A                      | G                      | G                      | 945                           |
| C                      | C                      | G                      | C                      | 574                           |
| C                      | C                      | G                      | G                      | 485                           |
| G                      | A                      | T                      | G                      | 547                           |
| C                      | C                      | C                      | A                      | 516                           |
| A                      | A                      | G                      | G                      | 713                           |
| G                      | T                      | A                      | G                      | 850                           |
| G                      | C                      | G                      | A                      | 439                           |
| A                      | G                      | A                      | A                      | 864                           |
| A                      | A                      | T                      | G                      | 1304                          |
| G                      | T                      | A                      | T                      | 352                           |
| T                      | T                      | T                      | G                      | 833                           |
| T                      | T                      | A                      | A                      | 392                           |
| A                      | G                      | C                      | T                      | 1264                          |
| G                      | T                      | T                      | T                      | 888                           |
| T                      | T                      | T                      | T                      | 410                           |
| C                      | C                      | T                      | A                      | 664                           |
| T                      | G                      | T                      | T                      | 574                           |

Table 5 (continued).

| <b>-45</b> | <b>-44</b> | <b>-43</b> | <b>-42</b> | <b>β-GAL</b>  |
|------------|------------|------------|------------|---------------|
| <b>G</b>   | <b>C</b>   | <b>C</b>   | <b>G</b>   | <b>VALUES</b> |
| A          | T          | T          | T          | 293           |
| C          | T          | T          | A          | 1149          |
| T          | A          | G          | T          | 480           |
| A          | T          | T          | A          | 902           |
| T          | T          | C          | G          | 927           |
| A          | A          | T          | T          | 729           |
| G          | T          | T          | G          | 745           |
| C          | G          | T          | C          | 849           |
| C          | C          | C          | T          | 934           |
| G          | C          | T          | T          | 551           |
| C          | A          | T          | T          | 433           |
| T          | G          | G          | T          | 580           |
| G          | C          | C          | C          | 1022          |
| T          | G          | C          | G          | 913           |
| G          | A          | C          | T          | 1077          |
| C          | T          | T          | T          | 549           |
| A          | C          | T          | C          | 756           |
| A          | C          | C          | C          | 917           |
| C          | T          | C          | T          | 728           |
| T          | C          | A          | G          | 323           |
| A          | A          | G          | A          | 549           |
| A          | A          | A          | A          | 579           |

The positions of the bases are given at the top of each column with respect to the transcription start site of +1. Immediately following that, in the first row in red are the wild-type bases at each of those positions. The β-gal values are the readings from one experimental assay and are given with respect to the wild-type set to 1000 and the negative control calculated to be 2.

## Observation of phenotypes

The phenotypes of the transformants were observed after incubation of the plates at 37°C for 16 hours. The colonies were scored relative to the wild-type strain MH20655 (positive control) which was red in color and the strain MH20656 (negative control) which had the vector plasmid pIA12, without P<sub>m</sub> and was white on the MacConkey plates. The phenotypes were as given in Table 6. There were red colonies, reds with a white halo around them and white colonies as a result of the different mutations. The red colonies indicated those with wild-type promoter activity. The red ones with a small white halo indicated a slightly reduced activity as compared with wild-type and colonies which were red with a large white halo indicated an even more reduced activity. White colonies indicated a defective phenotype similar to the negative control with almost no detectable promoter activity.

## *In vivo* transactivation assay

After one hour of Mor expression from the P<sub>lacUV5</sub> promoter, the ability of Mor to activate transcription from P<sub>m</sub> with the different mutations was assayed by measuring β-galactosidase activity. They were done in parallel with a wild-type control. They are discussed with the results of the gel shifts as the effect of each mutation is discussed. The β-gal values are an average of two assays and are relative to the wild-type activity from wild-type P<sub>m</sub> set to 1000.

## Gel retardation assay results

Gel retardation assays were performed to test the ability of His-Mor to bind to the promoters with specific mutations. They were done at two concentrations of wild-type His-mor, 200 ng and 400 ng to determine the shift caused by the lowest concentration of the protein. Each gel had a control with 30-mer oligonucleotide which had the wild-type promoter and the intensity of the shift by the mutagenised oligonucleotides was scored relative to the control set to 100. The effects of gel shifts due to the mutations are discussed as a percentage of the wild-type in most cases except in cases where there was no detectable shift to assign a percent value to it. The results of the gel shift in correlation with the plate phenotypes and *in vivo* β-galactosidase activity levels are discussed on the following pages from Figure 15 through Figure 26. They are arranged as per the sequence of the mutations and the gels.

Both the *in vivo* transactivation assay values and gel retardation assay results are discussed with each mutation and also in the summarized results and discussion section.

Table 6: Phenotypes of strains with specific promoter mutations.

| <b>Mutation</b> | <b>Mutation alterations</b>  | <b>Strain</b> | <b>Colony phenotype</b>                        |
|-----------------|------------------------------|---------------|------------------------------------------------|
| 1               | -46 A to C                   | MH20561       | Red with small white halo                      |
| 2               | -41 G to T                   | MH20616       | Red with small white halo                      |
| 3               | -48 A to C and<br>-39 T to G | MH20620       | White                                          |
| 4               | -37 A to C                   | MH20624       | Red                                            |
| 5               | -52 C to G                   | MH20564       | Red with small white halo                      |
| 6               | -52 C to A and<br>-35 C to T | MH20627       | Red                                            |
| 7               | -34 C to T                   | MH20630       | Red                                            |
| 8               | -53 A to G                   | MH20567       | Red with small white halo                      |
| 9               | -33 G to T                   | MH20600       | Red                                            |
| 10              | -54 A to C                   | MH20570       | Red with larger white halo than others         |
| 11              | -32 G to T                   | MH20603       | Red                                            |
| 12              | -31 C to A                   | MH20607       | Red                                            |
| 13              | -56 T to A and<br>-31 C to T | MH20610       | Red                                            |
| 14              | -57 G to T                   | MH20575       | Red with small white halo                      |
| 15              | -57 G to T and<br>-30 T to A | MH 20614      | Red with small white halo                      |
| 16              | -50 G to T and<br>-41 G to T | MH20631       | Red with small white halo                      |
| 17              | -50 G to T and<br>-46 A to C | MH20576       | Red with small white halo                      |
| 18              | -42 through -45<br>AAAA      | MH20584       | Red with small white halo                      |
| 19              | -42 through -45<br>TAAT      | MH20587       | Red with large white halo, almost red centered |
| 20              | -42 through -45<br>GGCC      | MH20590       | Red with small white halo                      |

Table 6 (continued).

| <b>Mutation</b>     | <b>Mutation alterations</b>              | <b>Strain</b> | <b>Colony phenotype</b> |
|---------------------|------------------------------------------|---------------|-------------------------|
| 21                  | -42 through -45<br>GGGG                  | MH20591       | Red                     |
| 22                  | -57 G to T, -53 A to G<br>and -31 C to T | MH20635       | Red                     |
| 23                  | -57 G to T and -53 A<br>to G             | MH20594       | Red                     |
| 24                  | -57 G to T and -31 C<br>to T             | MH20638       | Red                     |
| 25                  | -53 A to G and -31 C<br>to T             | MH20641       | Red                     |
| 26                  | -50 G to T                               | MH20597       | Red                     |
| 27                  | -48 A to T and -39 T<br>to A             | MH20644       | White                   |
| 28                  | -49 to -47 ATT and<br>-40 to -38 AAT     | MH20646       | White                   |
| 29                  | -50 to -46 TATTC<br>and -40 to -38 AAT   | MH20651       | White                   |
| Positive<br>Control | Wild-type P <sub>m</sub>                 | MH20655       | Red                     |
| Negative<br>control | Without P <sub>m</sub>                   | MH20656       | White                   |

Figure 15. Gel shift for mutations 1, 2 and 3

Mutation 1 (-46 A to C): This mutation led to red colonies with a small white halo around them indicating somewhat lower activity, which correlated with a  $\beta$ -galactosidase activity of 518 relative to the wild-type activity set to 1000. The gel shift for this mutation revealed a slightly higher binding as compared to that of the wild-type at both concentrations of His-Mor (400 and 800 ng). Thus, changing this A to a C is tolerated by the promoter without drastically changing its activity.

Mutation 2 (-41 G to T): This mutation gave red colonies with a small white halo around them indicating lower activity which correlated with its  $\beta$ -galactosidase activity of 377 relative to the wild-type activity set to 1000. The gel retardation assay with this mutation showed significantly lower binding as compared to the wild-type at both the concentrations of His-Mor (400 and 800 ng). This weak binding to the promoter DNA would explain its lower capacity to activate transcription from the promoter, as indicated by its lower  $\beta$ -galactosidase activity. This indicates that the G in this position is important for Mor binding.

Mutation 3 (-48 A to C and -39 T to G): This mutation gave white colonies on MacConkey plates, indicating a severely defective phenotype similar to the negative control with a deleted promoter  $P_m$ . Mor failed to bind to the promoter DNA as seen from the gel shift data at levels even as high as 800 ng of protein, and this explained its almost undetectable  $\beta$ -galactosidase activity of 0.3. As wild-type His-Mor failed to bind to the promoter with this mutation, it could not activate transcription from that promoter as it could not recruit RNA polymerase to the promoter.

These specific bases seem to be critical for the promoter binding and transactivation function of His-Mor. It also indicates that the specific bases are important for Mor recognition and binding rather than just symmetry because even though this mutation maintained the symmetry with a G/C combination as opposed to the wild-type of A/T combination, yet it failed to bind His-Mor.

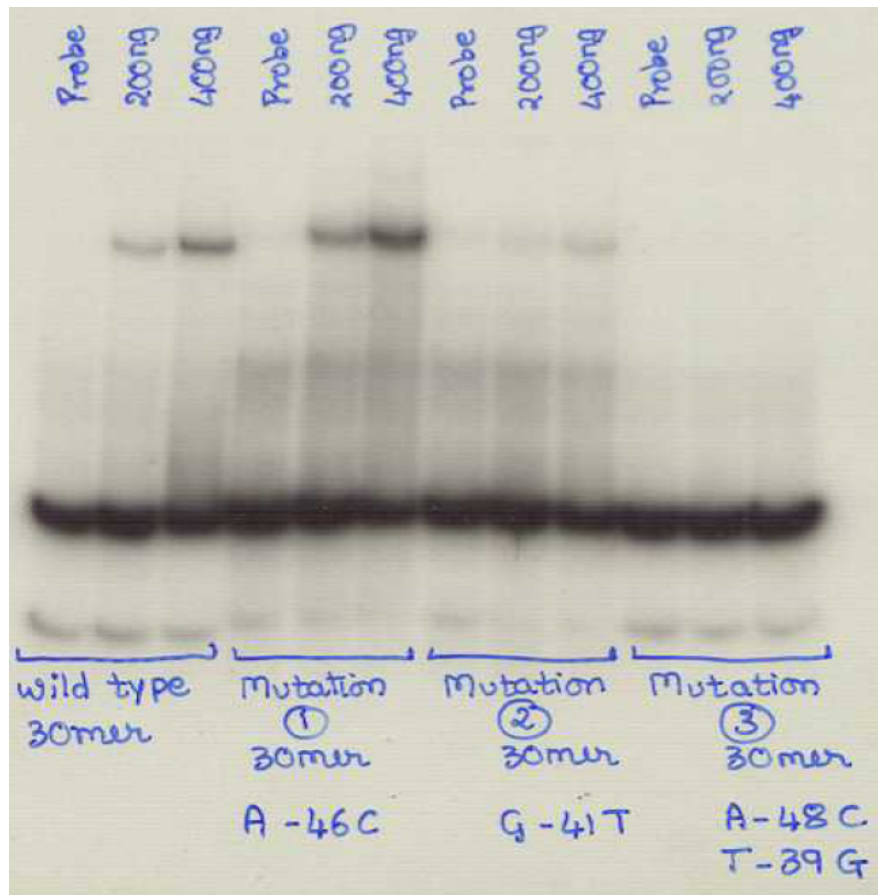




Figure 16: Gel shift for mutations 4, 5 and 6.

Mutation 4 (-37 A to C): This mutant exhibited a red phenotype on MacConkey plates, indicating a wild-type phenotype and functionally active promoter. This was further justified by its His-Mor binding which was similar to the wild-type promoter at both concentrations of His-Mor (400 and 800 ng), and its *in vivo*  $\beta$ -galactosidase activity of 951, that was very close to the wild-type promoter activity. Even though a symmetrical change was made at this position with respect to -50G centered at -43.5, it did not increase the binding of His-Mor. Thus, symmetry at this position may not contribute significantly to Mor binding and activation, as it tolerates both C and A at this position without much effect.

Mutation 5 (-52 C to G): This mutation imparted a red phenotype with a small white halo around the colonies indicating a slightly lower promoter activity which correlated with its  $\beta$ -galactosidase activity of 639 relative to the wild-type activity set to 1000. The gel retardation assay indicated that the promoter with this mutation had the capacity to bind wild-type His-Mor to about 80% of the wild-type promoter at both concentrations of His-Mor (400 and 800 ng). This position lies just outside the Mor-binding region as indicated in footprinting analysis (Kahmeyer-Gabbe and Howe, 1996). Making this position symmetrical with respect to -35C centered at -43.5 reduced its capacity to bind His-Mor and gave a lower promoter activity as compared to wild-type. This indicates the contribution of this position to Mor binding and promoter activity.

Mutation 6 (-52 C to A and -35 C to T): This double mutation imparted a red phenotype with a small white halo around the colonies similar to that of mutation 5, indicating a slightly lower promoter activity, which correlated with its  $\beta$ -galactosidase activity of 597 relative to the wild-type activity set to 1000. The gel retardation assay showed that the promoter with this double mutation had the capacity to bind wild-type His-Mor to about 70% of the wild-type promoter, which explains its lower level of *in vivo* promoter activity as measured in the  $\beta$ -galactosidase assay. This mutation and mutation 5, both indicate the specific contribution of -52C to Mor binding and promoter activity. Also, it is not the symmetry at these positions that is important for Mor binding but the presence of the specific bases -52C and -35C that stabilize Mor binding to  $P_m$ .

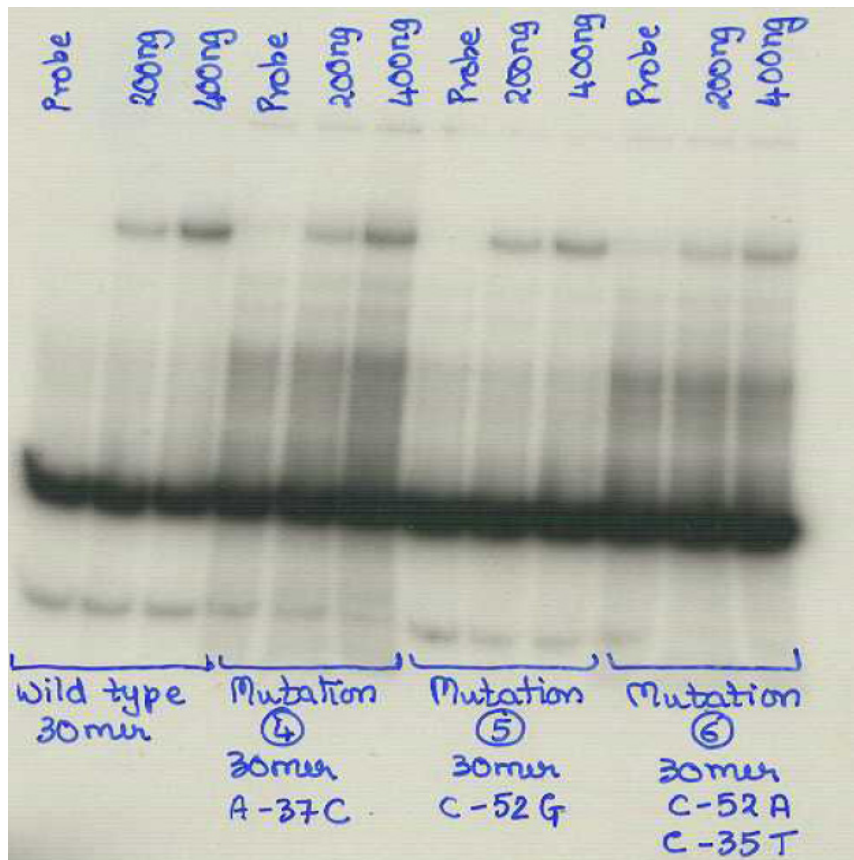


Figure 17: Gel shift for mutations 7, 8 and 9.

Mutation 7 (-34 C to T): This mutant exhibited a red (wild-type) phenotype on MacConkey plates indicating a functionally active promoter. This conclusion was further supported by its binding to His-Mor and its *in vivo*  $\beta$ -galactosidase activity of 809, both of which were similar to those of wild-type promoter. Even though a symmetrical change was made at this position with respect to -53A centered at -43.5, it did not increase the binding of His-Mor. Thus, symmetry at this position may not contribute significantly to Mor binding and activation as it tolerates both C and T at this position without much effect. This position lies just outside the Mor-binding region of  $P_m$  and as expected, it does not significantly contribute to Mor binding. The  $\beta$ -galactosidase activity indicated that the nucleotide at this position is also not very important for RNA polymerase binding and that it can tolerate base changes.

Mutation 8 (-53 A to G): This mutation imparted a red phenotype with a small white halo around the colonies similar to that of mutation 5, indicating a slightly lower promoter activity which correlated with its  $\beta$ -galactosidase activity of 630 relative to the wild-type activity set to 1000. The gel retardation assay indicated that the promoter with this mutation had the capacity to bind wild-type His-Mor to about 80% of the wild-type promoter which correlates with its lower level of *in vivo* promoter activity as measured in the  $\beta$ -galactosidase assay. This mutation also introduced a symmetrical substitution similar to mutation 7, but with a G/C combination as opposed to the A/T combination that was introduced by mutation 7 at these positions. Even though a promoter with this mutation binds His-Mor almost as effectively as wild-type, this position probably contributes to transactivation through interaction with RNA polymerase. As noted earlier, this position lies in the region predicted to interact with one of the  $\alpha$ -CTD subunits of the RNAP (Ma and Howe, 2004). This would explain why a mutation at this position bound the protein but displayed a lower *in vivo* activity.

Mutation 9 (-33 G to T): This mutation imparted a red phenotype with a small white halo around each colony similar to that of mutation 8, indicating a slightly lower promoter activity which correlated with its  $\beta$ -galactosidase activity of 654. The gel retardation assay indicated that the promoter with this mutation had the capacity to bind wild-type His-Mor to about 80% of the wild-type promoter which explains its lower level of *in vivo* promoter activity as measured in the  $\beta$ -galactosidase assay. This mutation introduces a symmetrical change at this position with respect to -54A. The closer-to-wild-type binding by His-Mor but significantly lower *in vivo* activity indicates that even though it lies beyond the Mor-binding site, -33G contributes to the activation of the promoter probably through its interaction with RNAP.



Figure 18: Gel shift for mutations 10, 11 and 12.

Mutation 10 (-54 A to C): This mutation imparted a red phenotype with a larger white halo around the colonies similar to mutation 8, indicating a lower promoter activity which correlated with its  $\beta$ -galactosidase activity of 335. The gel retardation assay indicated that the promoter with this mutation had the capacity to bind wild-type His-Mor to about 85% of the wild-type promoter at both concentrations of His-Mor (400 and 800 ng). The *in vivo* activity due to this mutation is much lower and does not seem to be because of the binding effect of Mor to the promoter with this mutation. This mutation also introduced a symmetrical substitution similar to mutation 9, but with a G/C combination as opposed to the A/T combination that was introduced by mutation 9 at these positions. The closer-to-wild-type binding by His-Mor but significantly lower *in vivo* activity indicates that this position probably contributes to the transactivation through interaction with RNA polymerase. As noted earlier, this position lies in the region predicted to interact with one of the  $\alpha$ -CTD subunits of the RNAP, just like position -53A that was affected by mutation 8 (Ma and Howe, 2004). This would explain why a mutation at this position bound the protein but displayed a lower *in vivo* activity.

Mutation 11 (-32 G to T): This mutation conferred a red (wild-type) phenotype on MacConkey plates, indicating a functionally active promoter. Its binding of His-Mor was similar to that of wild-type promoter and its *in vivo*  $\beta$ -galactosidase activity of 1133 was slightly higher than the wild-type promoter activity. Even though a symmetrical change was made at this position with respect to -55A centered at -43.5, it did not increase the binding of His-Mor. Thus, symmetry at this position may not contribute to Mor binding and activation as both T and G are tolerated at this position, a finding understandable by the observation that this position lies outside the Mor-binding region of  $P_m$ . The higher *in vivo* activity of the promoter due to this mutation indicates the possible contribution of this position to activation through interaction with RNAP which is increased by substituting a T at this position.

Mutation 12 (-31 C to A): Colonies with this mutation exhibited a red (wild-type) phenotype on MacConkey plates, indicating a functionally active promoter. The *in vivo*  $\beta$ -galactosidase activity was 1139, slightly higher than wild-type promoter activity. The gel shift showed a lower binding capacity of the promoter with this mutation to His-Mor; 60% to that of the wild-type promoter at both concentrations of His-Mor (400 and 800 ng). Even though a symmetrical change was made at this position with respect to -56T centered at -43.5; it did not increase the binding of His-Mor. The lower binding of His-Mor could be an experimental artifact as this position lies well outside the Mor-binding region. Thus, symmetry at this position may be irrelevant to Mor binding. The higher *in vivo* activity of the promoter with this mutation indicates the possible contribution of this position to activation through interaction with RNAP, which is increased by substituting an A at this position.

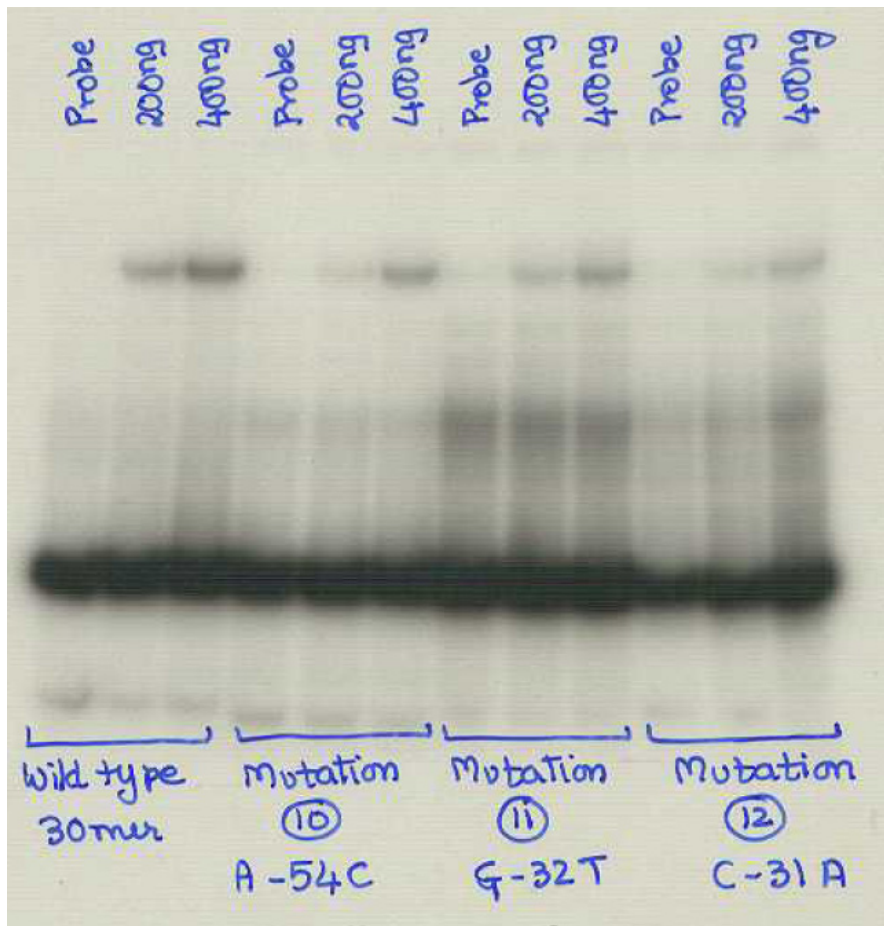


Figure 19: Gel shift for mutations 13, 14 and 15.

Mutation 13 (-56 T to A and -31 C to T): This mutation led to a red (wild-type) phenotype on MacConkey plates indicating a functionally active promoter. The *in vivo*  $\beta$ -galactosidase activity was 1272, higher than the wild-type promoter activity observed for -31 C to T alone. The gel retardation assay revealed similar to wild-type binding. The closer-to-wild-type binding by His-Mor at both concentrations (400 and 800 ng) but significantly higher *in vivo* promoter activity indicates that this position probably contributes to transactivation through interaction with RNA polymerase. As noted earlier, -56 lies in the region predicted to interact with one of the  $\alpha$ -CTD subunits of RNAP, just like position -53A and -54A that were affected by mutation 8 and mutation 10 respectively (Ma and Howe, 2004). The difference was that mutations 8 and 10 negatively affected the promoter activity while this mutation contributed positively to promoter activity. This would explain why a mutation at this position bound His-Mor at levels similar to the wild-type promoter but displayed a higher *in vivo* activity. Recall that mutation -31 C to T alone gave a higher than wild-type promoter activity, it could indicate a positive contribution of the single substitution in this mutation to binding with the  $\sigma$ -CTD of the RNAP more effectively. Hence, if at all mutation -56 T to A contributes to promoter activity, it would have a positive effect.

Mutation 14 (-57 G to T): Colonies with this mutation exhibited a red phenotype on MacConkey plates indicating a functionally active promoter. Its binding to His-Mor was similar to the wild-type promoter and its *in vivo*  $\beta$ -galactosidase activity was 902, very close to the wild-type promoter activity. The similarity to wild-type binding can possibly be explained by the observation that this position is well outside of the Mor-binding region of  $P_m$ . The high *in vivo* activity of the promoter due to this mutation indicates a possible contribution of this position to activation through its interaction with the RNAP  $\alpha$ -CTD subunit which is increased by substituting a T at this position; correlating with the observation that this position lies in the region predicted to interact with one of the  $\alpha$ -CTD subunits of the RNAP (Ma and Howe, 2004). This would explain why a mutation at this position bound the protein but displayed a high *in vivo* activity.

Mutation 15 (-57 G to T and -30 T to A): This mutation imparted a red phenotype with a small white halo around the colonies, indicating somewhat lower promoter activity than the wild-type, which correlated with its  $\beta$ -galactosidase activity of 647. The gel retardation assay indicated that the promoter with this mutation had the capacity to bind wild-type His-Mor to about 80% of the wild-type promoter at both concentrations of His-Mor (400 and 800 ng) which partially explains its lower level of *in vivo* promoter activity as measured in the  $\beta$ -galactosidase assay. The closer-to-wild-type binding by His-Mor but significantly lower *in vivo* activity indicates that these positions contribute to the activation of the promoter probably through its interaction with RNAP. Since mutation 14 indicated the contribution from the individual substitution at -57 which gave a wild-type activity, the effect of this double substitution is probably due to the contribution of the -30 position. Changing the -30 T to A reduces the promoter activity, probably indicating an important contribution of the -30T to RNAP interaction through its  $\sigma$ -CTD subunit which binds to this region of the promoter.

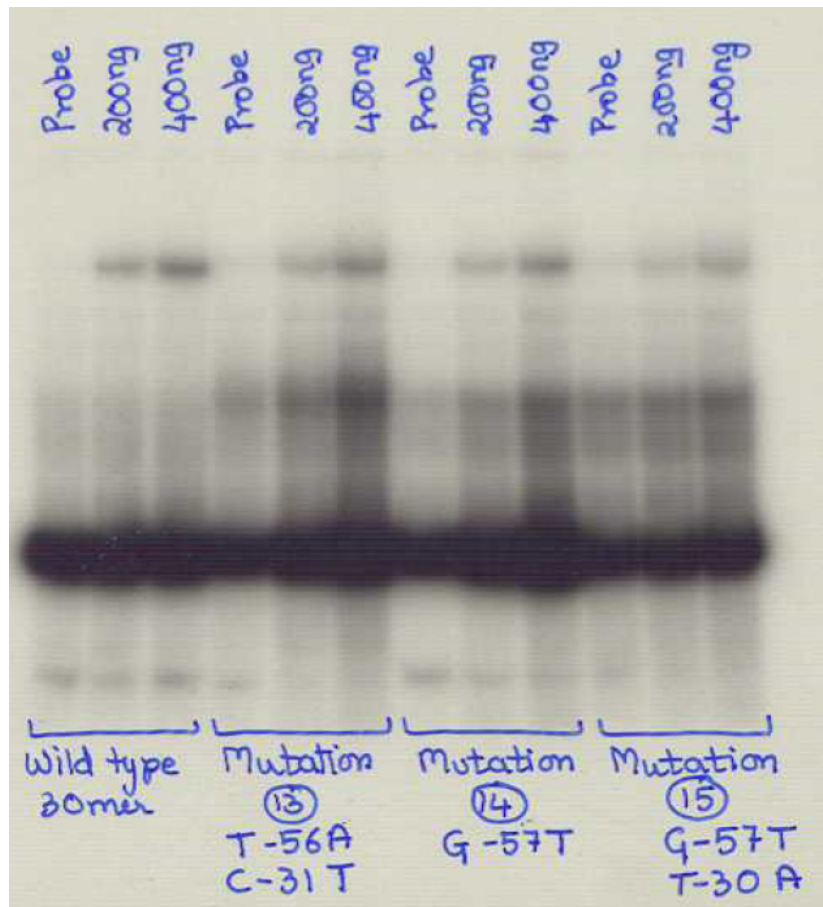




Figure 20: Gel shift for mutations 16 and 17.

Mutation 16 (-50 G to T and -41 G to T): This mutation imparted a red phenotype with a small white halo around the colonies indicating a lower activity, which correlated with its  $\beta$ -galactosidase activity of 421. The gel retardation assay with this mutant gave only about 50% of the wild-type promoter binding to His-Mor at both concentrations of His-Mor (400 and 800 ng). This weak binding to the promoter DNA would explain its lower capacity to activate transcription from the promoter indicated by almost 60% reduction in  $\beta$ -galactosidase activity. It indicates the importance of the nucleotides in these positions for Mor binding. Even though the 6-bp segments that bind Mor were made perfectly symmetrical to each other with an increased AT-rich sequence, binding was negatively affected by the mutations, indicating that Mor binding to promoter DNA in this region depends on specific bases and not the symmetry factor in binding the two monomers of the Mor dimer. When compared to the effect of the single substitutions at these positions, mutation 2 (-41 G to T) and mutation 26 (-50 G to T) both of which reduced Mor binding to promoter DNA, the effect of this mutation seems to be the combined effect of the two single mutations.

Mutation 17 (-50 G to T and -46 A to C): This double mutation imparted a red phenotype with a small white halo around the colonies, indicating a slightly lower than wild-type promoter activity which correlated with its  $\beta$ -galactosidase activity of 671. The *in vitro* binding assay for this mutation was noteworthy in the fact that it repeatedly gave significantly higher binding by His-Mor (almost three times the binding as compared to the wild-type promoter). These mutations make the 6-bp segments perfectly symmetrical to each other and it is these segments on either side of the spacer GCCG that binds the two monomers of the Mor dimer. Mor binding to the promoter with this double mutation imparts a high level of stability to the Mor- $P_m$  complex indicated by the dramatically higher binding in the gel shift assay. The hypothesis that increased Mor binding would increase the promoter activity is refuted in this case. Since Mor recruits the RNAP to  $P_m$ , the interactions between Mor and RNAP should be optimum for efficient recruitment of RNAP to the promoter and also optimum for the release of  $\sigma 70$  during promoter clearance and transcription elongation by the core RNAP. The higher binding of Mor might interfere with the efficient release of  $\sigma 70$  during promoter clearance by lowering its dissociation from the promoter. This might possibly explain why this mutation gave higher binding with His-Mor but lower *in vivo*  $\beta$ -galactosidase activity.



Figure 21: Gel shift for mutations 18 and 19.

Mutation 18 (-42 through -45 AAAA): This mutation led to red colonies with a small white halo around them indicating lower promoter activity which correlated with its  $\beta$ -galactosidase activity of 539. The gel retardation assay with this mutant promoter showed significantly lower binding, only 40 % as compared to the wild-type promoter. This weak binding to the promoter DNA would explain its reduced ability to activate transcription from the promoter as indicated by its lower  $\beta$ -galactosidase activity. Since side chains of Q68 and Y70 from the  $\beta$ -strand of Mor are proposed to interact with the minor groove, these nucleotides in the minor groove may be important for Mor binding. The A-rich sequence may change the dimensions of the minor groove altering the stereochemistry of Mor interactions within the minor groove and the two major grooves on either side.

Mutation 19 (-42 through -45 TAAT): This mutation resulted in red colonies with large white halos, indicating lower promoter activity consistent with a  $\beta$ -galactosidase activity of 344. In the gel retardation assay, this mutation led to significantly lower Mor binding (~40%) as compared to the wild-type. The weak binding of His-Mor to the promoter DNA with this mutation would explain its reduced transcription activity. Since side chains of Q68 and Y70 from the  $\beta$ -strand of Mor are proposed to interact with the minor groove, these bases in the minor groove may be important for Mor binding. The A-rich sequence may change the dimensions of the minor groove, altering the stereochemistry of Mor interactions at the minor groove and in the two major grooves on either side.

The effect of the mutations 18 and 19, which make the minor groove spacer AT-rich, indicates the preference for GC-rich sequences in the region. Thus it is not the symmetry in this spacer that is important, but the presence of GC-rich sequences.

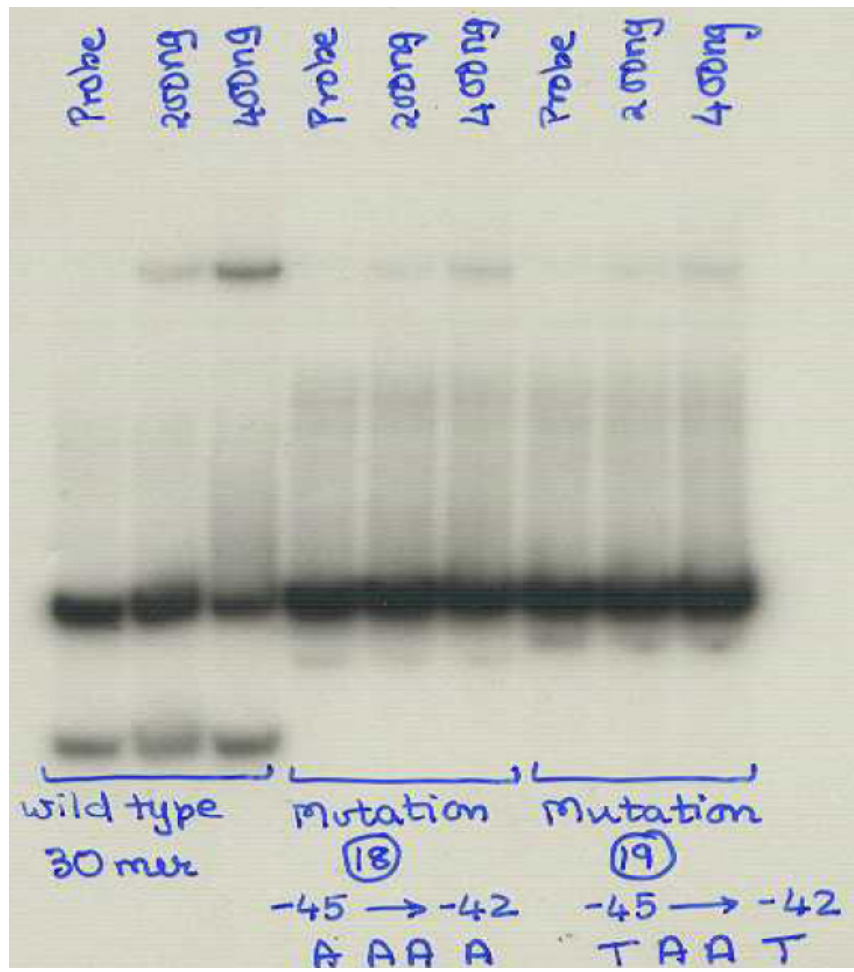


Figure 22: Gel shift for mutations 20 and 21.

Mutation 20 (-42 through -45 GGCC): The mutant with these substitutions gave red colonies with small white halos around them indicating slightly lower promoter activity which correlated with its  $\beta$ -galactosidase activity of 618. The gel shift for this mutation revealed a lower binding, of about 60% as compared to that of the wild-type at both the concentrations of His-Mor used (400 and 800 ng). Thus the lower promoter activity due to this mutation is possible because of the reduced binding of His-Mor. The difference in the *in vivo* activity correlates with the difference in the binding of His-Mor at the promoter with this mutation. The fact that this mutation maintains the GC-richness of the minor groove spacer but significantly reduces binding by His-Mor indicates that it is not the GC-richness alone, but the specific bases in the minor groove that are important. The symmetry introduced by the wild-type may be important for Mor binding.

Mutation 21 (-42 through -45 GGGG): This mutation imparted a red (wild-type) phenotype to the colonies on MacConkey plates indicating a functionally active promoter. This was consistent with its binding, which was similar to the wild-type promoter at both the concentrations of His-Mor used (400 and 800 ng) and its *in vivo*  $\beta$ -galactosidase activity of 768. The lower  $\beta$ -galactosidase activity may indicate the importance of these bases in the minor groove.

Recall that most of the base substitutions in the minor groove gave red colonies on MacConkey plates; leading to the conclusion that the bases may not contribute significantly to Mor binding and function. Comparing mutations 18 and 19 with mutations 20 and 21 indicates that a somewhat GC-rich sequence may be preferred in the minor groove spacer, though a mutation to AT-rich sequences does not very drastically change the activity. Comparing mutation 20 and 21 in terms of their binding and  $\beta$ -galactosidase activity, it is possible that the outer G's might be more important than the inner C's of the GCCG spacer. Even though minor groove interactions appear to be important for Mor binding, mutations of the bases in the minor groove do not affect Mor binding as much mutations of the amino acids Q68 and Y70 (Kumaraswami, 2005). Hence the amino acids probably interact with the minor groove spacer in a base non-specific manner.

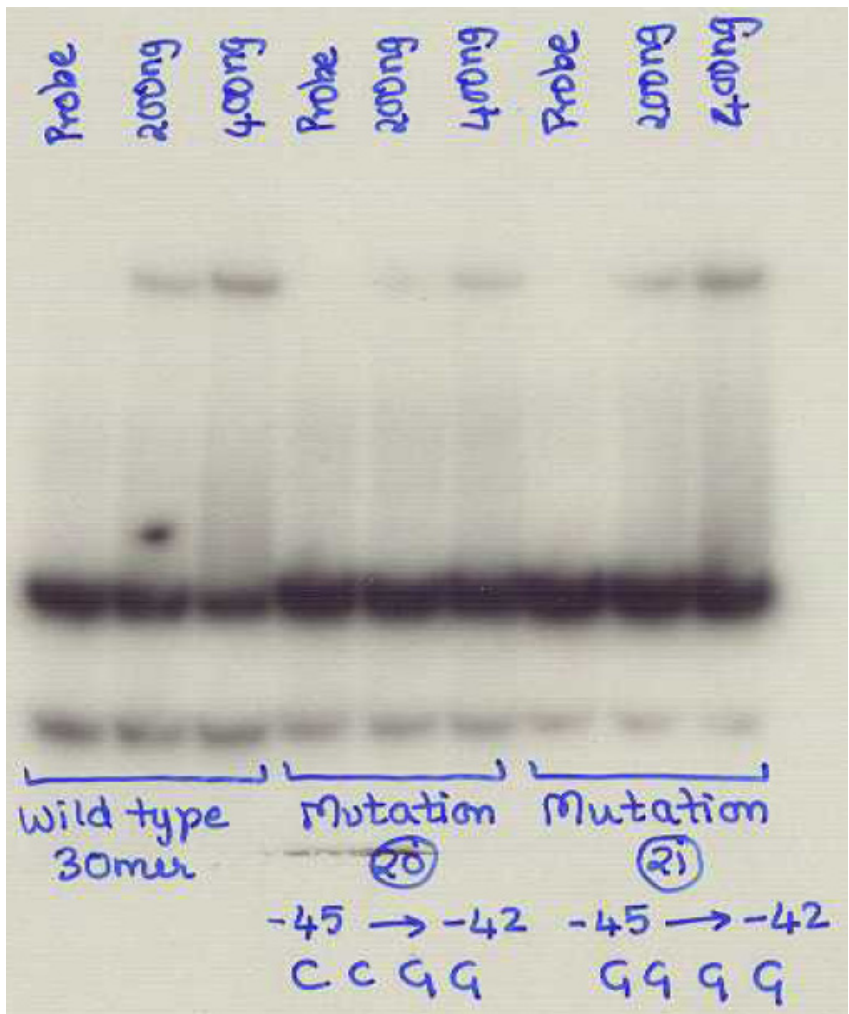


Figure 23: Gel shift for mutations 22 and 23.

Mutation 22 (-57 G to T, -53 A to G and -31 C to T): The mutant with these base changes mutation exhibited a red phenotype on MacConkey plates indicating it has a functionally active promoter. This conclusion was further supported by its His-Mor binding to a level similar to that of the wild-type promoter and an *in vivo*  $\beta$ -galactosidase activity of 1021, which was close to the wild-type promoter activity. Even though the single mutations individually at these positions recorded a high  $\beta$ -galactosidase activity, the effect is not cumulative as indicated from the triple mutation. The similarity to wild-type binding can possibly be explained by the fact that all three bases changed are outside of the Mor-binding region of  $P_m$ . The high *in vivo* activity of the promoter due to this mutation indicates the possible contribution of these positions to activation through interaction with RNAP. Even though a promoter with these mutations binds His-Mor as effectively as wild-type, these positions probably contribute to the transactivation through interaction with RNA polymerase. As noted earlier, these positions lie in the region predicted to interact with RNAP; -53 and -57 are in the region of the promoter predicted to interact with one of the  $\alpha$ -CTD subunits of the RNAP and -31 is in the region that interacts with  $\sigma$ -CTD of the RNAP (Artsimovitch and Howe, 1996). This would explain why a mutation at this position bound the protein as efficiently as wild-type promoter and displayed a high *in vivo* activity.

Mutation 23 (-57 G to T and -53 A to G): This mutation conferred a red phenotype on MacConkey plates indicating a functionally active promoter. This was further supported by its binding assay which showed similarity to wild-type promoter. The *in vivo*  $\beta$ -galactosidase activity was lower than the wild-type, only 642 relative to the wild-type of 1000. The similar to wild-type binding can possibly be explained by the observation that this position is outside of the Mor-binding region of  $P_m$ . Even though a promoter with these mutations binds His-Mor as effectively as wild-type, the low *in vivo* activity of the promoter due to this mutation indicates the possible contribution of these positions to activation through interaction with RNAP. As noted earlier, these positions lie in the region predicted to interact with RNAP; -53 and -57 are in the region of the promoter predicted to interact with one of the  $\alpha$ -CTD subunits of the RNAP (Ma and Howe, 2004). This would explain why this mutation bound the protein as efficiently as the wild-type promoter but displayed a lower *in vivo* activity.

Comparing mutations 22 and 23, the higher promoter activity of mutation 22 is probably due to contribution from -31T which is in the region interacting with  $\sigma$ -CTD of the RNAP than the other two (-57 G to T and -53 A to G).





Figure 24: Gel shift for mutations 24 and 25.

Mutation 24 (-57 G to T and -31 C to T): This mutation imparted a red phenotype to the colonies on MacConkey plates indicating a functionally active promoter. The *in vivo*  $\beta$ -galactosidase activity was 1272, that was higher than the wild-type promoter activity as observed for -31 C to T. The gel retardation assay revealed a similarity to wild-type binding. The closer-to-wild-type binding by His-Mor at both the concentrations (400 and 800 ng) but significantly higher *in vivo* activity indicates that these positions probably contribute to transactivation through interaction with RNA polymerase. These positions lie in the region predicted to interact with RNAP subunits; position -57 is in the region of the promoter predicted to interact with one of the  $\alpha$ -CTD subunits of the RNAP and position -31 is the region that is predicted to interact with  $\sigma$ -CTD of the RNAP (Artsimovitch and Howe, 1996; Ma and Howe, 2004). This would explain why mutation at these positions bound the protein as efficiently as wild-type promoter but displayed a significantly higher *in vivo* activity. Comparing this with mutations 22 and 23 also indicates the significant contribution of the -31T to promoter activity.

Mutation 25 (-53 A to G and -31 C to T): This mutation conferred a red phenotype to colonies on MacConkey plates indicating a functionally active promoter which correlated with the *in vivo*  $\beta$ -galactosidase activity of 816, which was close to the wild-type promoter activity. The gel retardation assay revealed a similarity to wild-type binding. The closer-to-wild-type binding by His-Mor at both the concentrations of His-Mor (400 and 800 ng) but slightly lower *in vivo* activity indicates that these positions probably contribute to the transactivation through interaction with RNA polymerase. These positions lie in the region predicted to interact with RNAP; -53 is in the region of the promoter predicted to interact with one of the  $\alpha$ -CTD subunits of the RNAP and the -31 is the region that interacts with  $\sigma$ -CTD of the RNAP (Artsimovitch and Howe, 1996; Ma and Howe, 2004). This would explain why a mutation at this position bound the protein as efficiently as wild-type promoter but displayed a slightly lower *in vivo* activity. Comparing the *in vivo* promoter activity due to 1) mutation 25 and mutation 23 and 2) mutation 25 with mutations 22 and 24, there is an indication of a significant contribution of -31T to promoter activity than that of the other 3 positions.

Comparison of mutations 22 through 25 with mutation 14 (-57 G to T) indicates a significant contribution by -57T though not as much as that by -31T. Comparing these mutations with mutation 13 (-56 T to A and -31 C to T) also underlines the previous observation. Comparing these with mutation 8 (-53 A to G), there is an indication that the contribution by -53A is not as significant as those of the other two (-57T and -31T). The observation that mutations 22 through 25 showed His-Mor binding similar to wild-type promoter, but variations in their *in vivo* activity also alludes to the possibility that mutations in these positions do not affect binding by Mor, but affect transcription activation through interaction with the RNAP subunits.

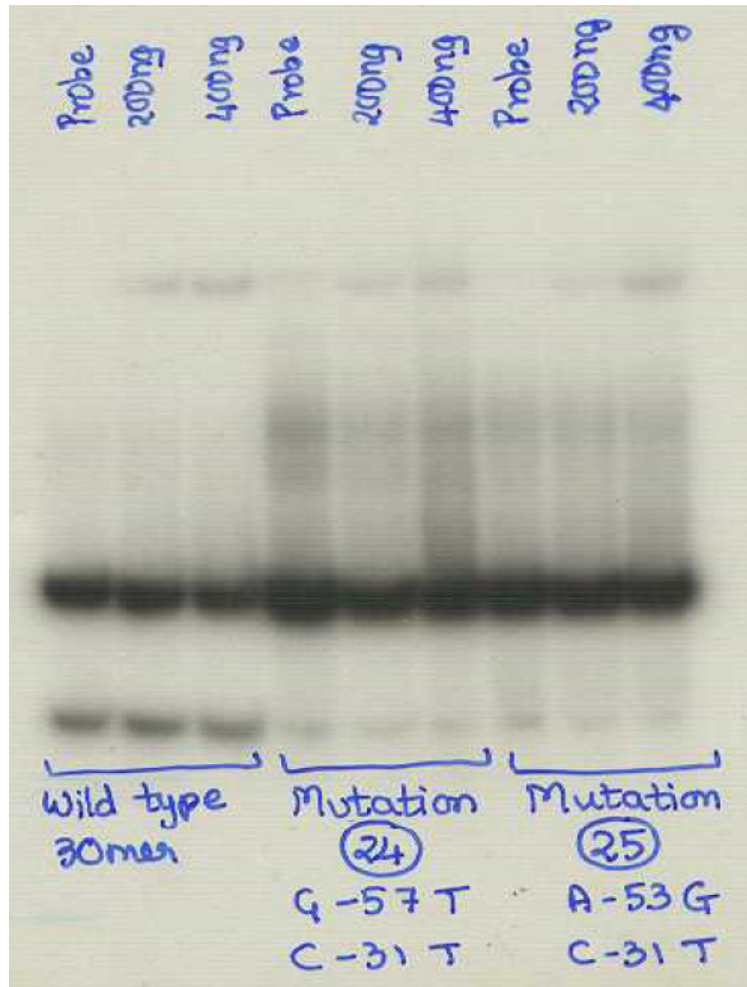


Figure 25: Gel shift for mutation 26.

Binding experiments with mutation 1 and mutation 17 on this gel were done to test the reproducibility of the higher binding of His-Mor to the promoter with these mutations and it was found to be so.

Mutation 26 (-50 G to T): This mutation conferred a red phenotype to the colonies on MacConkey plates indicating a functionally active promoter. This was further supported by its binding to His-Mor, which was just slightly lower to that of wild-type promoter and its *in vivo*  $\beta$ -galactosidase activity of 863, that was also slightly lower than the wild-type promoter activity. This mutation was designed to test the individual contribution by -50T to mutation 17 (-50 G to T and -46 A to C), as mutation 17 gave twice the wild-type binding by His-Mor in a gel shift *in vitro* binding assay. Base change -46C contributes significantly to mutation 17 as seen from binding with mutation 1 (-46 A to C), but contribution by the individual mutation -50 G to T is not significant. It might be that the combination of -50T and -46C, which made the 6-bp segments on either side of the 4-bp spacer (-42 through -45) is what makes it bind so efficiently to His-Mor by highly stabilizing the Mor-P<sub>m</sub> complex. Thus, the specific bases in the region of Mor binding are important rather than the factor of symmetry as indicated by comparing mutations 16 and 17, which both made the segments perfectly symmetrical, though using different substitutions.

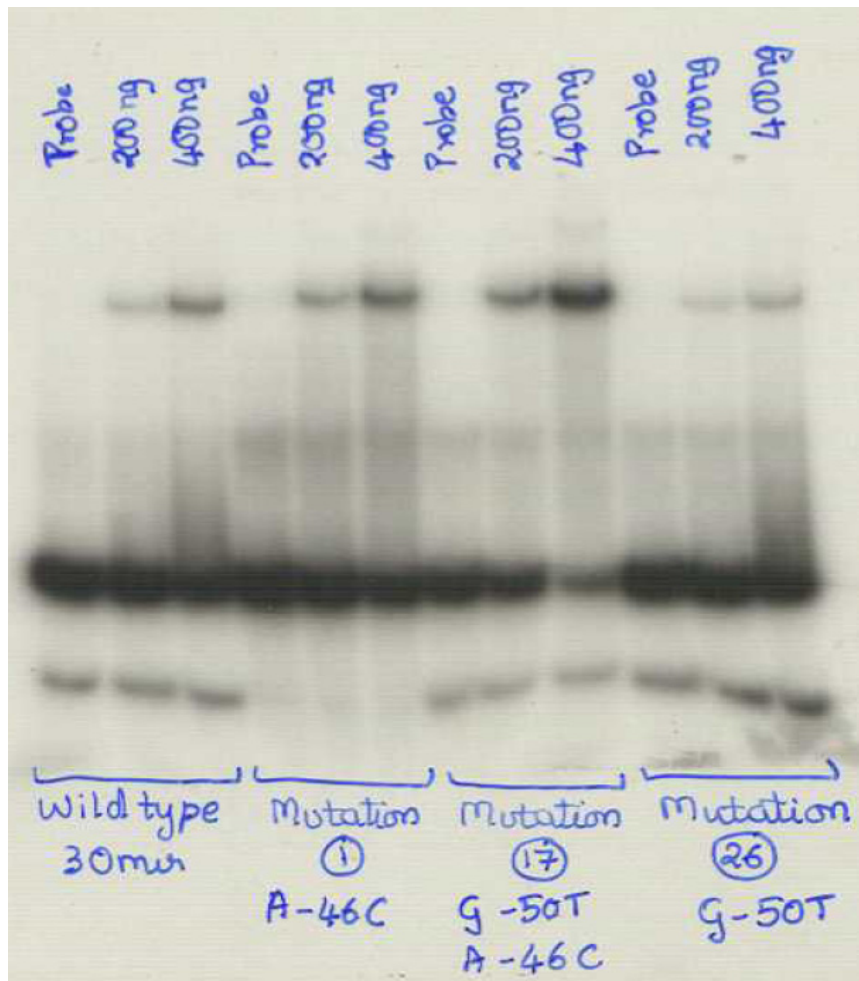


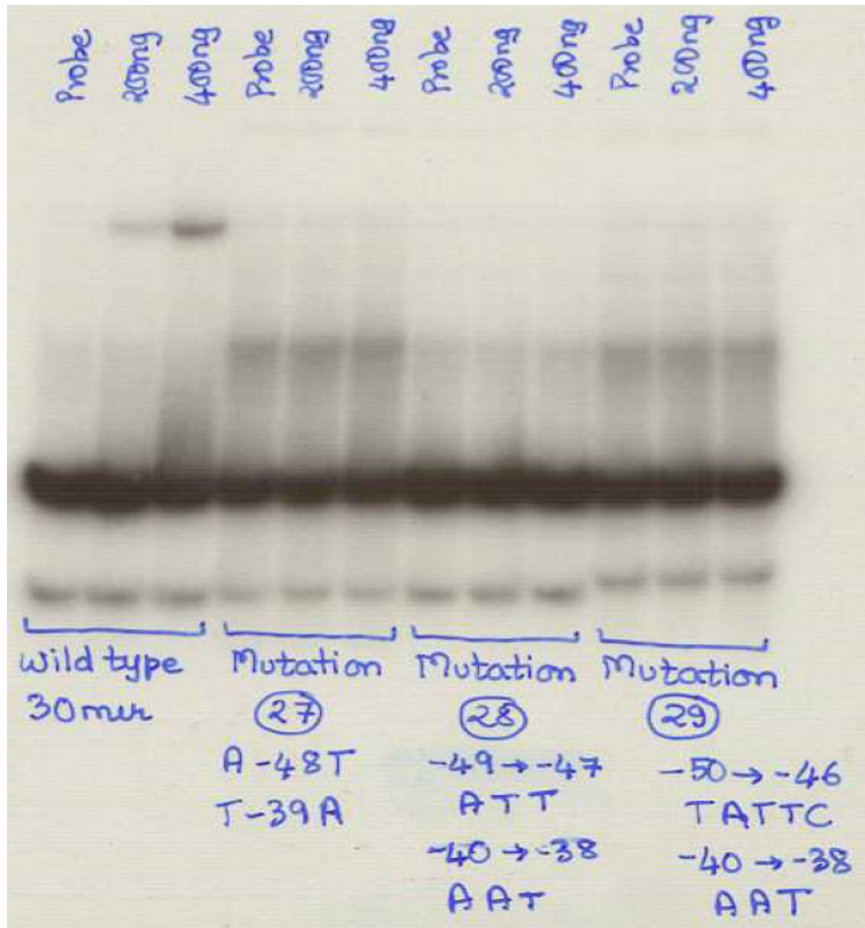
Figure 26: Gel shift for mutations 27, 28 and 29.

Mutation 27 (-48 A to T and -39 T to A): This mutation imparted white color to the colonies on MacConkey plates indicating a severely defective phenotype similar to the mutant with a deleted promoter. His-Mor failed to bind to the promoter DNA with this mutation at levels even as high as 800 ng of protein, as seen from the gel shift data and this explained its almost undetectable  $\beta$ -galactosidase activity of 0.5. It also indicates that the bases are important for Mor recognition and binding and it is not just the symmetry element because even though we maintained the symmetry by swapping the nucleotides between these positions in the already symmetrical wild-type promoter, it failed to bind His-Mor.

Mutation 28 (-49 to -47 ATT and -40 to -38 AAT): This mutation also led to white colonies on MacConkey plates indicating a severely defective phenotype similar to the mutant with a deleted promoter. His-Mor failed to bind to the promoter with this mutation at levels even as high as 800 ng of protein DNA, as seen from the gel shift data and this explained its almost undetectable  $\beta$ -galactosidase activity of 0.8. It also indicates that the bases are important for Mor recognition and binding and it is not just the symmetry element because even though we maintained the symmetry by swapping the segments between these positions in the already symmetrical wild-type promoter, it failed to bind His-Mor. Thus, the bases in these segments seem to be extremely critical for binding with Mor as it does not tolerate any changes as shown by mutations 27, 28 and 3 (-48 A to C and -39 T to G).

Mutation 29 (-50 to -46 TATTC and -40 to -38 AAT): This mutation also led to white colonies on MacConkey plates indicating a severely defective phenotype similar to the mutant with a deleted promoter. His-Mor failed to bind to the promoter DNA with this mutation at levels even as high as 800 ng of protein, as seen from the gel shift data and this explained its almost undetectable  $\beta$ -galactosidase activity of 0.8. This mutation is a combination of mutation 28 (-49 to -47 ATT and -40 to -38 AAT) and mutation 17 (-50 G to T and -46 A to C). Even though mutation 17 bound His-Mor extremely efficiently, mutation 29 failed to do so indicating the significant contribution of the segments -49 to -47 TAA and -40 to -38 TTA to mutation 17. It is not just -50T and -46C of mutation 17 that contributes to its high binding, but it is the effect of the mutations in combination with the other bases in the 6-bp segment on either side of the 4-bp spacer.

The analysis of the mutations 3, 27, 28 and 29 together reveal the critical role played by the specific bases in the Mor-binding region of  $P_m$  which lie as 2 imperfectly inverted symmetrical elements on either side of the 4-bp spacer. They also indicate that it is the presence of the specific bases in these positions that are important for Mor binding and not just the factor of symmetry at these positions. As wild-type His-Mor failed to bind to the promoter with these mutations, it could not activate transcription from that promoter as it could not recruit RNA polymerase to the promoter. These specific bases seem to be critical for the promoter binding and function of His-Mor.



### **The length of the double-stranded oligonucleotides optimum for Mor binding**

Different length oligonucleotides were designed starting with an 18-mer centered at -43.5. Then two nucleotides were added corresponding to their positions in  $P_m$ , one on either side of the 18-mer to obtain other oligonucleotides in the series: 20-mer, 22-mer, 24-mer, 26-mer, 28-mer and 30-mer. The list of the oligonucleotides used for this experiment is given in Table 4. All the bottom strand oligonucleotides were radioactively labeled with  $\gamma\text{-P}^{32}$  ATP, and gel shift reactions were set up with each double-stranded pair, assaying with wild-type His-Mor binding at varying concentrations: 200 ng, 400 ng and 800 ng. This experiment was done to identify the lowest length that would give an optimum gel shift. The rationale behind designing the different length oligonucleotides and using different levels of protein is as follows: there is a possibility that a shorter length oligonucleotide which would fail to bind detectably to His-Mor at lower concentrations, could bind His-Mor at higher protein concentrations, giving a detectable shift in the assay. In contrast, it would help us determine the lowest concentration of protein which would give an optimum detectable shift with longer oligonucleotides. This can also be used as an indicator to determine the most critical bases that define Mor-promoter DNA interactions and the minimum length of the oligonucleotide needed to produce a detectable shift in the assay.

The pictures of the gels of this experiment are given in Figure 27. The following observations were made from this experiment:

1. His-Mor failed to shift the 16-mer oligonucleotide duplex, probably because its length is not sufficient for protein binding or that the double-stranded DNA complexes and/or the His-Mor DNA complexes are unstable, hence undetectable at experimental conditions.
2. All other length oligonucleotides bound His-Mor at all three protein levels used, with minor differences.
3. At the levels of His-Mor protein used, increasing amounts of the protein gave increasing intensities of bound probe, giving a gradually increasing gradient from 200 ng to 400 ng to 800 ng.
4. All the probes from at least a 22-mer and increasing to the 30-mer showed similarity in their binding to the protein with no observable differences.
5. The capacity of binding of the 18-mer probe to the protein was lower than all other size oligonucleotides at the three levels of protein used, indicating that even though the binding of His-Mor to the 18-mer is detectable, this interaction is probably more stable and/or more detectable in longer oligonucleotides.

In order to study the effect of mutations in the context of a longer stretch of  $P_m$ , the 30-mer oligonucleotide was used for the study of mutations from -57 to -30 of  $P_m$ .

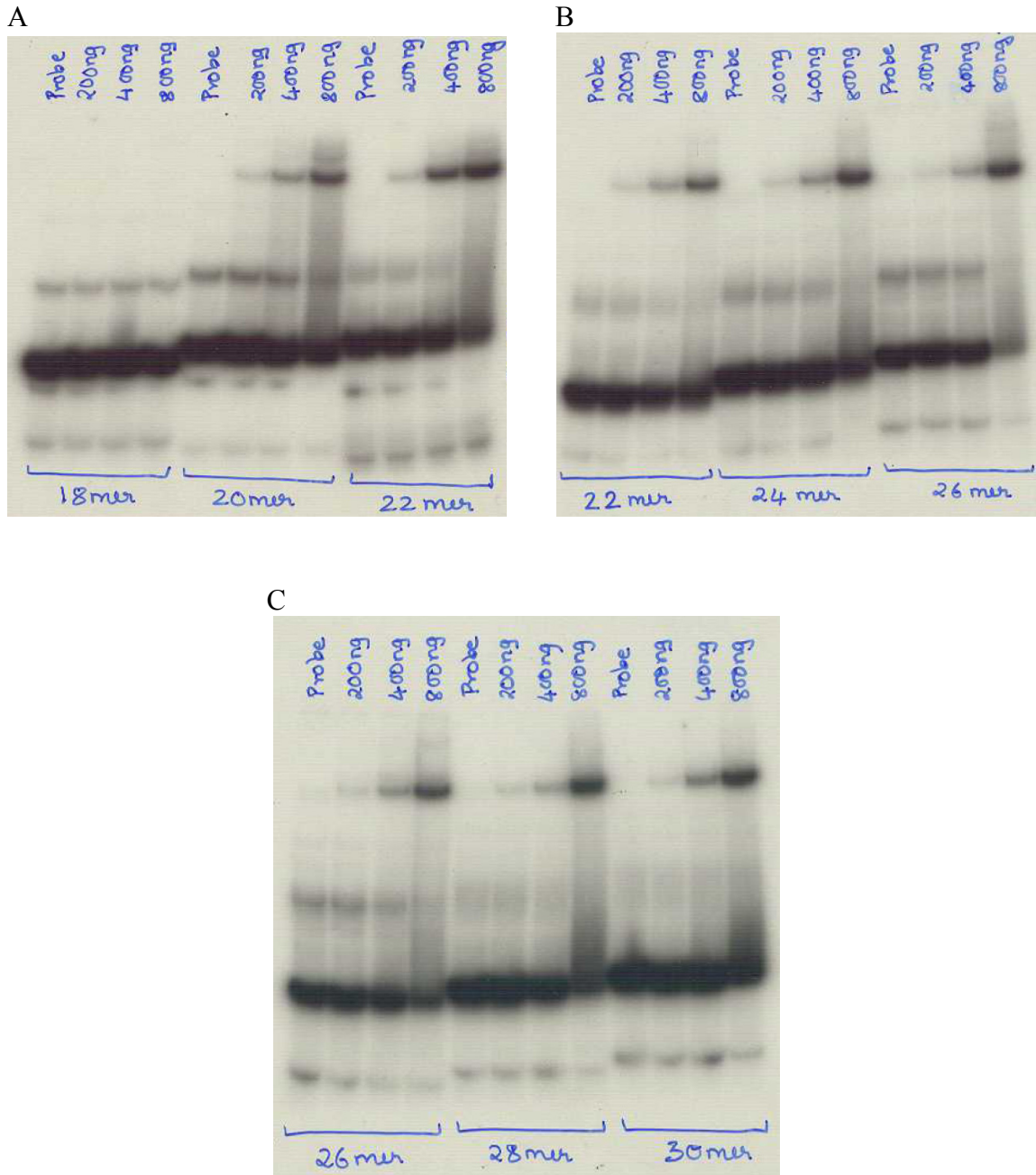


Figure 27: Gel shifts with different size oligonucleotides.

A Gel shift with 18, 20 and 22-mer oligonucleotides

B Gel shift with 22, 24 and 26-mer oligonucleotides

C Gel shift with 26, 28 and 30-mer oligonucleotides



### **Test whether Mor binding stabilizes the duplex oligonucleotide**

For this experiment, the combined mutation -50T and -46C that gave the highest binding to Mor were used. The hypothesis was that if this interaction between Mor and DNA stabilizes the duplex, then we would get detectable and good shift in the assay with shorter length oligonucleotides and/or at lower reaction times. For this a 16-mer, 18-mer and 20-mer with the combined mutations -50T and -46C within the promoter,  $P_m$  were used with wild-type His-Mor at 200 ng and 400 ng amounts and the reaction was set up at three time points: 10, 20 and 30 min.

The pictures of the gels of this experiment are given in Figure 28. The following observations were made from this experiment:

1. The 16-mer and 18-mer did not show detectable His-Mor binding at both the protein levels and at all the time points.
2. The binding of His-Mor to the 20-mer at 10 min reaction time was barely detectable.
3. At 400 ng of protein, the 20-mer was detectably bound after 20min reaction time and the level of binding increased at the 30 min time point.
4. The same trend was observed with 200 ng of protein as with 400 ng though at lower intensities.
5. The control 30-mer bound very efficiently to His-Mor even at 200 ng protein level and as early as 10 min incubation time, and the amount of bound probe increased with an increase in the reaction incubation time.

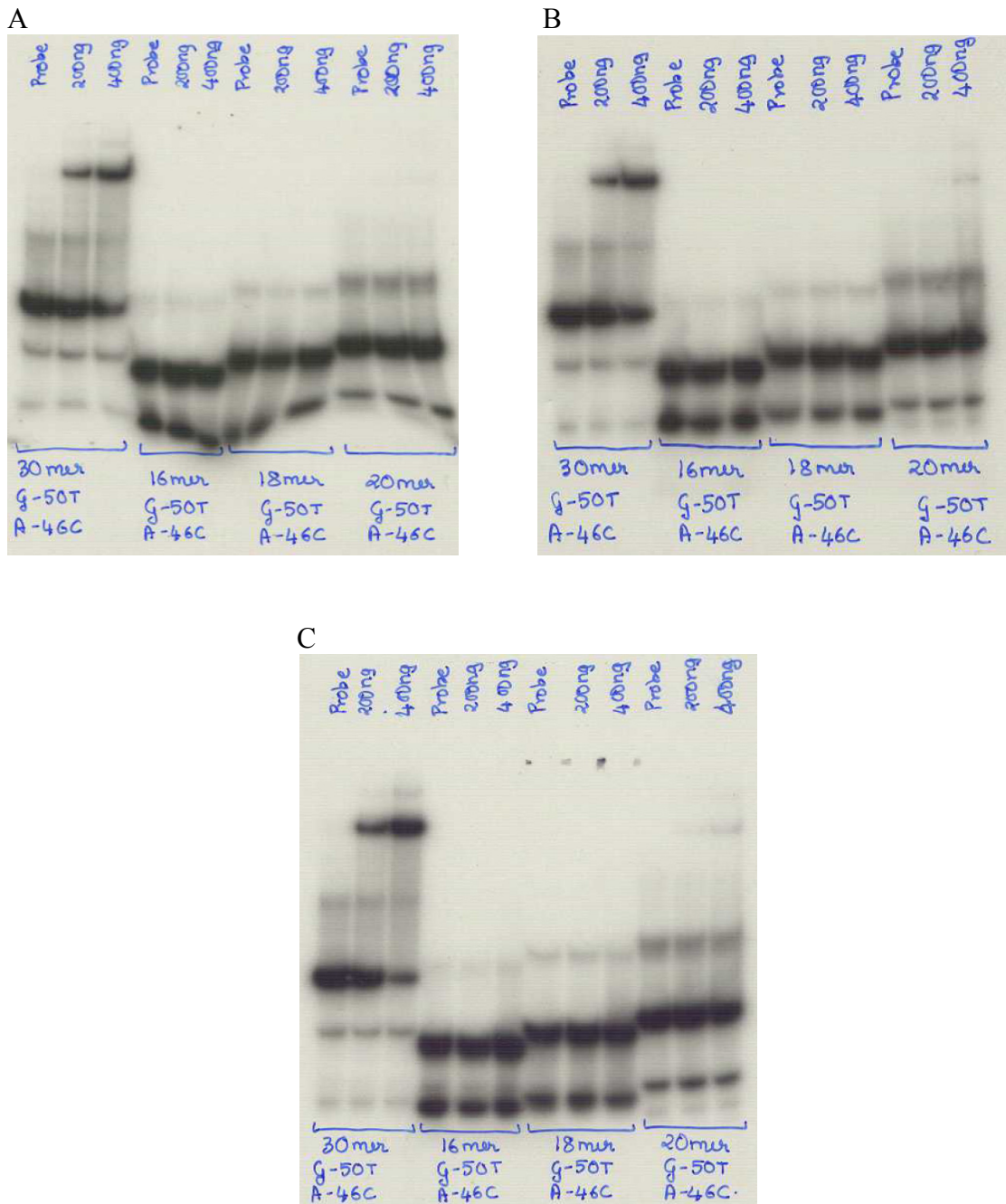


Figure 28: Gel shift to test Mor-oligo duplex stability.

## Chapter 3: Summarized results and discussion

The results of plate phenotyping, *in vivo* transactivation assays and *in vitro* binding assays for all the specific mutations are presented in Table 7. For the convenience of understanding, positions -30 to -57 of P<sub>m</sub> promoter is divided into three regions: the 4-bp minor groove spacer region, the 6-bp dyad-symmetry sequences and the region flanking the Mor-binding sequences of P<sub>m</sub>.

### Results of the spacer region mutations (Figure 29)

About 60 mutants isolated with mutations in the spacer region gave red colonies, indicating a wild-type phenotype. Functionally these can be classified as wild-type or at least close-to-wild-type. Two GC-rich mutants were chosen for further analysis. GGGG gave red colonies while CCGG had given an almost wild-type phenotype. In addition, two more AT-rich mutants were isolated by targeted mutagenesis and analyzed further. The AAAA mutation gave colonies which were red with a small white halo while the TAAT mutation imparted almost a red-centered phenotype to the colonies. The  $\beta$ -galactosidase activities recorded for these mutants correlated with their phenotypes. The red one had almost wild-type activity; the reds with a small white halo gave less than wild-type activity and the red centered one had less than half of the wild-type activity. None of the 60 isolated mutants imparted a severely defective phenotype or completely abolished Mor binding. However, a clear gradation was seen with respect to Mor binding and *in vivo* activity in the four mutations tested for them. This indicates that even though the specificity of the bases in the 4-bp spacer region 5' GCCG 3' is not stringent, the effect of mutagenesis of this region is detectable. Comparing this with the effect of mutagenesis in the bases of the major groove of the Mor-binding site, it can be stated that even though the minor groove bases do contribute to facilitate Mor binding, they are not as important as the major groove bases. Thus, major groove interactions of Mor with DNA play a major role at the promoter and the minor groove has a minor contribution.

### Results of the mutations in the Mor-binding segments of the promoter (Figure 30)

The bases TAA in the positions -49, -48 and -47 and their symmetrical positions occupied by bases TTA in positions -40, -39 and -38 are very interesting in the fact that they do not tolerate any base changes. These positions are perfectly symmetrical in the natural promoter; the bases and perhaps the symmetry in these positions seem to play a major role in strengthening Mor-DNA interactions. This was supported by the observation that any mutations in these positions exhibited a severe defect in Mor binding and hence activation of P<sub>m</sub>.

In this study, no mutagenesis was attempted in the positions -51 and -36 of P<sub>m</sub> as any mutations in these positions were previously found to be severely defective in activation by Mor (Artsimovitch, 1996).

Table 7: Summarized results of mutants: phenotypes,  $\beta$ -gal values and gel shift assays.

| <b>Mutation number</b> | <b>Mutation</b>                             | <b>Phenotype</b>                               | <b><math>\beta</math>-gal value</b> | <b>Gel shift assay</b> |
|------------------------|---------------------------------------------|------------------------------------------------|-------------------------------------|------------------------|
| 1                      | -46 A to C                                  | Red with small white halo                      | 518                                 | +++++                  |
| 2                      | -41 G to T                                  | Red with small white halo                      | 377                                 | ++                     |
| 3                      | -48 A to C and<br>-39 T to G                | White                                          | 0.4                                 | -                      |
| 4                      | -37 A to C                                  | Red                                            | 951                                 | +++++                  |
| 5                      | -52 C to G                                  | Red with small white halo                      | 639                                 | ++++                   |
| 6                      | -52 C to A and<br>-35 C to T                | Red with small white halo                      | 597                                 | +++                    |
| 7                      | -34 C to T                                  | Red                                            | 809                                 | ++++                   |
| 8                      | -53 A to G                                  | Red with small white halo                      | 630                                 | ++++                   |
| 9                      | -33 G to T                                  | Red with small white halo                      | 654                                 | ++++                   |
| 10                     | -54 A to C                                  | Red with larger white halo than others         | 335                                 | ++++                   |
| 11                     | -32 G to T                                  | Red                                            | 1133                                | ++++                   |
| 12                     | -31 C to A                                  | Red                                            | 1139                                | ++++                   |
| 13                     | -56 T to A and<br>-31 C to T                | Red                                            | 1272                                | +++++                  |
| 14                     | -57 G to T                                  | Red                                            | 902                                 | +++++                  |
| 15                     | -57 G to T and<br>-30 T to A                | Red with small white halo                      | 647                                 | +++                    |
| 16                     | -50 G to T and<br>-41 G to T                | Red with small white halo                      | 421                                 | +++                    |
| 17                     | -50 G to T and<br>-46 A to C                | Red with small white halo                      | 671                                 | +++++                  |
| 18                     | -42 through<br>-45 AAAA                     | Red with small white halo                      | 539                                 | ++                     |
| 19                     | -42 through<br>-45 TAAT                     | Red with large white halo, almost red centered | 344                                 | ++                     |
| 20                     | -42 through<br>-45 GGCC                     | Red with small white halo                      | 618                                 | +++                    |
| 21                     | -42 through<br>-45 GGGG                     | Red                                            | 768                                 | +++++                  |
| 22                     | -57 G to T, -53<br>A to G and -31<br>C to T | Red                                            | 1021                                | +++++                  |

Table 7 (continued).

| <b>Mutation number</b>                         | <b>Mutation</b>                              | <b>Phenotype</b> | <b><math>\beta</math>-gal value</b> | <b>Gel shift assay</b> |
|------------------------------------------------|----------------------------------------------|------------------|-------------------------------------|------------------------|
| 23                                             | -57 G to T and<br>-53 A to G                 | Red              | 642                                 | +++++                  |
| 24                                             | -57 G to T and<br>-31 C to T                 | Red              | 1488                                | +++++                  |
| 25                                             | -53 A to G and<br>-31 C to T                 | Red              | 816                                 | +++++                  |
| 26                                             | -50 G to T                                   | Red              | 863                                 | +++                    |
| 27                                             | -48 A to T and<br>-39 T to A                 | White            | 0.5                                 | -                      |
| 28                                             | -49 to -47<br>ATT and<br>-40 to -38<br>AAT   | White            | 0.8                                 | -                      |
| 29                                             | -50 to -46<br>TATTC and<br>-40 to -38<br>AAT | White            | 0.8                                 | -                      |
| Positive Control with wild-type P <sub>m</sub> | No mutation                                  | Red              | 1000                                | +++++                  |
| Negative control without P <sub>m</sub>        | Deletion of P <sub>m</sub>                   | White            | 5                                   | -                      |

$\beta$ -gal Value standardized to the wild-type set as 1000.

+++++ is wild-type binding.

++++ is 75 to 90% of wild-type.

+++ is 50 to 75% of wild-type.

++ is 30 to 50% of wild-type.

+++++ is very high binding almost 200% of wild-type binding.

- is no detectable binding similar to the negative control.



Base changes are also not tolerated at -41G, as any mutation at this position gave weaker Mor binding and hence lower activation of  $P_m$ .

Base changes at the symmetrical positions -37A and -50G did not affect Mor binding and promoter activity. These positions lie just within the imperfect dyad symmetry, and they do not seem to play as significant a role in Mor-promoter interactions as the other bases of the dyad symmetrical elements.

Positions -46A and -41G do contribute to Mor binding; mutation at -46 gave a significantly higher Mor binding, while mutation at -41 negatively affected Mor binding. These positions lie just outside the 4-bp spacer region and just within the dyad symmetrical elements of the Mor monomer binding sites.

Mutations -50T and -46C together and -46C alone exhibited a remarkable binding capacity to Mor, emphasizing the importance of these positions. This observation may be useful while trying to crystallize Mor-DNA binary complexes to understand structural aspects of their interactions. Surprisingly, this mutation gave a lower level of promoter activity. Since Mor is proposed to recruit RNAP to the promoter through interactions with the  $\alpha$ -CTD and  $\sigma 70$  (Artsimovitch and Howe, 1996; Ma and Howe, 2004), strengthening of Mor-DNA interactions may strengthen Mor-RNAP interactions; this may negatively affect release of  $\sigma 70$  during promoter clearance. Thus, the interactions between Mor and DNA should be optimum not only to recruit RNAP to the promoter but also for the subsequent release of RNAP subunits to facilitate transcription. This fact leads to the hypothesis that Mor binds the promoter to direct and recruit RNAP to the promoter and subsequently has to be released from the DNA as transcription begins.

A majority of the activators influence the recruitment of RNAP in the transcription initiation. The activators either increase RNAP binding to the promoter or, when it is already bound, influence isomerization from the closed to open promoter complex. For example, CAP increases the initial binding of RNAP at the *lac* promoter, and  $\lambda$ cI protein positively influences the rate of isomerization at the  $\lambda P_{RM}$  promoter. A few activators have also been shown to act at a post-recruitment step of transcription initiation, as observed in the case of Arc, which enhances promoter clearance from the  $P_{ant}$  promoter of bacteriophage P22 during late lytic growth. A given activator can also influence different steps of transcription initiation in different promoters. Most of the CAP-dependent promoters are regulated at the RNAP recruitment step; however, at the *malT* promoter, CAP enhances promoter clearance. Using the CAP and mutant *lac* promoter, it has been demonstrated that regulatory proteins could act as activators or repressors in different steps of the transcription initiation pathway, depending on the energetic differences of the intermediate complexes. In the vast repertoire of transcription activators, there are few that act at multiple steps of transcription initiation at a single promoter. The most well known example is CAP-mediated activation of the initial binding of RNAP and subsequent isomerization at the *gal P1* promoter. Recruitment of RNAP and subsequent promoter clearance at the  $\lambda P_R$  promoter are enhanced by DnaA. Fis has been reported to sequentially stimulate transcription initiation steps at the *tyrT* promoter by facilitating initial binding of RNAP, unwinding of DNA at the transcription

start point, and subsequent promoter clearance. At the  $P_{RE}$  promoter of the  $\lambda$ phage, cII protein enhances both the formation and stability of the RNAP-promoter open complex. Transactivator C protein of phage Mu acting at the late promoter  $P_{mom}$  has also been indicated to belong to a small group of activators acting at multiple steps of promoter-polymerase interactions in a single promoter (Chakraborty and Nagaraja, 2006). Thus, based on the above observation, Mor could belong to the group of transcriptional activators affecting multiple steps of promoter-polymerase interactions at promoter  $P_m$  including its effect on promoter clearance.

### **Results of the mutations in the region beyond Mor-binding segments of the promoter (Figures 31, 32 and 33)**

Positions -52 and -35 of  $P_m$  do contribute to Mor binding as well as promoter activity, indicating that these positions have a contribution to Mor binding as they lie just outside the Mor-binding sites on either side.

Positions -57 through -53 which lie upstream of the Mor-binding site and positions -34 through -30 which are downstream of the Mor-binding site do not have a dramatic effect on Mor binding, but they do affect promoter activity. The observation that mutations in this region affect promoter activity may reflect a contribution of these regions in atypical interaction with the subunits of RNAP.

The region -29 to -33 was previously predicted to overlap the region that undergoes a structural change in the form of a distortion which confers flexibility to the DNA and is probably necessary for the precise alignment of Mor and RNAP. The fact that this region may occur in the Mor-RNAP interface also raises the possibility of a contribution of this region to promoter activity by affecting either Mor-RNAP interactions or RNAP-DNA interactions (Artsimovitch and Howe, 1996).

Bases -57 to -54 lie in the region that was previously predicted to be the UP-like element on the promoter distal side of the Mor-binding site (Ma and Howe, 2004). The results of mutagenesis of this region is consistent with the prediction that one of the  $\alpha$ -CTD subunits of RNAP interacts with this region, and the effect of mutations in this region to promoter activity is probably due to this interaction, which is either strengthened or weakened depending on the type of base changes in these positions (Ma, 2004). The observation that positions -53 and -54 contribute more than positions -56 and -57 is also consistent with the previous observations (Ma, 2004).

Thus, region -57 to -30 of  $P_m$  is important for activity of the promoter either directly or indirectly (Figure 34). The direct effect of the spacer region and the imperfect dyad symmetrical element and the bases just flanking them is revealed from Mor-binding experiments. The indirect contribution of the positions beyond the Mor-binding site on either side can be hypothesized from activity assays and previous observations and predictions for their involvement in interactions with the RNAP subunits or their predicted effects on conformational changes in this region.







### General conclusions from P<sub>m</sub> mutagenesis

|        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 5b     |        | 4b     |        |        |        | 2      |        |        |        | 1      |        |        |        | 2      |        |        |        | 4a     |        |        |        | 5a     |        |        |        |        |        |
| 5<br>7 | 5<br>6 | 5<br>5 | 5<br>4 | 5<br>3 | 5<br>2 | 5<br>1 | 5<br>0 | 4<br>9 | 4<br>8 | 4<br>7 | 4<br>6 | 4<br>5 | 4<br>4 | 4<br>3 | 4<br>2 | 4<br>1 | 4<br>0 | 3<br>9 | 3<br>8 | 3<br>7 | 3<br>6 | 3<br>5 | 3<br>4 | 3<br>3 | 3<br>2 | 3<br>1 | 3<br>0 |
| G      | T      | A      | A      | A      | C      | A      | G      | T      | A      | A      | A      | G      | C      | C      | G      | G      | T      | T      | A      | A      | T      | C      | C      | G      | G      | C      | T      |
|        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
|        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
|        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |

1. Contributes to Mor binding in the minor groove spacer region, but not as critical as major groove bases.
2. Extremely critical for Mor binding in the major groove, no mutations are tolerated.
3. Critical major groove bases for interaction with Mor; coincides with the dyad symmetry Mor-binding segments.
4. Some contribution to Mor binding, and some to interaction with RNAP
  - a) probably with  $\sigma$  subunit of RNAP
  - b) probably with  $\alpha$ -CTD of RNAP
5. Contributes to interaction with RNAP subunits
  - a) probably with  $\sigma$  subunit of RNAP
  - b) probably with  $\alpha$ -CTD of RNAP

Figure 34: General conclusions from mutagenesis of the middle promoter.

The regions are color coded with the font colors corresponding to the colors of the segments marked. Bases from -30 to -57 upstream of the transcription start site are given with their numbers given above the bases.

## Major findings from the current work

The findings describe the impact of this work on our understanding of Mor- $P_m$  interactions.

1. Demonstrated some contribution of the minor groove spacer region to Mor binding.
2. Identified positions extremely critical for Mor interactions at the promoter.
3. Identified positions that have a modest contribution to interaction with activator Mor.
4. Predicted positions of the promoter that affect transactivation. This could be through direct interaction with RNAP subunits or through indirect conformational changes.
5. Higher affinity binding of Mor to promoter correlates with reduced promoter activity probably due to ineffective promoter clearance. This indicates a dual role for activator Mor at the promoter: recruitment of RNAP to the promoter and release of core RNAP during promoter clearance and transcription initiation.

## Future directions

The following possible experiments are proposed to dissect the protein-protein and protein-DNA interactions between Mor,  $P_m$ , and the different subunits of RNA polymerase and to define their respective roles in middle promoter regulation.

1. To identify amino acid positions of Mor that interact with regions of promoter.
2. To investigate and further understand interactions between RNAP subunit and promoter positions.
3. To find out whether positions beyond Mor-binding region directly contribute to RNAP subunit interactions or indirectly contribute by affecting conformational changes at Mor-DNA-RNAP interfaces.
4. In most cases, protein-protein contacts between the activator and different subunits of RNAP, mainly the  $\alpha$ -CTD, are the basis for transcription activation (Dove et al, 2000). The role of protein-protein contacts between Mor- $\alpha$ -CTD and Mor- $\sigma$ -CTD in the efficient recruitment of RNAP to  $P_m$  and in Mor-dependent  $P_m$  activation is well established (Artsimovitch et al, 1996). A structure-guided targeting of the side chains of specific residues of Mor can be carried out to identify the RNAP contact sites in Mor. As indicated by structural homology, Mor binds to the major groove using primarily the residues from the tip of the recognition helix,  $\alpha 5$ . When Mor binds to DNA in an “ends-on mode”, two regions of the protein would be in close proximity to the C-terminal domains of  $\alpha$  and  $\sigma$  of RNAP. The amino acid side chains on the external face of preceding helix  $\alpha 4$ . In ends-on mode of base recognition, the body of the preceding helix is left out of the major groove and solvent-exposed. Contacts involving minimal protein-protein interface is not uncommon in activator-RNAP contacts. Another characteristic feature of ends-on DNA-binding is that most of the C-terminal part

of recognition helix sticks out of the major groove and away from DNA. The basic-rich C-terminal tail of Mor presents an ideally located surface for electrostatic interactions between Mor and different subunits of RNAP. The frequency of the defective phenotypes in the mutant library might not be as drastic as it is seen in case of residues that are predicted to be involved in base-specific interactions. So isolation of phenotypically selected mutations and a detailed functional characterization of the mutant proteins would be necessary to define their role in protein-protein contacts, thereby transcription activation.

5. To test the prediction between inverse correlation between affinity of Mor binding and effective promoter clearance.
6. UV-photo cross-linking and mass spectrometry: Identification of the specific contacts between Mor with middle promoter is essential to understand the basis for the promoter recognition. Interaction between DNA-binding proteins and their binding site generally involves a combination of base-specific interactions and non-specific interactions with phosphate backbone and sugar. The sequence-specific DNA-binding proteins, such as Mor, discriminate their binding site from a vast majority of random sequences by specific interactions, whereas the non-specific interactions provide the stability to the binary complex. UV-photo cross-linking in tandem with mass spectrometry have emerged as a useful tool in the identification of specific interactions. The technique generally involves UV cross-linking of DNA with modified bases to the side chains of amino acids it interacts with. UV irradiation of derivatized nucleobases such as 5'-iodouracil generates highly reactive intermediates that form zero-length cross links to the amino acid side chains in the vicinity. This covalently cross-linked heteroconjugates can be subjected to protease digestion and/or nuclease digestion to reduce the size of the peptide or nucleotide fragments. The purified fragments can be analyzed on a Matrix-Assisted Laser Desorption/Ionisation Time-Of-Flight Mass Spectrometry (MALDI-TOF MS).
7. Findings can be used to optimize the positions of the promoter with respect to those that affect Mor-binding and  $\alpha$ -CTD-binding for crystallographic studies on
  - (a) Mor-DNA binary complex
  - (b) Mor-DNA- $\alpha$ -CTD ternary complex.

## List of References

Access Excellence @ the National Health. <http://www.accessexcellence.org/RC/VL/GG/central.php>. Accessed June 18, 2008.

Adelman K, Yuzenkova J, La Porta A, Zenkin N, Lee J, Lis JT, Borukhov S, Wang MD, Severinov K. Molecular mechanism of transcription inhibition by peptide antibiotic microcin J25. *Mol Cell*. 2004, 14(6): 753-62.

Aiyar SE, Gourse RL, Ross W. Upstream A-tracts increase bacterial promoter activity through interactions with the RNA polymerase alpha subunit. *Proc Natl Acad Sci USA*. 1998, 95(25): 14652-7.

Artsimovitch I. Activation of Middle Transcription of Phage Mu. 1996. University of Tennessee Health Science Center, Memphis.

Artsimovitch I, Howe M. Transcription activation by the bacteriophage Mu Mor protein: analysis of promoter mutations in P<sub>m</sub> identifies a new region required for promoter function. *Nucl Acids Res*. 1996, 24(3): 450-7.

Artsimovitch I, Kahmeyer-Gabbe M, Howe MM. Distortion in the spacer region of P<sub>m</sub> during activation of middle transcription of phage Mu. *Proc Natl Acad Sci USA*. 1996, 93(18): 9408-13.

Artsimovitch I, Murakami K, Ishihama A, Howe MM. Transcription activation by the bacteriophage Mu Mor protein requires the C-terminal regions of both alpha and sigma70 subunits of *Escherichia coli* RNA polymerase. *J Biol Chem*. 1996, 271(50): 32343-8.

Artsimovitch I, Vassylyev DG. Is it easy to stop RNA polymerase? *Cell Cycle*. 2006, 5(4): 399-404. Review.

Artsimovitch I, Vassylyeva MN, Svetlov D, Svetlov V, Perederina A, Igarashi N, Matsugaki N, Wakatsuki S, Tahirov TH, Vassylyev DG. Allosteric modulation of the RNA polymerase catalytic reaction is an essential component of transcription control by rifamycins. *Cell*. 2005, 122(3): 351-63.

Au TK, Agrawal P, Harshey RM. Chromosomal integration mechanism of infecting Mu virion DNA. *J Bacteriol*. 2006, 188(5): 1829-34.

Banerjee S, Chalissery J, Bandey I, Sen R. Rho-dependent transcription termination: more questions than answers. *J Microbiol*. 2006, 44(1): 11-22. Review.

Barne KA, Bown JA, Busby SJ, Minchin SD. Region 2.5 of the *Escherichia coli* RNA polymerase sigma70 subunit is responsible for the recognition of the 'extended-10' motif at promoters. *EMBO J*. 1997, 16(13): 4034-40.

- Bétermier M, Poquet I, Alazard R, Chandler M. Involvement of *Escherichia coli* FIS protein in maintenance of bacteriophage Mu lysogeny by the repressor: control of early transcription and inhibition of transposition. *J Bacteriol.* 1993, 175(12): 3798-811.
- Bewley C, Gronenborn A, Clore G. Minor groove-binding architectural proteins: structure, function and DNA recognition. *Annu Rev Biophys Biomol Struct.* 1998, 27: 105-31.
- Borukhov S, Lee J, Laptenko O. Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol Microbiol.* 2005, 55(5): 1315-24. Review.
- Borukhov S, Nudler E. RNA polymerase holoenzyme: structure, function and biological implications. *Curr Opin Microbiol.* 2003, 6(2): 93-100. Review.
- Borukhov S, Nudler E. RNA polymerase: the vehicle of transcription. *Trends Microbiol.* 2008, 16(3): 126-34. Review.
- Borukhov S, Severinov K. Role of the RNA polymerase sigma subunit in transcription initiation. *Res Microbiol.* 2002, 153(9): 557-62. Review.
- Bowers CW, Dombroski AJ. A mutation in region 1.1 of sigma70 affects promoter DNA binding by *E. coli* RNA polymerase holoenzyme. *EMBO J.* 1999, 18(3): 709-16.
- Browning DF, Busby SJ. The regulation of bacterial transcription initiation. *Nat Rev Microbiol.* 2004, 2(1): 57-65. Review.
- Busby S, Ebright RH. Promoter structure, promoter recognition and transcription activation in prokaryotes. *Cell.* 1994, 79(5): 743-6.
- Callaci S, Heyduk E, Heyduk T. Core RNA polymerase from *E. coli* induces a major change in the domain arrangement of sigma70 subunit. *Mol Cell.* 1999, 3: 229-38.
- Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A, Darst SA. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell.* 2001, 104(6): 901-12.
- Campbell EA, Masuda S, Sun JL, Muzzin O, Olson CA, Wang S, Darst SA. Crystal structure of the *Bacillus stearothermophilus* anti-sigma factor SpoIIAB with the sporulation sigma factor sigmaF. *Cell.* 2002, 108(6): 795-807.
- Campbell EA, Muzzin O, Chlenov M, Sun JL, Olson CA, Weinman O, Trester-Zedlitz ML, Darst SA. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell.* 2002, 9(3): 527-39.
- Campbell EA, Pavlova O, Zenkin N, Leon F, Irschik H, Jansen R, Severinov K, Darst SA. Structural, functional and genetic analysis of sorangicin inhibition of bacterial RNA

polymerase. EMBO J. 2005, 24(4): 674-82.

Campbell EA, Tupy JL, Gruber TM, Wang S, Sharp MM, Gross CA, Darst SA. Crystal structure of *Escherichia coli* sigmaE with the cytoplasmic domain of its anti-sigma RseA. Mol Cell. 2003, 11(4): 1067-78.

Campbell EA, Westblade LF, Darst SA. Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. Curr Opin Microbiol. 2008, 11(2): 121-7.

Carpousis AJ, Gralla JD. Interaction of RNA polymerase with lacUV5 promoter DNA during mRNA initiation and elongation. Footprinting, methylation and rifampicin-sensitivity changes accompanying transcription initiation. J Mol Biol. 1985, 183(2): 165-77.

Caruthers MH, Beaucage SL, Becker C, Efcavitch JW, Fisher EF, Galluppi G, Goldman R, deHaseth P, Matteucci M, McBride L. Deoxyoligonucleotide synthesis via the phosphoramidite method. Gene Amplif Anal. 1983, 3: 1-26.

Chaconas G, Harshey RM, Sarvetnick N, Bukhari AI. Predominant end-products of prophage Mu DNA transposition during the lytic cycle are replicon fusions. J Mol Biol. 1981, 150(3): 341-59.

Chaconas G, Kennedy DL, Evans D. Predominant integration end products of infecting bacteriophage Mu DNA are simple insertions with no preference for integration of either Mu DNA strand. Virology. 1983, 128(1): 48-59.

Chakraborty A, Nagaraja V. Dual role for transactivator protein C in activation of *mom* promoter of bacteriophage Mu. J Biol Chem. 2006, 281(13): 8511-7.

Chiang LW, Howe MM. Mutational analysis of a C-dependent late promoter of bacteriophage Mu. Genetics. 1993, 135(3): 619-29.

Chopra I. Bacterial RNA polymerase: a promising target for the discovery of new antimicrobial agents. Curr Opin Investig Drugs. 2007, 8(8): 600-7. Review.

Ciampi MS. Rho-dependent terminators and transcription termination. Microbiology. 2006, 152(9): 2515-28. Review.

Collado-Vides J, Magasanik B, Gralla JD. Control site location and transcriptional regulation in *Escherichia coli*. Microbiol Rev. 1991, 55(3): 371-94. Review.

Cramer P. Common structural features of nucleic acid polymerases. Bioessays. 2002, 24(8): 724-9.

Cramer P. Multisubunit RNA polymerases. Curr Opin Struct Biol. 2002, 12(1): 89-97.



- deHaseth PL, Zupancic ML, Record MT Jr. RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J Bacteriol.* 1998, 180(12): 3019-25.
- Dombroski A, Walter W, Gross C. Amino-terminal amino acids modulate sigma-factor DNA-binding activity. *Genes Dev.* 1993, 12: 2446-55.
- Dove SL, Darst SA, Hochschild A. Region 4 of  $\sigma$  as a target for transcription regulation. *Mol Microbiol.* 2003, 48(4): 863-74.
- Dove SL, Huang FW, Hochschild A. Mechanism for a transcriptional activator that works at the isomerization step. *Proc Natl Acad Sci USA.* 2000, 97(24): 13215-20.
- Ebright RH. RNA Polymerase: Structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol.* 2000, 304(5): 687-98.
- Ebright RH, Busby S. The *Escherichia coli* RNA polymerase [alpha] subunit: structure and function. *Curr Opin Gen & Dev.* 1995, 5(2): 197-203.
- Erie DA. The many conformational states of RNA polymerase elongation complexes and their roles in the regulation of transcription. *Biochim Biophys Acta.* 2002, 1577(2): 224-39. Review.
- Estrem ST, Ross W, Gaal T, Chen ZW, Niu W, Ebright RH, Gourse RL. Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes Dev.* 1999, 13(16): 2134-47.
- Faellen M, Toussaint A. Isolation of conditional defective mutants of temperate phage Mu-1 and deletion mapping of the Mu-1 prophage. *Virology.* 1973, 54(1): 117-24.
- Gabrielian A, Pongor S. Correlation of intrinsic DNA curvature with DNA property periodicity. *FEBS Lett.* 1996, 393(1): 65-8.
- Gezvain K, Gruber TM, Mooney RA, Gross CA, Landick R. A hydrophobic patch on the flap-tip helix of *E. coli* RNA polymerase mediates sigma (70) region 4 function. *J Mol Biol.* 2004, 343(3): 569-87.
- Giphart-Gassler M, Wijffelman C, Reeve J. Structural polypeptides and products of late genes of bacteriophage Mu: characterization and functional aspects. *J Mol Biol.* 1981, 145(1): 139-63.
- Gloor G, Chaconas G. Sequence of bacteriophage Mu N and P genes. *Nucl Acids Res.* 1988, 16(11): 5211-2.
- Gloor G, Chaconas G. The bacteriophage Mu N gene encodes the 64-kDa virion protein which is injected with, and circularizes, infecting Mu DNA. *J Biol Chem.* 1986, 261(35):

16682-8.

Goosen N, van de Putte P. Regulation of Mu transposition I. Localization of the presumed recognition sites for HimD and Ner functions controlling bacteriophage Mu transcription. *Gene*. 1984, 30(1-3): 41-6.

Goosen N, van de Putte P. Role of ner protein in bacteriophage Mu transposition. *J Bacteriol*. 1986, 167(2): 503-7.

Gourse R, Ross W, Gaal T. UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol Microbiol*. 2000, 37(4): 687-95.

Gourse RL, Ross W, Rutherford ST. General pathway for turning on promoters transcribed by RNA polymerases containing alternative sigma factors. *J Bacteriol*. 2006, 188(13): 4589-91.

Gross CA, Chan C, Dombroski A, Gruber T, Sharp M, Tupy J, Young B. The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harbor Symposia on Quantitative Biology*. 1998, 63: 141-55.

Grundy FJ, Howe MM. Involvement of the invertible G segment in bacteriophage Mu tail fiber biosynthesis. *Virology*. 1984, 134(2): 296-317.

Grundy FJ, Howe MM. Morphogenetic structures present in lysates of amber mutants of bacteriophage Mu. *Virology*. 1985, 143(2): 485-504.

Harley CB, Reynolds RP. Analysis of *E. coli* promoter sequences. *Nucleic Acids Res*. 1987, 15(5): 2343-61.

Harshey RM, Bukhari AI. Infecting bacteriophage Mu DNA forms a circular DNA-protein complex. *J Mol Biol*. 1983, 167(2): 427-41.

Harshey RM, Jayaram M. The Mu transpososome through a topological lens. *Crit Rev Biochem Mol Biol*. 2006, 41(6): 387-405. Review.

Hattman S. Unusual transcriptional and translational regulation of the bacteriophage Mu *mom* operon. *Pharmacol Ther*. 1999, 84(3): 367-88. Review.

Hattman S, Ives J, Margolin W, Howe MM. Regulation and expression of the bacteriophage Mu *mom* gene: mapping of the transactivation (dad) function to the C region. *Gene*. 1985, 39(1): 71-6.

Hawley DK, McClure WR. Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucl Acids Res*. 1983, 11(8): 2237-55. Review.

Heisig P, Kahmann R. The sequence and *mom*-transactivation function of the C gene of bacteriophage Mu. *Gene*. 1986, 43(1-2): 59-67.

Hook-Barnard I, Johnson XB, Hinton DM. *Escherichia coli* RNA polymerase recognition of a sigma70-dependent promoter requiring a -35 DNA element and an extended -10 TGn motif. *J Bacteriol*. 2006, 188(24): 8352-9.

Howe MM. Prophage deletion mapping of bacteriophage Mu-1. *Virology*. 1973, 54: 93-101.

Hsu LM, Vo NV, Kane CM, Chamberlin MJ. *In vitro* studies of transcript initiation by *Escherichia coli* RNA polymerase. 1. RNA chain initiation, abortive initiation, and promoter escape at three bacteriophage promoters. *Biochemistry*. 2003, 42(13): 3777-86.

Husnain SI, Thomas MS. The UP element is necessary but not sufficient for growth rate-dependent control of the *Escherichia coli* *guaB* promoter. *J Bacteriol*. 2008, 190(7): 2450-7.

Ishihama A. Functional modulation of *Escherichia coli* RNA polymerase. *Annu Rev Microbiol*. 2000, 54: 499-518. Review.

Ishihama A. Protein-protein communication within the transcription apparatus. *J Bacteriol*. 1993, 175(9): 2483-9. Review.

Ishihama A. Subunit of assembly of *Escherichia coli* RNA polymerase. *Adv Biophys*. 1981, 14: 1-35. Review.

Ishihama A, Fujita N, Glass RE. Subunit assembly and metabolic stability of *E. coli* RNA polymerase. *Proteins*. 1987, 2(1): 42-53.

Jiang Y. Mutational analysis of C protein: the late gene activator of bacteriophage Mu. 1999, University of Tennessee Health Science Center, Memphis.

Jones S, van Heyningen P, Berman HM, Thornton JM. Protein-DNA interactions: A structural analysis. *J Mol Biol*. 1999, 287(5): 877-96.

Juang YL, Helmann JD. A Promoter melting region in the primary [sigma] factor of *Bacillus subtilis*: Identification of functionally important aromatic amino acids. *J Mol Biol*. 1994, 235(5): 1470-88.

Juang YL, Helmann JD. Pathway of promoter melting by *Bacillus subtilis* RNA polymerase at a stable RNA promoter: effects of temperature, delta protein and sigma factor mutations. *Biochemistry*. 1995, 34(26): 8465-73.

Kahmann R. The *mom* gene of bacteriophage Mu. *Curr Top Microbiol Immunol*. 1984, 108: 29-47. Review.

- Kahmann R, Rudt F, Koch C, Mertens G. G inversion in bacteriophage Mu DNA is stimulated by a site within the invertase gene and a host factor. *Cell*. 1985, 41(3): 771-80.
- Kahmann R, Seiler A, Wulczyn FG, Pfaff E. The *mom* gene of bacteriophage mu: a unique regulatory scheme to control a lethal function. *Gene*. 1985, 39(1): 61-70.
- Kahmeyer-Gabbe M, Howe M. Regulatory factors acting at the bacteriophage Mu middle promoter. *J Bacteriol*. 1996, 178(6): 1585-92.
- Kamp D, Kahmann R, Zipser D, Broker TR, Chow LT. Inversion of the G DNA segment of phage Mu controls phage infectivity. *Nature*. 1978, 271(5645): 577-80.
- Keilty S, Rosenberg M. Constitutive function of a positively regulated promoter reveals new sequences essential for activity. *J Biol Chem*. 1987, 262(13): 6389-95.
- Korzheva N, Mustaev A. Transcription elongation complex: structure and function. *Curr Opin in Microbiol*. 2001, 4(2): 119-25. Review.
- Krause HM, Higgins NP. On the Mu repressor and early DNA intermediates of transposition. *Cold Spring Harb Symp Quant Biol*. 1984, 49: 827-34.
- Krause HM, Higgins NP. Positive and negative regulation of the Mu operator by Mu repressor and *Escherichia coli* integration host factor. *J Biol Chem*. 1986, 261(8): 3744-52.
- Krause HM, Rothwell MR, Higgins NP. The early promoter of bacteriophage Mu: definition of the site of transcript initiation. *Nucl Acids Res*. 1983, 11(16): 5483-95.
- Krummel B, Chamberlin MJ. RNA chain initiation by *Escherichia coli* RNA polymerase. Structural transitions of the enzyme in early ternary complexes. *Biochemistry*. 1989, 28(19): 7829-42.
- Krummel B, Chamberlin MJ. Structural analysis of ternary complexes of *Escherichia coli* RNA polymerase. Individual complexes halted along different transcription units have distinct and unexpected biochemical properties. *J Mol Biol*. 1992, 225(2): 221-37.
- Kumar A, Malloch RA, Fujita N, Smillie DA, Ishihama A, Hayward RS. The minus 35-recognition region of *Escherichia coli* sigma70 is inessential for initiation of transcription at an "extended minus 10" promoter. *J Mol Biol*. 1993, 232(2): 406-18.
- Kumaraswami M. Structural and functional characterization of Phage Mu Mor gene. 2005, University of Tennessee Health Science Center, Memphis.
- Kumaraswami M, Howe MM, Park HW. Crystal structure of the Mor protein of bacteriophage Mu, a member of the Mor/C family of transcription activators. *J Biol Chem*. 2004, 279(16): 16581-90.

Kuznedelov K, Korzheva N, Mustaev A, Severinov K. Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA-DNA hybrid length. *EMBO J.* 2002, 21(6): 1369-78.

Kuznedelov K, Minakhin L, Niedziela-Majka A, Dove SL, Rogulja D, Nickels BE, Hochschild A, Heyduk T, Severinov K. A role for interaction of the RNA polymerase flap domain with the sigma subunit in promoter recognition. *Science.* 2002, 295(5556): 855-7.

Li M, Moyle H, Susskind M. Target of the transcriptional activation function of phage lambda cI protein. *Science.* 1994, 263: 75-7.

Liebart JC, Ghelardini P, Paolozzi L. Conservative integration of bacteriophage Mu DNA into pBR322 plasmid. *Proc Natl Acad Sci USA.* 1982, 79(14): 4362-6.

Lonetto M, Gribskov M, Gross C. The sigma70 family: sequence conservation and evolutionary relationships. *J Bacteriol.* 1992, 174: 3843-9.

Lonetto MA, Rhodius V, Lamberg K, Kiley P, Busby S, Gross C. Identification of a contact site for different transcription activators in region 4 of the *Escherichia coli* RNA polymerase  $\sigma$ 70 subunit. *J Mol Biol.* 1998, 284(5): 1353-65.

Ma J. Protein-DNA and protein-protein interactions regulating Mu middle promoter activation. 2004, University of Tennessee Health Science Center, Memphis.

Ma J, Howe MM. Binding of the C-terminal domain of the alpha subunit of RNA polymerase to the phage Mu middle promoter. *J Bacteriol.* 2004, 186(23): 7858-64.

Malhotra A, Severinova E, Darst S. Crystal structure of a sigma70 subunit fragment from *E. coli* RNA polymerase. *Cell.* 1996, 87: 127-36.

Margolin W, Howe MM. Localization and DNA sequence analysis of the C gene of bacteriophage Mu, the positive regulator of Mu late transcription. *Nucl Acids Res.* 1986, 14(12): 4881-97.

Margolin W, Howe MM. Activation of the bacteriophage Mu lys promoter by Mu C protein requires the sigma70 subunit of *Escherichia coli* RNA polymerase. *J Bacteriol.* 1990, 172(3): 1424-9.

Margolin W, Rao G, Howe MM. Bacteriophage Mu late promoters: four late transcripts initiate near a conserved sequence. *J Bacteriol.* 1989, 171(4): 2003-18.

Marrs CF, Howe MM. Kinetics and regulation of transcription of bacteriophage Mu. *Virology.* 1990, 174(1): 192-203.

- Mathee K, Howe MM. Bacteriophage Mu Mor protein requires sigma70 to activate the Mu middle promoter. *J Bacteriol.* 1993, 175(17): 5314-23.
- Mathee K, Howe MM. Identification of a positive regulator of the Mu middle operon. *J Bacteriol.* 1990, 172(12): 6641-50.
- Mathee K, Hughes KT. The anti-sigma factors. *Annu Rev of Microbiol.* 1998, 52: 231-86.
- Mathew R, Chatterji D. The evolving story of the omega subunit of bacterial RNA polymerase. *Trends Microbiol.* 2006, 14(10): 450-5.
- Mauro SA, Pawlowski D, Koudelka GB. The role of the minor groove substituents in indirect readout of DNA sequence by 434 repressor. *J Biol Chem.* 2003, 278(15): 12955-60.
- McAllister WT, Raskin CA. The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. *Mol Microbiol.* 1993, 10(1): 1-6. Review.
- Meng W, Belyaeva T, Savery NJ, Busby SJ, Ross WE, Gaal T, Gourse RL, Thomas MS. UP element-dependent transcription at the *Escherichia coli* rrnB P1 promoter: positional requirements and role of the RNA polymerase alpha subunit linker. *Nucl Acids Res.* 2001, 29(20): 4166-78.
- Miller J. *Experiments in Molecular Genetics.* Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. 1972, 352-5.
- Mitchell JE, Oshima T, Piper SE, Webster CL, Westblade LF, Karimova G, Ladant D, Kolb A, Hobman JL, Busby SJ, Lee DJ. The *Escherichia coli* regulator of sigma70 protein, Rsd, can up-regulate some stress-dependent promoters by sequestering sigma70. *J Bacteriol.* 2007, 189(9): 3489-95.
- Mo Y. RNA polymerase recruitment, dimerization and  $\alpha$ -CTD dependent activation by members of Mor/C family of transcriptional activators. 2004, University of Tennessee Health Science Center, Memphis.
- Moravek Z, Neidle S, Schneider B. Protein and drug interactions in the minor groove of DNA. *Nucl Acids Res.* 2002, 30(5): 1182-91.
- Mukhopadhyay J, Sineva E, Knight J, Levy RM, Ebright RH. Antibacterial peptide microcin J25 inhibits transcription by binding within and obstructing the RNA polymerase secondary channel. *Mol Cell.* 2004, 14(6): 739-51.
- Murakami KS, Darst SA. Bacterial RNA polymerases: the whole story. *Curr Opin Struct Biol.* 2003, 13(1): 31-9.

Murakami KS, Masuda S, Campbell EA, Muzzin O, Darst SA. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science*. 2002, 296(5571): 1285-90.

Murakami KS, Masuda S, Darst SA. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science*. 2002, 296(5571): 1280-4.

Nakai H, Doseeva V, Jones JM. Handoff from recombinase to replisome: insights from transposition. *Proc Natl Acad Sci USA*. 2001, 98(15): 8247-54. Review.

Nickels BE, Dove SL, Murakami KS, Darst SA, Hochschild A. Protein-protein and protein-DNA interactions of sigma70 region 4 involved in transcription activation by lambda<sub>da</sub>I. *J Mol Biol*. 2002, 324(1): 17-34.

Nickels BE, Hochschild A. Regulation of RNA polymerase through the secondary channel. *Cell*. 2004, 118(3): 281-4. Review.

Nudler E. Transcription elongation: structural basis and mechanisms. *J Mol Biol*. 1999, 288(1): 1-12. Review.

Pato ML, Howe MM, Higgins NP. A DNA gyrase-binding site at the center of the bacteriophage Mu genome is required for efficient replicative transposition. *Proc Natl Acad Sci USA*. 1990, 87(22): 8716-20.

Perederina A, Svetlov V, Vassilyeva MN, Tahirov TH, Yokoyama S, Artsimovitch I, Vassilyev DG. Regulation through the secondary channel structural framework for ppGpp-DksA synergism during transcription. *Cell*. 2004, 118(3): 297-309.

Pérez-Martín J, de Lorenzo V. Clues and consequences of DNA bending in transcription. *Annu Rev Microbiol*. 1997, 51: 593-628. Review.

Pérez-Martín J, Espinosa M. Correlation between DNA bending and transcriptional activation at a plasmid promoter. *J Mol Biol*. 1994, 241(1): 7-17.

Pérez-Martín J, Espinosa M. Protein-induced bending as a transcriptional switch. *Science*. 1993, 260(5109): 805-7.

Pérez-Martín J, Rojo F, de Lorenzo V. Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. *Microbiol Rev*. 1994, 58(2): 268-90. Review.

Polyakov A, Severinova E, Darst SA. Three-dimensional structure of *E. coli* core RNA polymerase: promoter binding and elongation conformations of the enzyme. *Cell*. 1995, 83(3): 365-73.

Ramesh V, Nagaraja V. Sequence-specific DNA binding of the phage Mu C protein: Footprinting analysis reveals altered DNA conformation upon protein binding. *J Mol*

Biol. 1996, 260(1): 22-33.

Ranquet C, Toussaint A, de Jong H, Maenhaut-Michel G, Geiselman J. Control of bacteriophage Mu lysogenic repression. *J Mol Biol.* 2005, 353(1): 186-95.

Richardson JP. Transcription termination. *Crit Rev Biochem Mol Biol.* 1993, 28(1): 1-30. Review.

Rojo F. Mechanisms of transcriptional repression. *Curr Opin Microbiol.* 2001, 4(2): 145-51.

Rojo F. Repression of Transcription initiation in bacteria. *J Bacteriol.* 1999, 181(10): 2987-91.

Rojo F, Mencía M, Monsalve M, Salas M. Transcription activation and repression by interaction of a regulator with the alpha subunit of RNA polymerase: the model of phage phi 29 protein p4. *Prog Nucleic Acid Res Mol Biol.* 1998, 60: 29-46. Review.

Ross W, Gosink KK, Salomon J, Igarashi K, Zou C, Ishihama A, Severinov K, Gourse RL. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science.* 1993, 262(5138): 1407-13.

Ross W, Gourse RL. Sequence-independent upstream DNA-alphaCTD interactions strongly stimulate *Escherichia coli* RNA polymerase-lacUV5 promoter association. *Proc Natl Acad Sci USA.* 2005, 102(2): 291-6.

Ross W, Schneider DA, Paul BJ, Mertens A, Gourse RL. An intersubunit contact stimulating transcription initiation by *E.coli* RNA polymerase: interaction of the alpha C-terminal domain and sigma region 4. *Genes Dev.* 2003, 17(10): 1293-307.

Sandulache R, Prehm P, Kamp D. Cell wall receptor for bacteriophage Mu G (+). *J Bacteriol.* 1984, 160(1): 299-303.

Seshasayee AS, Bertone P, Fraser GM, Luscombe NM. Transcriptional regulatory networks in bacteria: from input signals to output responses. *Curr Opin Microbiol.* 2006, 9(5): 511-9. Review.

Severinov K, Fenyö D, Severinova E, Mustaev A, Chait BT, Goldfarb A, Darst SA. The sigma subunit conserved region 3 is part of "5'-face" of active center of *Escherichia coli* RNA polymerase. *J Biol Chem.* 1994, 269(33): 20826-8.

Sevostyanova A, Feklistov A, Barinova N, Heyduk E, Bass I, Klimasauskas S, Heyduk T, Kulbachinskiy A. Specific recognition of the -10 promoter element by the free RNA polymerase sigma subunit. *J Biol Chem.* 2007, 282(30): 22033-9.



Sharp MM, Chan CL, Lu CZ, Marr MT, Nechaev S, Merritt EW, Severinov K, Roberts JW, Gross CA. The interface of sigma with core RNA polymerase is extensive, conserved and functionally specialized. *Genes Dev.* 1999, 13(22): 3015-26.

Shore SH, Howe MM. Bacteriophage Mu T mutants are defective in synthesis of the major head polypeptide. *Virology.* 1982, 120(1): 264-8.

Shultzaberger RK, Chen Z, Lewis KA, Schneider TD. Anatomy of *Escherichia coli* sigma70 promoters. *Nucl Acids Res.* 2007, 35(3): 771-88.

Sivan S, Zaritsky A, Kagan-Zur V. Replication forks of *Escherichia coli* are not the preferred sites for lysogenic integration of bacteriophage Mu. *J Bacteriol.* 1988, 170(7): 3089-93.

Stoddard SF, Howe MM. Characterization of the C operon transcript of bacteriophage Mu. *J Bacteriol.* 1990, 172(1): 361-71.

Stoddard SF, Howe MM. DNA sequence within the Mu C operon. *Nucl Acids Res.* 1987, 15(17): 7198.

Stoddard SF, Howe MM. Localization and regulation of bacteriophage Mu promoters. *J Bacteriol.* 1989, 171(6): 3440-8.

Symonds NA, Toussaint A, Van de Putte P, Howe MM. Phage Mu. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 1987.

Toussaint A, Faelen M, Desmet L, Allet B. The products of gene A of the related phages Mu and D108 differ in their specificities. *Mol Gen Genet.* 1983, 190(1): 70-9.

Toussaint A, Lecocq JP. Sensitivity of bacteriophage Mu-1 development to rifampicin and streptolydigin. *Mol Gen Genet.* 1974, 129(2): 185-8.

Travers A. Reading the minor groove. *Nat. Struct. Biol.* 1995, 2(8): 615-8.

Tunitskaya VL, Kochetkov SN. Structural-functional analysis of bacteriophage T7 RNA polymerase. *Biochemistry (Mosc).* 2002, 67(10): 1124-35. Review.

Uptain SM, Kane CM, Chamberlin MJ. Basic mechanisms of transcript elongation and its regulation. *Annu Rev Biochem.* 1997, 66: 117-72. Review.

van de Putte P, Cramer S, Giphart-Gassler M. Invertible DNA determines host specificity of bacteriophage Mu. *Nature.* 1980, 286(5770): 218-22.

van de Putte P, Giphart-Gassler M, Goosen N, Goosen T, van Leerdam E. Regulation of integration and replication functions of bacteriophage Mu. *Cold Spring Harb Symp Quant Biol.* 1981, 45(1): 347-53.

- Van Leerdam E, Karreman C, van de Putte P. *Ner*, a *cro*-like function of bacteriophage Mu. *Virology*. 1982, 123(1): 19-28.
- Vassilyev DG, Sekine S, Laptenko O, Lee J, Vassilyeva MN, Borukhov S, Yokoyama S. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6Å resolution. *Nature*. 2002, 417(6890): 712-9.
- Vassilyev DG, Vassilyeva MN, Perederina A, Tahirov TH, Artsimovitch I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*. 2007, 48(7150): 157-62.
- Vassilyev DG, Vassilyeva MN, Zhang J, Palangat M, Artsimovitch I, Landick R. Structural basis for substrate loading in bacterial RNA polymerase. *Nature*. 2007, 48(7150): 163-8.
- Vo NV, Hsu LM, Kane CM, Chamberlin MJ. *In vitro* studies of transcript initiation by *Escherichia coli* RNA polymerase. 2. Formation and characterization of two distinct classes of initial transcribing complexes. *Biochemistry*. 2003, 42(13): 3787-97.
- Vo NV, Hsu LM, Kane CM, Chamberlin MJ. *In vitro* studies of transcript initiation by *Escherichia coli* RNA polymerase. 3. Influences of individual DNA elements within the promoter recognition region on abortive initiation and promoter escape. *Biochemistry*. 2003, 42(13): 3798-811.
- von Hippel PH, Pasmán Z. Reaction pathways in transcript elongation. *Biophys Chem*. 2002, 101-102: 401-23.
- Wassarman KM. 6S RNA: a small RNA regulator of transcription. *Curr Opin Microbiol*. 2007, 10(2): 164-8. Review.
- Wijffelman C, Gassler M, Stevens WF, van de Putte P. On the control of transcription of bacteriophage Mu. *Mol Gen Genet*. 1974, 131(2): 85-96.
- Wijffelman C, Lotterman B. Kinetics of Mu DNA synthesis. *Mol Gen Genet*. 1977, 51(2): 169-74.
- Wijffelman C, van de Putte P. Transcription of bacteriophage Mu. An analysis of the transcription pattern in the early phase of phage development. *Mol Gen Genet*. 1974, 135(4): 327-37.
- Witkowski RT, Hattman S, Newman L, Clark K, Tierney DL, Penner-Hahn J, McLendon G. The zinc coordination site of the bacteriophage Mu translational activator protein complex. *J Mol Biol*. 1995, 247(4): 753-64.
- Wulczyn FG, Bölker M, Kahmann R. Translation of the bacteriophage Mu *mom* gene is

- positively regulated by the phage com gene product. *Cell*. 1989, 57(7): 1201-10.
- Wulczyn FG, Kahmann R. Translational stimulation: RNA sequence and structure requirements for binding of Com protein. *Cell*. 1991, 65(2): 259-69.
- Young BA, Gruber TM, Gross CA. Views of transcription initiation. *Cell*. 2002, 109(4): 417-20.
- Zenkin N, Kulbachinskiy A, Yuzenkova Y, Mustaev A, Bass I, Severinov K, Brodolin K. Region 1.2 of the RNA polymerase sigma subunit controls recognition of the -10 promoter element. *EMBO J*. 2007, 26(4): 955-64.
- Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell*. 1999. 98(6): 811-24.
- Zhao Z. Effects of  $P_{lys}$  promoter sequence on transcription activation by the C protein of bacteriophage Mu. 1999. University of Tennessee Health Science Center, Memphis.
- Zhou D, Yang R. Global analysis of gene transcription regulation in prokaryotes. *Cell Mol Life Sci*. 2006, 63(19-20): 2260-90. Review.

## Appendix: Determining the effect of N- and C-terminal deletions of Mor

In the Mor/C family of proteins the least conserved regions are the N-terminal and C-terminal portions of as shown in Figure 35. These are also the portions of Mor that are not visible in the crystal structure. It is important to note, however, that these sequences are not random; they simply show a smaller number of matches than the central region. Their potential importance to Mor function is underscored by the fact that deletion of the 26 N-terminal or 9 C-terminal amino acids, not seen in the structure, render Mor non-functional (Kumaraswami et al, 2004). Thus, the N- and C-terminal regions of Mor may be important for its structure and stability, binding to DNA or transcription activation through interactions with RNA polymerase. In order to determine the functional relevance of these terminal regions, serial deletions were made from both ends. For the N-terminus, 5, 10, 15, 20 and 25 amino acids were deleted and at the C-terminus, 3, 6, 9, 12 and 15 amino acids were deleted.

Primers with the above deletions were used in PCR reactions to create the deleted *mor* genes. Table 8 shows the sequences of the oligonucleotides that were used in this study. The N-terminal deletion primers retained the codon for the initiating methionine following the His-tag and the C-terminal deletion primers retained the stop codon. The deletion was made in two steps as shown in Figure 36: For the N-terminus, the top strand deletion primer was used with the bottom strand wild-type primer ZAO 3 in PCR amplification reaction to create PCR product 1. At the same time, the bottom strand primer with overlap to the N-terminal deletion primers was used with MUT 13 in a separate PCR reaction to create PCR product 2. The resulting PCR products 1 and 2 with overlapping ends were used as templates in another PCR reaction, and amplification was achieved using top and bottom strand wild-type primers, MUT 13 and ZAO 3, respectively. A similar strategy was used to make the C-terminal deletions with the difference that the deletion bottom-strand primers were used with MUT 13 for amplification, while the top strand overlap primer was used with ZAO 3. The PCR reaction combining the first two PCR products was similar to that for the N-terminal deletions. The template for the first and second PCR amplification reactions was the wild-type pIA69 which carries the *his-mor* gene under the influence of an IPTG-inducible  $P_{lacUV5}$  promoter. After purification of the final PCR products with a Qiaquick PCR purification kit, they were sequentially digested with *NdeI* and *BamHI* at 37°C for 3 hours each. The digested products were again purified with a Qiaquick PCR purification kit and ligation reactions were set up with the similarly digested vector pIA69, and incubated at 16°C overnight. The ligation mixture was transformed into MH13435 carrying the  $P_m-lacZ$  fusion plasmid pIA14. The resulting transformants were selected on MacConkey lactose indicator plates with 50  $\mu$ M IPTG, chloramphenicol (25  $\mu$ g/ml) and ampicillin (40  $\mu$ g/ml); the promoter plasmid, pIA14 confers ampicillin resistance and the protein plasmid, pIA69 confers chloramphenicol resistance. Candidate clones were purified and sequenced to identify the deletions. The phenotypes resulting from the N- and C-terminal deletions are listed in Table 9.

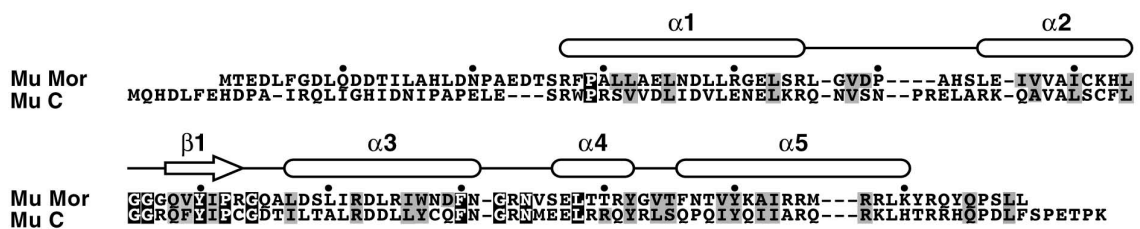


Figure 35: Amino acid and secondary structure alignment for Mor and C proteins of bacteriophage Mu.

$\alpha$ -helix is shown as an oval and  $\beta$ -strand as an arrow. Identical amino acids are shown in black boxes; chemically similar residues are shaded in gray. Dots indicate 10 amino acid intervals in Mor.

Modified with permission from American Society for Biochemistry and Molecular biology. Kumaraswami M, Howe MM, Park HW. Crystal structure of the Mor protein of bacteriophage Mu, a member of the Mor/C family of transcription activators. J Biol Chem. 2004, 279(16): 16581-90.

Table 8: Oligonucleotides used for N- and C-terminal deletions of His-Mor.

| Primer | Sequence (from 5' to 3' end)                 | Comments                                                                       |
|--------|----------------------------------------------|--------------------------------------------------------------------------------|
| KRI 1  | GATGACGATAAGATG Δ<br>GGTGATCTGCAGGATGACAC    | For deletion of 5 amino acids from the N-terminus of Mor.                      |
| KRI 2  | GATGACGATAAGATG Δ<br>GACACCATCCTGGCACAT      | For deletion of 10 amino acids from the N-terminus of Mor.                     |
| KRI 3  | GATGACGATAAGATG Δ<br>CATCTTGACAATCC          | For deletion of 15 amino acids from the N-terminus of Mor.                     |
| KRI 4  | GATGACGATAAGATG Δ<br>GCCGAGGACACGTCACGCTTT   | For deletion of 20 amino acids from the N-terminus of Mor.                     |
| KRI 20 | GATGACGATAAGATG Δ<br>CGCTTCCGGCACTGCTGGCGGAG | For deletion of 25 amino acids from the N-terminus of Mor.                     |
| KRI 5  | CATCTTATCGTCATCGTCGAG                        | Bottom strand primer with overlapping region with N-terminal deletion primers. |
| KRI 6  | CGGCTTTACTGTTTA Δ<br>GGGCTGGTACTGGCGGTA      | For deletion of 3 amino acids from the C-terminus of Mor.                      |
| KRI 7  | CGGCTTTACTGTTTA Δ<br>CTGGCGGTATTTCA          | For deletion of 6 amino acids from the C-terminus of Mor.                      |
| KRI 21 | CGGCTTTACTGTTTA Δ<br>TTTCAGCCGTCG            | For deletion of 9 amino acids from the C-terminus of Mor.                      |
| KRI 22 | CGGCTTTACTGTTTA Δ<br>TCGCATGCGCGAAT          | For deletion of 12 amino acids from the C-terminus of Mor.                     |
| KRI 23 | CGGCTTTACTGTTTA Δ<br>GCGAATGGCTTTGTACACCGTA  | For deletion of 15 amino acids from the C-terminus of Mor.                     |
| KRI 8  | TAAACAGTAAAGCCGGTTAATCC                      | Top strand primer with overlapping region with N-terminal deletion primers.    |
| MUT 13 | CCAGGCTTTACACTTTATGC                         | Wild-type top strand primer beyond 5' end of <i>mor</i> gene.                  |
| ZAO 3  | ACCTGAAGTCAGCCCCATAC                         | Wild-type bottom strand primer beyond 3' end of <i>mor</i> gene.               |

The oligonucleotide sequences are written from the 5' to the 3' end.  
 Δ indicates the position of the deletion from the *mor* gene.

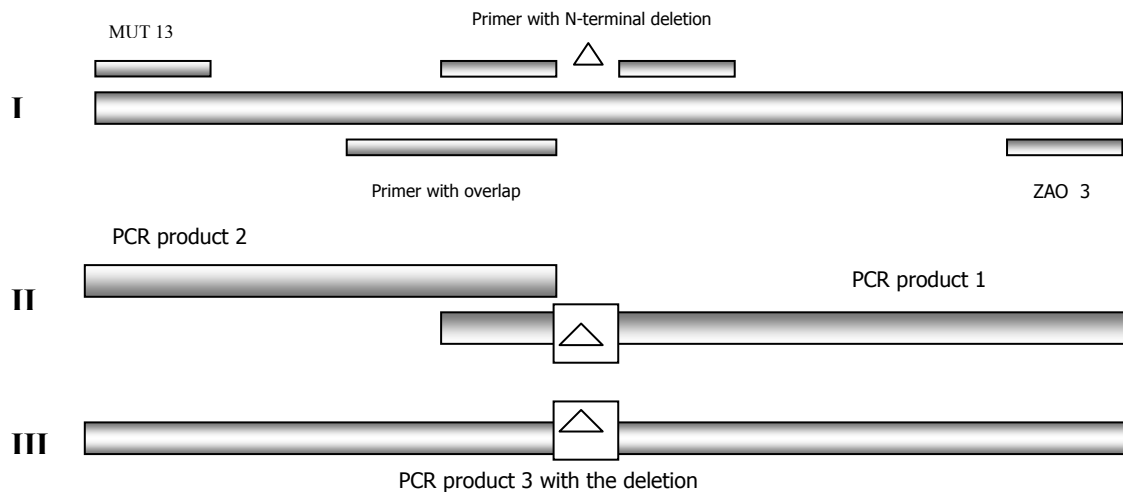


Figure 36: Schematic diagram of the strategy used for deletion of amino acids from the N-terminus of Mor.

The shaded block represents the part of pIA69 with the *mor* gene. The  $\Delta$  shows the region of deletion. Layer I shows the position of the primers used for first set of amplification reactions. Layer II shows the products of the first two amplifications, their regions of overlap, and the region deleted. Layer III shows the final PCR product containing the deletion in the *mor* gene.

Table 9: Resulting phenotypes of the N- and C-terminal deletion proteins.

| <b>N-terminal deletions of Mor</b> | <b>Phenotype of the colony with the deletion</b> |
|------------------------------------|--------------------------------------------------|
| Δ 5 amino acids                    | Red                                              |
| Δ 10 amino acids                   | Red                                              |
| Δ 15 amino acids                   | White                                            |
| Δ 20 amino acids                   | White                                            |
| Δ 25 amino acids                   | White                                            |

| <b>C-terminal deletions of Mor</b> | <b>Phenotype of the colony with the deletion</b> |
|------------------------------------|--------------------------------------------------|
| Δ 3 amino acids                    | Red                                              |
| Δ 6 amino acids                    | Red                                              |
| Δ 9 amino acids                    | White                                            |
| Δ 12 amino acids                   | White                                            |
| Δ 15 amino acids                   | White                                            |

Δ indicates deletion.



Protein over-expression and small-scale protein purification for gel shift assay were performed as described in material and methods. Labeled probe was prepared using a  $\gamma$ -P<sup>32</sup> labeled primer MLK 7 and unlabeled IRI 21 in a PCR reaction using pIA14, containing P<sub>m</sub> sequences from -61 to +10 as template. The labeled probe was purified with a Qiaquick PCR purification kit. A 20  $\mu$ l reaction volume containing about 50 counts per second of radioactive probe, 50 ng of calf-thymus DNA, and different concentrations of the protein in binding buffer (20 mM Tris-HCl pH 7.9, 50 mM NaCl, 5% glycerol and 1 mM DTT) was incubated at room temperature for 20 minutes. The reaction mixture was resolved on a 10% non-denaturing, native acrylamide gel containing 0.5 X TBE and run in 0.5 X TBE buffer at 260 V for 3 hours at 4°C. Initial exposure of the gels to X-OMAT BioMax-MR films was done without drying.

DNA-binding assays with the C-terminal deleted His-Mor proteins showed that proteins with  $\Delta$ 3C and  $\Delta$ 6C at the C-terminus retained DNA-binding ability correlating with their red colony phenotype. DNA-binding assay for Mor proteins carrying C-terminal deletions is shown in Figure 37. Both  $\Delta$ 3C and  $\Delta$ 6C proteins bound the probe less than the wild-type and the  $\Delta$ 3C protein bound the probe less than  $\Delta$ 6C protein at both concentrations of protein used. For the two deletions and the wild-type, 800 ng of protein bound the probe more than 400 ng of the probe as indicated by the intensity of the shift.  $\Delta$ 9C,  $\Delta$ 12C and  $\Delta$ 15C did not bind the DNA probe at concentrations even as high as 800 ng of His-Mor. Thus, the contribution by the last 6 amino acids of Mor is less critical than the others at the C-terminus as their deletion gave us functional protein in terms of binding to DNA as well as transactivation as indicated by red colony phenotype. It would be useful to investigate the importance of the amino acids at the C-terminus in terms of their contribution to Mor structure and function.

An experiment was done to observe the shift by lower concentration of the protein. The gel shift for this experiment is shown in Figure 38. As observed previously,  $\Delta$ 6 C-terminus of His-Mor gave functional protein with respect to DNA-binding and hence activation determined by the red color of the colony.  $\Delta$ 6C bound the probe less than the wild-type. For  $\Delta$ 6C and the wild-type, the amount of probe bound to the protein gradually increases as the amount of protein increases in the reaction as indicated by the intensity of the shift.  $\Delta$ 6C bound the probe even at concentration as low as 200 ng of protein, but  $\Delta$ 9C did not bind the DNA probe at concentrations even as high as 800 ng of His-Mor.

Similar to the C-terminus, DNA-binding experiments were done for proteins deleted from the N-terminus. DNA-binding assays with the deleted His-Mor proteins showed that proteins with  $\Delta$ 5N and  $\Delta$ 10N at the N-terminus retained DNA-binding ability consistent with their red colony phenotype. Both  $\Delta$ 5N and  $\Delta$ 10N proteins bound the probe less than the wild-type and  $\Delta$ 10N protein bound the probe more than  $\Delta$ 5N at both concentrations of protein used. For the two deletions and the wild-type, 800 ng of protein bound the probe more than 400 ng of the probe as indicated by the intensity of the shift.  $\Delta$ 15N,  $\Delta$ 20N and  $\Delta$ 25N did not bind the DNA probe at concentrations even as high as 800 ng of His-Mor. Thus, the contribution by the first 10 amino acids of Mor is less critical than the others at the N-terminus as their deletion gave us functional protein in

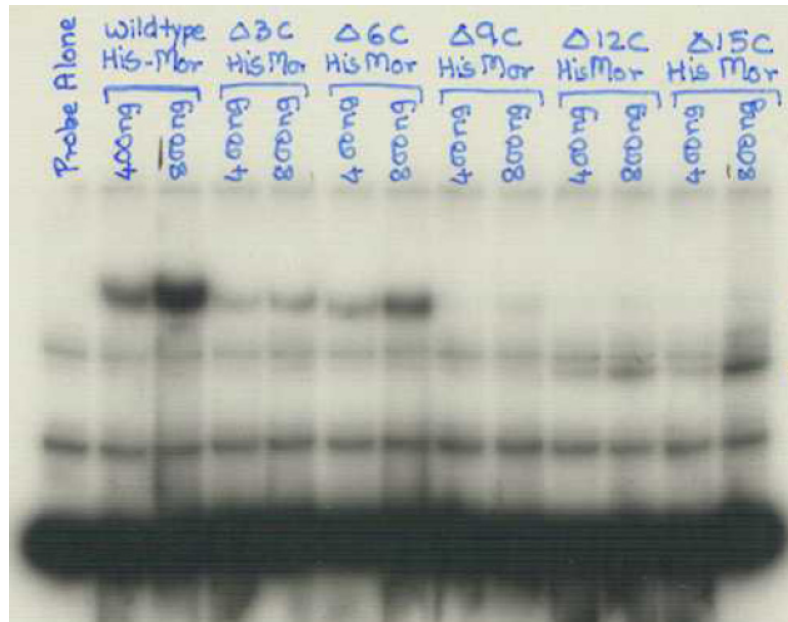


Figure 37: Gel shift for Mor proteins carrying C-terminal deletions.

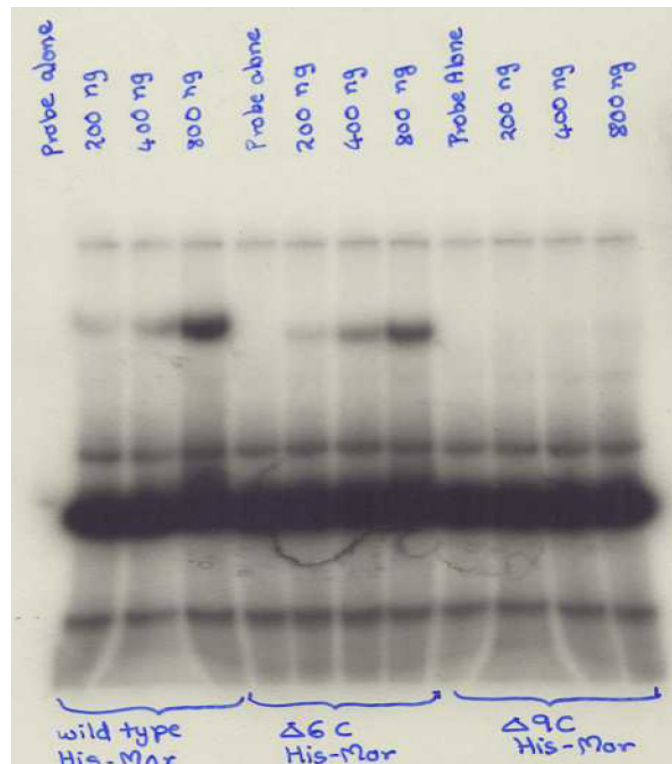


Figure 38: Gel shift for  $\Delta 6C$  and  $\Delta 9C$  Mor proteins.

terms of binding to DNA as well as transactivation as indicated by red colony phenotype. DNA-binding assay for Mor proteins carrying N-terminal deletions is shown in Figure 39. It would be useful to investigate the importance of the amino acids at the N-terminus in terms of their contribution to Mor structure and function.

An experiment was done to observe the shift by lower concentration of the protein. The gel shift for this experiment is shown in Figure 40. As observed previously,  $\Delta 10N$  of His-Mor gave functional protein with respect to DNA-binding and hence activation determined by the red color of the colony.  $\Delta 10N$  bound the probe less than the wild-type. For  $\Delta 10N$  and the wild-type, the amount of probe bound to the protein gradually increases as the amount of protein increases in the reaction as indicated by the intensity of the shift.  $\Delta 10N$  bound the probe even at concentration as low as 200 ng of protein, but  $\Delta 15N$  did not bind the DNA probe at concentrations even as high as 800 ng of His-Mor.

It was observed reproducibly in the N-terminal deletion and C-terminal deletion gels that a band appeared as the number of amino acid deleted increased. So, an experiment was done to test if any other proteins from the cell supernatant bind the DNA probe and result in a shift in the probe. A constant amount of wild-type His-Mor (400 ng) was used with increasing amount of cell supernatant. The supernatants were from a strain that did not have *his-mor* gene in its plasmid. Since there are no additional bands that appear to be shifted with concentrations as high as 6400 ng of the supernatant used, it could be concluded that no other protein from the cell supernatant fraction other than His-Mor bound the probe as seen in Figure 41.

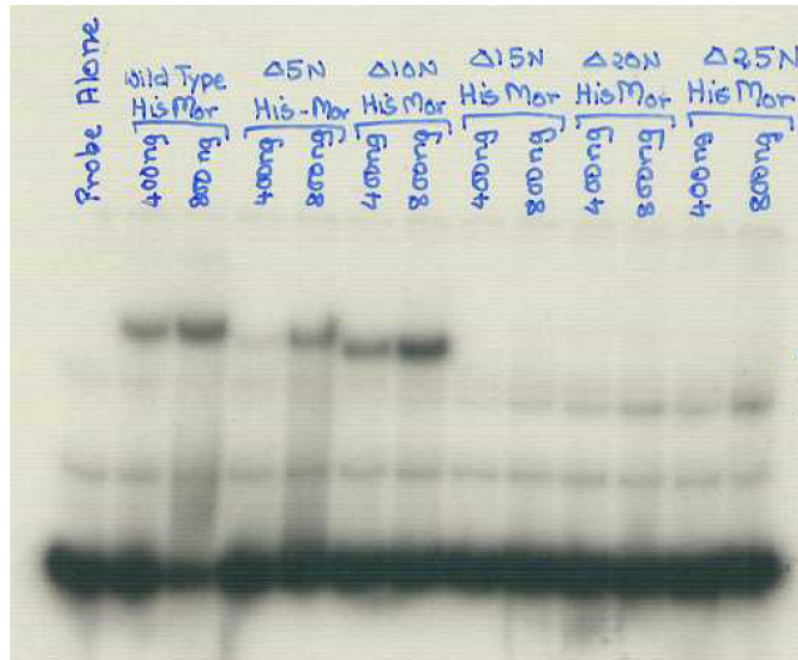


Figure 39: Gel shift for Mor proteins carrying N-terminal deletions.

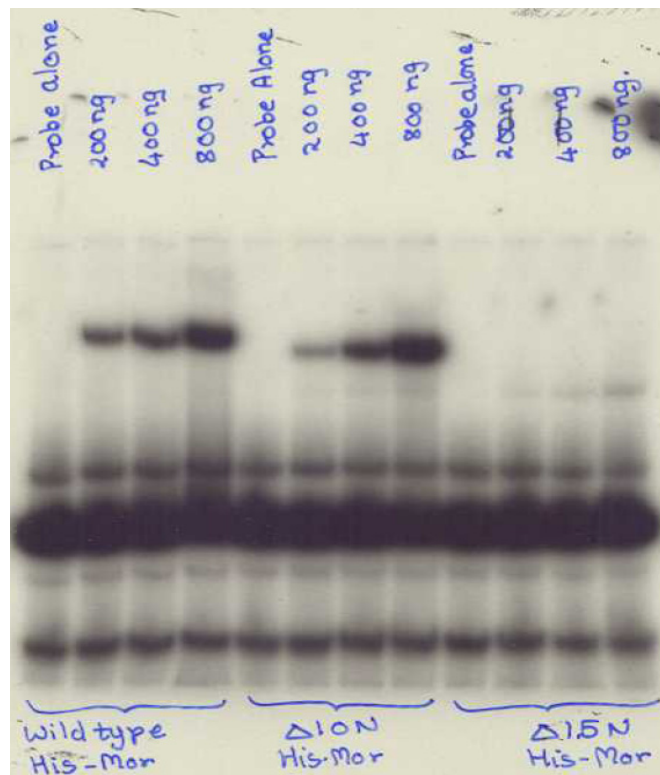


Figure 40: Gel shift for  $\Delta 10N$  and  $\Delta 15N$  Mor proteins.

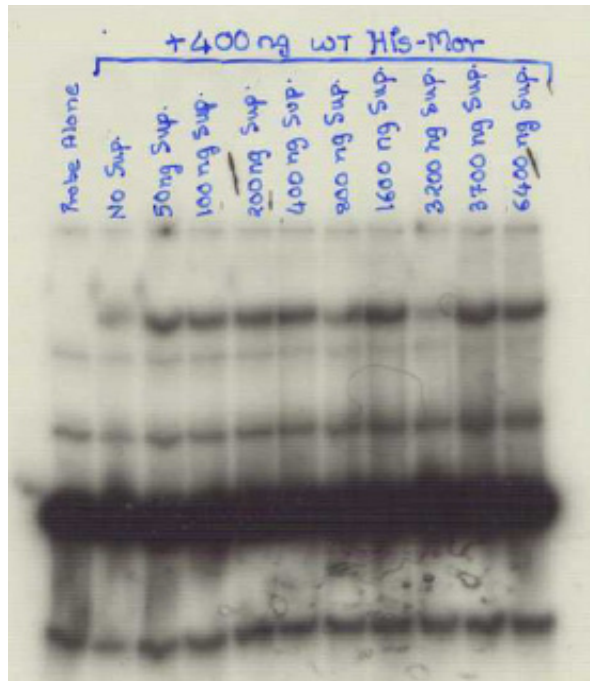


Figure 41: Gel shift with increasing amounts of supernatant.

## Vita

Kartik Iyer was born in Mumbai, India on November 24, 1980. He studied in Crescent High School, Pune, India till Class X and then joined Abasaheb Garware College, Pune, India for Class XI and XII. In 1998, he went on to join the Sri Sathya Sai Institute of Higher Learning, Prasanthi Nilayam, India for higher education. This is the university under the chancellorship of the spiritual master, Sri Sathya Sai Baba. He graduated with a gold medal for *Bachelor of Science* degree in Biosciences in 2001 and then a *Master of Science* degree in Biosciences in 2003. He then worked as a research assistant at National Chemical Laboratory, Pune, India in the Division of Biological Sciences.

In the fall of 2005, he entered the graduate program at the University of Tennessee Health Science Center, Memphis, USA through the Integrated Program in Biomedical Sciences and joined the lab of Dr. Martha Howe directly for his research work. After the decision to go back to India was made in December 2007, Dr. Howe agreed to confer the *Master of Science* degree for the work in the lab. He submitted the thesis entitled “Interaction of Bacteriophage Mu Middle Transcription Activator Protein Mor with Promoter DNA” for the degree of Master of Science with a concentration in Microbial Pathogenesis, Immunology and Inflammation within the Integrated Program in Biomedical Sciences. He will receive his MS degree in December 2008.

He has decided to go back to India and will be at Amritapuri, the ashram of Mata Amritanandamayi Devi. He would be teaching at the School of Biotechnology in Amrita University.