Computer Science and Artificial Intelligence Laboratory
Technical Report

MIT-CSAIL-TR-2003-017
AITR-2003-017
August 31, 2003

Fluorescence Assay for Polymerase Arrival Rates

Austin Che

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

Fluorescence Assay for Polymerase Arrival Rates

by

Austin Che

B.S. and A.B., Stanford University (2001)

Submitted to the Department of Electrical Engineering
and Computer Science in partial fulfillment of the
requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2003

© Massachusetts Institute of Technology 2003. All rights
reserved.

Certified by: Thomas F. Knight, Jr.
Senior Research Scientist
Thesis Supervisor

Accepted by: Arthur C. Smith
Chairman, Department Committee on Graduate Students

Fluorescence Assay for Polymerase Arrival Rates

by

Austin Che

Submitted to the Department of Electrical Engineering
and Computer Science on August 28, 2003, in partial fulfillment of
the requirements for the degree of Master of Science

Abstract

To engineer complex synthetic biological systems will require modular design, assembly, and characterization strategies. The RNA *polymerase arrival rate* (PAR) is defined to be the rate that RNA polymerases arrive at a specified location on the DNA. Designing and characterizing biological modules in terms of RNA polymerase arrival rates provides for many advantages in the construction and modeling of biological systems.

PARMESAN is an *in vitro* method for measuring polymerase arrival rates using pyrrolo-dC, a fluorescent DNA base that can substitute for cytosine. Pyrrolo-dC shows a detectable fluorescence difference when in single-stranded versus double-stranded DNA. During transcription, RNA polymerase separates the two strands of DNA, leading to a change in the fluorescence of pyrrolo-dC. By incorporating pyrrolo-dC at specific locations in the DNA, fluorescence changes can be taken as a direct measurement of the polymerase arrival rate.

Thesis Supervisor: Thomas F. Knight, Jr.
Title: Senior Research Scientist

Acknowledgments

Thanks to Richard Ebright (Rutgers) for plasmids to purify *E. coli* RNA polymerase and Chi-Wang Lin and Alice Ting (MIT) for rescuing me from some fluorescence frustrations.

Discussions with members of the Synthetic Biology group have been invaluable, especially Sri Kosuri, Julie Norville, Randy Rettberg, and Reshma Shetty. Also work on transcriptional modeling with Bogdan Fedeleş provided some of the initial insights into this thesis.

Thanks to Rajeev Surati for encouragement and the offer to send me to Europe if I would just finish my thesis before a deadline. I'm sorry to disappoint you, but I do have a PhD thesis to write and maybe you could send me to some place a bit more exotic as motivation.

Thank you for reading this and the next (many) pages. A small but substantial fraction of my life has disappeared into these pages. Comments are always welcome.

For everyone who has ever given me a smile and brightened my day, you've brought me the joy and hope that makes life worth living.

Finally, Tom Knight, the best adviser I could have hoped for and who introduced me to the life of biology. Without life, we'd all be dead and I would never have written this thesis. So on with life.

Table of Contents

Abstract **3**

Acknowledgments **5**

Table of Contents **7**

1 Introduction **13**
- 1.1 Motivation 13
- 1.2 Thesis Organization 14
- 1.3 Abbreviations 14

2 Background **17**
- 2.1 Transcription Overview 17
  - 2.1.1 RNA Polymerase 18
  - 2.1.2 Promoters 19
  - 2.1.3 Operators 20
- 2.2 Transcription Process 20
  - 2.2.1 Closed Complex 21
  - 2.2.2 Open Complex 21
  - 2.2.3 Initiation 21
  - 2.2.4 Elongation 22
  - 2.2.5 Pausing 23
  - 2.2.6 Termination 23
- 2.3 Transcription Assay Techniques 24
  - 2.3.1 Fluorescence 24
  - 2.3.2 *in vivo* Methods 26
    - Reporter Proteins 26
    - Enzyme Assays 26
    - Fluorescence Assays 26

RNA Measurements . . . 27
Relative Strengths . . . 27
2.3.3 *in vitro* Methods . . . 28
RNA Measurements . . . 28
Radioactive Labeling Methods . . . 28
Fluorescent Measurements . . . 29
Measuring DNA Binding . . . 29

3 PARMESAN Theory 31
3.1 Modularity . . . 31
3.1.1 Characterization . . . 32
3.2 PAR . . . 33
3.2.1 Definition . . . 33
3.2.2 Modules . . . 33
Examples . . . 34
3.2.3 Assumptions . . . 37
3.2.4 Motivation . . . 38
Conventional Models . . . 39
Rates vs. Levels . . . 40
Transcription vs. Translation . . . 41
Modularity . . . 44
3.2.5 Design and Assembly . . . 45
Design . . . 45
Assembly . . . 45
3.3 PARMESAN . . . 48
3.3.1 Pyrrolo-dC . . . 48
3.3.2 Transcription Assay . . . 49
Theory . . . 49
Measurement Issues . . . 50
3.4 Applications . . . 51
3.4.1 Promoter Library . . . 51
3.4.2 Controlling Gene Expression . . . 51
3.4.3 Transfer Curves . . . 52
3.4.4 Modeling Biological Networks . . . 53
3.4.5 Synthetic Biology . . . 53

4 Measurement Methods 55
4.1 Protocol Overview . . . 55
4.1.1 BioBricks . . . 55
4.2 Materials . . . 56
4.2.1 RNA Polymerase Purification . . . 57
4.3 Oligos . . . 57

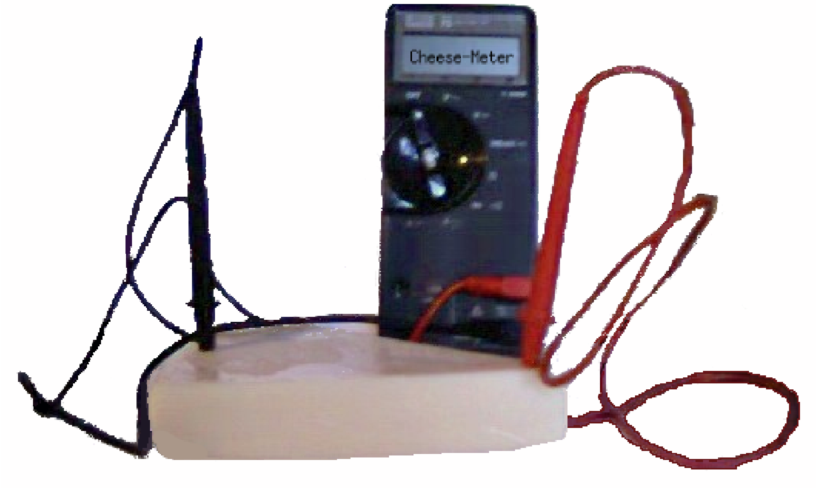
4.3.1 Oligo Synthesis 57
4.3.2 BB-R-P 58
4.3.3 TTT/TTB 58
4.3.4 TT-10P 60
4.3.5 his-6P 60
4.3.6 PrrnB 62
4.4 Experiments 62
4.4.1 Promoters 63
4.4.2 PCR 63
4.4.3 Ligations 63
4.4.4 Transcription Experiments 64
Transcription Buffer 64
Transcription Reactions 64
Assay Protocol 64

5 Results **65**
5.1 Fluorescence Detection 65
5.1.1 Plates 65
Dynex High Binding Plates 66
Dynex Non-treated Polystyrene 67
Costar Polypropylene Plates 69
Corning NBS Plates 71
Greiner Fluotrac 200 Plates 71
5.1.2 Machine 73
Plate Background 73
Pyrrolo-dC Fluorescence 76
5.2 Transcription Experiments 77
5.2.1 PrrnB 77
5.2.2 BB-R-P 79

6 Discussion **83**
6.1 Potential Problems 83
6.1.1 Template DNA 83
6.1.2 Fluorescence Measurements 84
6.1.3 Reaction Conditions 84
Temperature 85
Salt Concentration 85
Enzyme Concentration 85
NTP Concentration 86
Other Conditions 86
6.2 Future Work 86
6.2.1 Method Soundness 86

6.2.2 Reaction Conditions 87
6.2.3 Strand Dependence 87
6.2.4 Transcription Regulators 87
6.2.5 Real-Time Kinetics 88
6.2.6 Pyrrolo-dC Characterization 89
6.2.7 Polymerase States 89
6.3 Method Comparison 90
6.3.1 Advantages 90
6.3.2 Disadvantages 91
6.3.3 *in vivo* Methods 91
Reporter Proteins 92
RNA Assays 92
6.3.4 *in vitro* Methods 93
Fluorescent Measurements 94
6.3.5 Computational Methods 94
6.3.6 Single Molecule Techniques 95
Single Molecule Transcription 95
Single Molecule PAR 96
6.4 Conclusion 96

References **97**


Cheese-Meter

Chapter 1

Introduction

1.1 Motivation

The desire of engineers is to design and build complex systems. With the most complex systems existing in Nature, understanding, simplifying, and engineering living systems is an irresistible challenge. Both biologists and engineers have begun the process of engineering computational units into living cells [38, 49, 156, 168].

However, the discipline of engineering synthetic biological systems to date has been lacking several important tools [143]. Especially desirable is a system of modularity, with methods for designing, assembling, and characterizing biological modules. Although some progress has been made in these areas, most items remain unsolved. This thesis addresses the specific problem of module characterization, in order to create modules useful for engineering complex biological systems.

Not only is there no biological equivalent of a voltmeter, there is not even a definition of a standard volt unit. As most biological networks begin with transcription, transcription is a natural boundary for describing modules. We propose a standard unit of RNA polymerase arrival rates (PAR), a measure of the rate that RNA polymerase arrives at a location on the DNA, to define the boundaries between modules.

Modules can be defined as having some number of input PARs and some number of output PARs. Transfer curves that relate the outputs as a function of the inputs can then be used to completely specify the behavior of a module. In addition, modules can be connected interchangeably as the input and output for all modules is specified in identical units.

PARMESAN is a method for measuring polymerase arrival rates in an *in vitro* fluorescence assay. It relies on a DNA base analogue, pyrrolo-dC, that shows differing fluorescence when in single-stranded versus double-stranded DNA. RNA polymerase temporarily separates the two strands of DNA during transcription, making single-stranded DNA from double-stranded DNA. By incorporating pyrrolo-dC at specific positions in the DNA, fluorescence changes during transcription reactions can be used to measure the rate of RNA polymerases arriving at that location.

1.2 Thesis Organization

Chapter 2 contains background information related to the transcription process and existing methods to measure transcription. Chapter 3 describes modules based on PARs, the PARMESAN method, and the theory motivating the measurements. Chapter 4 discusses the details of the actual methods used and Chapter 5 contains some results. Finally, Chapter 6 contains an analysis of the progress made and a comparison with other methods.

1.3 Abbreviations

The following abbreviations will be used in this thesis:

Standard abbreviations:
A: adenine
bp: base pair
BSA: bovine serum albumin
C: cytosine
DNA: deoxyribonucleic acid
dNTP: deoxyribonucleoside triphosphate
dsDNA: double-stranded DNA
G: guanine
nt: nucleotide
NTP: ribonucleoside triphosphate
oligo: oligonucleotide
PCR: polymerase chain reaction
RNA: ribonucleic acid
RNAP: RNA polymerase
ssDNA: single-stranded DNA
T: thymine

U: uracil

Specific to this thesis:
BB-F: BioBricks forward primer
BB-R-P: BioBricks reverse primer with 1 pyrrolo-dC
BB-R: BioBricks reverse primer
P: pyrrolo-dC
PAPS: polymerase arrivals per second
PAR: polymerase arrival rate
PARMESAN: polymerase arrival rate measurement method
TTB: Transcription Tester Bottom
TTT: Transcription Tester Top

Chapter 2

Background

In biological systems, the primary function of gene expression is the conversion of the genetic information stored as DNA into proteins, the major functional units that maintain a cell. RNA serves as an intermediate step between DNA and proteins. During *transcription*, the information stored in DNA is converted into RNA messages by the DNA-dependent RNA polymerase. Then through *translation*, RNA can be made into proteins by the ribosomes and associated machinery.

This chapter provides the background that forms the basis for the rest of the thesis, with an overview of the interactions between RNA polymerase and DNA during transcription and existing techniques for assaying transcription.

2.1 Transcription Overview

By convention, DNA bases are numbered relative to the location of transcription initiation. The initiation point is numbered $+1$, and bases that come after, also known as the downstream region, are numbered with positive numbers. Bases before the initiation point, the upstream region, are numbered negatively, with the base immediately preceding the initiation point numbered -1. There is no base numbered zero.

The standard convention used here is to print DNA from left to right in the 5' to 3' direction and for the top strand to contain the promoter. Thus, the top DNA strand is the expected *non-template* strand, and the RNA polymerase will use the bottom DNA strand as the *template* strand.

For studying the transcription process, much work has been done with T7 RNA polymerase, as it contains a single subunit and is well characterized. However, T7 promoters do not have much variation and their regulation is much less interesting. Unless otherwise specified, the RNA polymerase and systems described in this thesis all come from *Escherichia coli*.

2.1.1 RNA Polymerase

E. coli RNA polymerase is a complex enzyme with several protein subunits. The complete RNA polymerase has two α subunits, a β subunit, a β' subunit, and a σ subunit. The core RNA polymerase consists of all the subunits except the σ subunit, that is $\alpha_2\beta\beta'$, and is capable of synthesizing RNA from nucleoside triphosphate (NTPs) molecules using DNA as a template. The α subunits are required for assembly of the core enzyme and play a role in promoter recognition. The β and β' subunits are important for binding DNA nucleotides and in catalyzing RNA synthesis [119].

With a σ subunit attached, the RNA polymerase is known as a holoenzyme. The role of the σ subunit is to allow the RNA polymerase to selectively initiate at specific regions of DNA [101]. The core enzyme binds tightly and non-specifically to dsDNA. Holoenzyme, on the other hand, binds loosely to dsDNA and tightly to specific regions of DNA.

Multiple σ-factors are present in almost all bacterial genomes. The most abundant σ factor in *E. coli*, transcribing genes fundamental to cell function, is called σ^{70}, with the 70 representing its molecular weight in kilodaltons. The total molecular weight of the holoenzyme containing σ^{70} is around 450 kilodaltons [30].

Other σ-factors are activated in response to certain events in the environment. For example, σ^{32} is activated in response to heat shock and σ^{54} is involved in nitrogen regulation [120]. The expression of different σ-factors allows a biological system to switch resources to transcribing different sets of genes depending on the conditions.

Two commonly used drugs targeting RNA polymerase are heparin and rifampicin. Heparin sequesters free RNA polymerase [131], preventing binding or initiation by the polymerase. Rifampicin specifically inhibits bacterial RNA polymerases by binding to the β subunit [121]. In the presence of rifampicin, RNA polymerase cannot initiate transcription chains but can bind to promoters and elongate existing RNA chains [102]. Streptolydigin is another drug that inhibits RNA polymerase by decreasing the elongation rate.

2.1.2 Promoters

Sequences, known as promoters, can direct the initial binding of RNA polymerase to the DNA, affecting the rate of transcription. Many promoters of differing strengths exist in biological systems. Although *promoter strength* can be defined in a variety of ways, a general definition of a promoter's strength is the number of RNA transcripts it produces per second [20].

As increased or decreased gene expression levels can be harmful for an organism, transcription must be tightly regulated. In *E. coli*, the transcription initiation rate varies by four orders of magnitude [119]. Some phage promoters, such as those from T4, are much stronger in *E. coli* than any native *E. coli* promoters [137], indicating that the full range of promoter strength that could be used is not normally found in *E. coli*.

Through statistical studies of many *E. coli* promoters, two highly conserved 6 bp regions located around −10 and −35 have been found. The −10 consensus sequence is `5’-TATAAT-3’` and the −35 consensus sequence is `5’-TTGACA-3’`. The sequence of most promoters match at least 7 of the 12 bp in the consensus sequence [119]. Reducing a promoter's degree of homology to consensus generally reduces its strength. However, promoters matching the consensus sequence too closely may not be able to initiate transcription effectively due to the extremely stable binding of RNA polymerase [37].

Another crucial factor in determining promoter strength is the number of bases and the sequence between the −10 and −35 regions [7, 139]. The spacer region between the two regions is commonly about 17 bp in length, but can range from 15–21 bp [119]. Extra `TG` sequences located slightly upstream of the −10 have also been shown to enhance promoter activity [17].

The strength of promoters can also depend on upstream regions, called UP elements found from −40 to −60 [125]. The strongest UP elements derived from *in vitro* selection have an A+T-rich consensus sequence containing two A-tracts with an intervening T-tract. [42]. The downstream region (DSR), from +1 to +20 can affect promoter strength 10-fold both *in vivo* and *in vitro* [66]. Promoter strength due to variations in the downstream region depend on the RNA polymerase concentration.

The interactions between RNA polymerase and these regions of DNA has been studied extensively. The −10 and −35 recognition hexamers of promoters is recognized by σ^{70} and UP elements are recognized by the C-terminal domain of the α subunit [126]. Evidence indicates the polymerase primarily interacts with the non-template strand [18, 44].

In one study, specific bases on two promoters in the -10 and -35 regions have been shown to make physical contact with RNA polymerase, with the physical contacts showing more homology across promoters than the DNA sequence itself [136]. Others have found large differences in promoter activity due solely to the curvature of DNA between the -35 and -10 region [27]. A strongly curved region of DNA next to a weak -10 region can have promoter activity even without the -35 region. The strongest promoters appear to be those that have the minor groove in the middle of the -10 region located on the inside of a curve.

Promoter sequences for *E. coli* have been compiled in several publicly available databases [36, 89, 55].

2.1.3 Operators

Regulators of transcription have been studied heavily [11, 131]. Some regulator proteins, called activators, can bind to *operator* sites on DNA and increase the rate of transcription. Other proteins, called repressors, bind to operator sites and have the opposite effect, inhibiting transcription.

Activators and repressors can affect any of the steps in transcription initiation. The simplest mechanism of repression is to have an operator site that overlaps with the promoter. The repressor and polymerase compete for the same piece of DNA, effectively lowering the free promoter concentration [119].

2.2 Transcription Process

The standard model for transcription initiation involves the following steps [34, 101, 119]. First, RNA polymerase binds a promoter to form the closed complex. The closed complex is then turned into an intermediate closed complex. In the last step, an active open complex is formed, initiating transcription. After a short RNA chain is transcribed, the polymerase switches to stable elongation.

In theory, any of the steps could be rate limiting and, therefore, determine the strength of the promoter [66]. No obvious correlation has been seen between the stability of the RNA polymerase-promoter complex and the *in vivo* promoter activity [16, 69], so the binding of RNA polymerase is probably not the major determinant of promoter strength. Promoter melting is also not limiting, as tested with premelted, bulged double-stranded DNA [60]. For the *rrnB* promoter, experiments suggest that the dissociation rate of the open complex may be the rate determining step for transcription [11]. Other results suggest that the conformational change of the RNA polymerase may be the rate limiting step [44, 63, 60].

2.2.1 Closed Complex

Transcription begins with RNA polymerase binding to DNA to form the closed complex. The complex is called closed because the promoter DNA remains double stranded [119]. The closed complex results from the RNA polymerase binding to DNA covering the -55 to -5 region, but likely involves sequence-specific contacts at the -35 region only [54].

The next step involves a polymerase rearrangement, extending the downstream DNA footprint of the polymerase to about $+20$. The σ subunit is believed to play a role in DNA strand separation and the transition to the open complex [54, 140].

2.2.2 Open Complex

The open complex is formed when a region of DNA is "melted", i.e. the strands are separated. For *E. coli* RNA polymerase, the melted DNA region extends from -10 to $+1$ [109]. The adenine base at -11 on the non-template strand is believed to play a crucial role in forming the open complex [54].

The open complex can be stable in some promoters and only transiently stable in others with measured half-lives ranging from minutes to hours [16, 54].

2.2.3 Initiation

Most promoters appear to have a set of potential initiation start sites, with one predominating [119]. RNA polymerase strongly prefers to initiate with a purine, an adenine or guanine, with adenine preferred over guanine. To initiate transcription, two template DNA bases, two NTPs, and a magnesium ion need to come together with the RNA polymerase.

The initiation complex is unstable. The RNA polymerase may form and release short abortive RNAs, from 2–12 nucleotides long, before promoter escape occurs and the polymerase switches to a stable elongation mode. Even *in vitro* with excess of all NTPs, abortive RNA transcripts are made [119]. After about 10 to 15 nucleotides, the σ subunit may be released, with the core RNA polymerase stably elongating the RNA chain [20]. However, some evidence indicates the σ subunit is not released in the majority of transcription complexes and is part of the elongation machinery [10, 110].

The maximum transcription initiation frequency can be estimated at about 1 initiation per second, assuming an elongation rate of 50 nt/s and a space requirement of 50 bp for a RNA polymerase molecule [101]. Transcription initiation rates vary from about 1 initiation per generation for *lac* repressor to 1 initiation per second for the ribosomal promoter, PrrnB [31, 152].

2.2.4 Elongation

Whereas the RNA polymerase contacts about 75 bases of DNA at the promoter, during elongation, the polymerase only contacts 30 to 40 bp of DNA [121].

The locally melted region of DNA during transcription is called the transcription bubble. In the transcription bubble, part of the nascent RNA is paired with the DNA template. For *E. coli* RNA polymerase, the transcription bubble is estimated at 17 ± 1 bp with a 12 ± 2 bp RNA:DNA hybrid region [48]. However, considerable debate exists about the length of this RNA:DNA hybrid, with other estimates of 8–9 bp [72, 115, 145] to less than 3 bp [105].

Average elongation rates for *E. coli* RNA polymerase can range from 30 to 100 nt/s *in vivo* [109]. RNA polymerase *in vitro* synthesizes about 10 to 35 nt/s [149] and has been measured at 17 nt/s when transcribing T7 DNA [21]. Elongation does not proceed at a constant rate, with the discontinuous movement of RNA polymerase observable by microscopy [28].

Unlike DNA polymerases, RNA polymerase is processive, meaning that once the RNA polymerase dissociates from the RNA and DNA, it cannot reassociate and resume synthesis [76]. Processivity requires the elongation complex be extremely stable with a half-life greater than several hours [121], allowing the transcription of long operons, on the order of 10^4 nucleotides in bacteria. But processivity also means transcription termination can be a complex process requiring massive destabilization [151].

The movement of the RNA polymerase and melting of the DNA is chemically powered from energy stored in NTPs, releasing one pyrophosphate per nucleotide incorporated into the RNA chain [149]. The RNA:DNA hybrid may also have a role in stabilizing interactions during elongation [135].

2.2.5 Pausing

RNA polymerase pauses at points in some sequences, contributing significantly to the time to synthesize a complete transcript [121]. Pausing slows RNA synthesis and provides a chance for other cell machinery to interact with the polymerase [154]. RNA polymerase is quite stable when stalled on a DNA template *in vitro* [160].

A pause refers to the stopping of RNA polymerase for a temporary and finite period of time. A more permanent block to elongation is transcriptional arrest. An arrested polymerase cannot efficiently resume elongation without the help of other molecules [149] and may slip backwards along the DNA [71].

Pause and arrest sites can be characterized by the half-life and the recognition efficiency [149]. The half-life is the time required for half of the stalled polymerase to resume elongation and the recognition efficiency is the percent of polymerases becoming stalled at the pause site. Half-lives range from too short to measure to several minutes. Recognition efficiencies range from a few percent to 90%.

Two well-studied pause sites are attenuation control regions for the *trp* and *his* amino acid biosynthesis operons [77]. The leader regions for these operons include a pause site where the RNA polymerase pauses, allowing transcription to be regulated by the translation machinery [45]. At the *trp* pause site, the half-life of the pause is about 4 seconds at *in vivo* conditions [161].

Many pause sites, including the *trp* and *his* leader regions, can form an inverted repeat hairpin or short stem-loop structure. The hairpin interacts with RNA polymerase to mediate pausing [6], but the hairpin is not the only important structure. The relevant region of DNA affecting pause strength is probably between 38 nt upstream of the pause site to 14 nt downstream [22, 80]. The transcription bubble for the polymerase at the *trp* and *his* pause sites has been determined using the sensitivity of single-stranded DNA to DEPC or $KMnO_4$ [79].

2.2.6 Termination

The role of transcriptional terminators in biological systems is complementary to that of promoters [122]. Termination occurs when the RNA polymerase releases both the DNA template and the RNA strand.

Some DNA sequences require the presence of a hexameric RNA-binding protein called *rho* to terminate transcription [109]. Intrinsic terminators do not require rho and are called rho-independent terminators. Rho-independent terminators are similar to pause sites, usually containing an inverted repeat forming a stable stem-loop hairpin structure. Cells appear to use rho-independent terminators for routine termination and rho-dependent terminators for error correction and in special cases [122].

When the RNA polymerase reaches an intrinsic termination signal, the polymerase slows down and pauses [52, 78]. The stability of the complex decreases and depending on the sequence following the hairpin, the RNA polymerase can either fall off, terminating the RNA chain or continue elongating again. Studies suggest that termination efficiency depends on the elongation rate [160]. In general, termination may depend on the competition between the free energy for normal elongation and termination [151].

Unlike pause sites, terminators usually have many As on the template strand after the stem-loop, leading to a RNA sequence containing many uracil bases. However, mutating the downstream region of the *trp* pause site to contain a U-rich RNA region did not convert it to a terminator, indicating that more is involved to termination than a hairpin followed by many U bases [80].

Terminators also differ from pause sites with fewer bases between the hairpin signal and the site where the polymerase pauses. There are about 7 to 9 bases for terminators instead of 10 or 11 for pause sites [24].

2.3 Transcription Assay Techniques

Many methods have been developed to study the interaction of RNA polymerase and promoters. Some of the most common and useful methods are described here. General background on fluorescence is also provided, as fluorescence is an often used tool to study molecular activity, including RNA polymerase and transcription [25].

2.3.1 Fluorescence

Fluorescent molecules are called *fluorophores* and each fluorophore has a characteristic light excitation and emission spectrum [99]. The emission wavelength is necessarily longer than the excitation wavelength, with the emission energy less than the excitation energy. Fluorescent

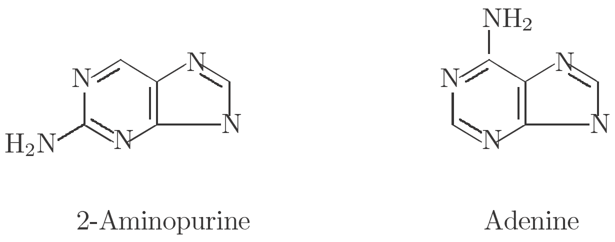


Figure 2.1: *2-Aminopurine is a fluorescent base useful for probing protein-DNA interactions that separate the strands of DNA.*

probes are easy to measure without significantly disturbing its environment.

Fluorescence is relatively sensitive with single molecule detection possible at picomolar concentrations. The fluorescence limits are due primarily to the background, such as Raman scattering from the solvent [134, 158]. Reduction in fluorescence can occur due to prolonged exposure to light, called fading or bleaching, or due to the presence of nearby molecules, called quenchers.

Other than measuring the fluorescent intensity of a fluorophore, there are other fluorescence techniques, such as fluorescence resonance energy transfer (FRET) or fluorescence recovery after photobleaching (FRAP) [99]. Another technique is to measure fluorescence anisotropy, also known as fluorescence polarization, which measures the rotation of molecules, with a molecule in a complex expected to rotate more slowly than when in solution alone [56].

The fluorescence of the normal nucleic acid bases is low, and, in proteins, only tryptophan usually has significant fluorescence. Thus, to study protein and DNA interactions, non-natural fluorescent probes are often used.

A fluorescent probe often used is the base analogue 2-aminopurine (2-AP) that is identical to adenine except the amino group is moved from the 6- to the 2- position (Figure 2.1). 2-Aminopurine substitutes for adenine and base pairs with T only slightly weaker than the A-T pair. In addition, the fluorescence of 2-aminopurine is sensitive to whether it is base paired. 2-Aminopurine has been used as a sensitive, real time method to measure the kinetics of *E. coli* RNA polymerase binding with DNA [44, 142], for measuring the kinetics of T7 RNA polymerase binding [58, 148], and for measuring DNA helicase activity [118].

2.3.2 *in vivo* Methods

Methods for measuring transcription in cells usually either focus on measuring protein or RNA levels.

Reporter Proteins

Reporter proteins are easily detectable from the rest of the cellular soup, making them used frequently for studying transcriptional activity. Many proteins are available with different characteristics [4, 83].

Common methods for detecting the activity of reporter proteins include enzymatic assays and fluorescence measurements. Some examples of using reporter proteins to measure promoter strength include green fluorescent protein (GFP), alkaline phosphatase, luciferase genes [94], β-galactosidase (*lacZ*) [107], chloramphenicol acetyltransferase (CAT) [57], *galK* [1, 108], and *trpA* [40].

Enzyme Assays

One way to measure the strength of expression of a reporter protein is to provide a substrate that interacts with the gene to form a measurable product. A common reporter gene is *lacZ*, coding for the enzyme β-galactosidase [11, 86, 107]. Various substrates can be provided that form colored products when hydrolyzed by β-galactosidase.

Other enzymatic assays exist to measure expression of β-lactamase, chloramphenicol acetyltransferase, and luciferases [43]. For example, firefly luciferase reacts with ATP and oxygen to produce light.

The most sensitive assays include chemiluminescent detection of alkaline phosphatase with a detection limit of 10^3 molecules and luciferase with a detection limit of 10^5 molecules [4]. β-Galactosidase has a detection limit on the order of 10^9 molecules.

Fluorescence Assays

Green fluorescent protein (GFP) from *Aequoria victoria* is extensively used as a reporter protein *in vivo*. The fluorescence of GFP can be measured easily. Using fluorescence-activated cell sorting (FACS) along with GFP, it is possible to select from a random library for promoter sequences [146]. The strength can then be measured using flow cytometry.

RNA Measurements

Protein reporters are an indirect measure of transcription. A more direct measurement of transcription would be to measure the RNA levels in a cell. PCR-based methods to quantitatively measure RNA levels include real-time reverse transcription PCR [127] and titration of the target RNA using competitive PCR [12].

Another approach is to label all RNA by incubating a culture briefly with radioactive [^{3}H]uridine [159]. It is then possible to hybridize with a known sequence and measure the amount of hybridization to determine the expressed RNA level.

One problem with using mRNA is that varying degradation rates can interfere with measurements. The half-lives of mRNAs can range from 40 seconds to 20 minutes [74]. When cloning a promoter region, the 5' transcription initiation point is often not known, and the sequence on the 5' end can have a large effect on RNA stability [13]. Mutated tRNAs or tRNAs from other organisms have been used as better reporter RNAs [93, 116]. The tRNAs are more stable than mRNAs as they fold to become resistant to degradation and mutations can be introduced to knock out any tRNA functional activity.

Relative Strengths

Measuring promoter strength relative to a fixed promoter can normalize for some experiment variability. One method for determining relative promoter strength is from the ratio of the activities of two promoters expressing different reporter proteins, such as β-lactamase under control of a test promoter and β-galactosidase under control of a fixed promoter [137].

Similarly, promoter activities can be measured as a ratio of RNA synthesis in relation to a standard under control of a fixed promoter. For example, it is possible to characterize promoter strengths relative to the β-lactamase promoter (P_{bla}) [31, 68]. In addition, transcription rates can be measured by taking samples at different times and plotting a time course of the relative mRNA expression [150].

The absolute promoter activity is more difficult to measure. For the *rrnB* promoter, the absolute activity was measured *in vivo* by taking the number of *rrn* transcripts per minute per unit of culture mass divided by the number of *rrn* genes per unit of culture mass [169, 170]. The measurements rely on the observation that the *rrn* operons produce stable rRNA present at a relatively constant fraction of the total RNA. To calculate absolute activity of *rrn* promoters, the number of copies of the gene is calculated based on the gene's location relative to the replication origin and the replication speed.

The absolute transcription activity of other promoters, measured as the number of transcription initiations per second, can be estimated based on their strength relative to the *rrnB* promoter [31, 84]. For example, the *lac* promoter can be estimated to have a transcription initiation frequency of about 4.6 initiations per minute.

2.3.3 *in vitro* Methods

Promoter measurements done outside of cells and in well-defined solutions, usually involve *in vitro* transcription and a way to assay the RNA products formed [147]. The major benefit of *in vitro* transcription is the ability to control the conditions of transcription.

For example, several methods have been used to synchronize transcription initiation, such as preincubating all components except NTPs for several minutes or preincubating all components except $MgCl_2$ and initiating synthesis by adding $MgCl_2$ and rifampicin [51]. By controlling when transcription begins, the effects of transcription can be more easily assayed.

RNA Measurements

A straightforward method for studying transcription *in vitro* is to set up a transcription reaction, stop the reaction after some amount of time, and run the product on a gel to quantify the amount of RNA generated. It is even possible to measure the strengths of several promoters in the same solution by having each transcribe a RNA sequence of a distinct size [65].

Radioactive Labeling Methods

RNA levels *in vitro* can be measured using radioactively labeled NTPs [97]. By using a standard DNA template such as T7 DNA, properties of RNA polymerase such as specific activity and elongation rate can be measured [21, 119].

To measure transcription initiation rate, experiments can be done with excess active RNA polymerase and a sufficiently short transcribed sequence. A short transcribed sequence ensures that the elongation rate will be quick in comparison to binding and initiation, and, thus, the incorporation of radioactive NTPs over time becomes a measure of the initiation rate.

The abortive initiation method also measures transcription initiations, and relies on only adding the first two NTPs needed by RNA polymerase to extend the RNA chain [100, 103]. Thus, the RNA polymerase would initiate RNA synthesis but would abort soon after due to the lack of NTPs. The rate of initiations can be measured by following radioactive ^{32}P labeled NTP.

Fluorescent Measurements

The abortive initiation method can be simplified by using fluorescent labeling rather than radioactive labeling. ANS (1-amino-naphthalene-5-sulfonate) is an example of a fluorescent label that has been used [14, 131]. The ANS fluorophore attached to the γ-phosphate of a NTP is a good substrate for *E. coli* RNA polymerase, with about 60-80% of the activity of an unmodified NTP [164]. In addition, when RNA polymerase cleaves the α-β phosphoryl bond during insertion of a NTP into a RNA strand, a change in the fluorescence spectrum is measurable, thus providing a continuous assay for transcription [130].

Another fluorescence method uses FRET to measure absolute distances during protein interactions with nucleic acids [87]. For example, the movement of RNA polymerase along DNA can be measured with FRET [110], by attaching one fluorophore to the polymerase and another fluorophore to the end of the DNA. Kinetic analysis of the polymerase movement should be possible with this method.

Multiple fluorescence probes can be used simultaneously, as long as the excitation and emission wavelengths do not overlap significantly. A real-time *in vitro* system has been developed using three different fluorescent probes that measures three properties simultaneously: protein/DNA complex formation, transcription bubble formation, and RNA production [33]. Tetramethylrhodamine is used to label the 5' end of DNA such that its fluorescence polarization properties change as a result of RNA polymerase binding the DNA. The transcription bubble is measured using 2-aminopurine that changes its fluorescence intensity when the strands of DNA are pulled apart. The third assay for RNA production is done with the ANS-like fluorophore, 5-amino-2-naphthalenesulfonic acid, connected to the γ-phosphate of the NTPs. This system has the advantage of obtaining information about binding and transcription simultaneously.

Measuring DNA Binding

Although the strength of DNA binding by RNA polymerase is usually not an accurate predictor of promoter strength, the binding characteristics does contribute to the behavior of a promoter. Methods to measure binding include inserting a fluorescent tryptophan analogue into σ^{70} and measuring its quenching due to binding [18], measuring the change in intrinsic protein fluorescence due to binding [58], and release assays measuring the stability of the RNA polymerase and DNA complex at points along a DNA strand [5].

Another method is the filter binding technique, using nitrocelullose fiters that bind protein-DNA complexes and not free DNA to assay for RNA polymerase-promoter complexes. Labeled DNA can be mixed with polymerase and the extent of binding can be measured [101].

Binding strengths can also be measured by a template competition assay [63]. The binding on a test promoter can be compared with another piece of DNA by initiating one round of transcription from both and comparing the relative transcription from each template.

With solid-phase transcription, the strength of polymerase binding to arbitrary regions of DNA can be measured by "walking" the polymerase to the desired location on a known sequence of DNA. The polymerase is hexahistidine-tagged and immobilized on Ni-NTA beads [70]. Immobilization allows for walking the polymerase to arbitrary positions on the DNA by alternating between providing limited NTPs and washing the beads to remove old NTPs.

Chapter 3

PARMESAN Theory

In this chapter, we present the theory motivating the experiments that follow to measure transcription events. The objective is to design, assemble, and characterize synthetic biological systems in a modular fashion. Modular design is facilitated by abstracting module inputs and outputs as RNA polymerase arrival rates. PARMESAN, an *in vitro* fluorescence assay, is one potential method that can be used for measuring polymerase arrival rates.

3.1 Modularity

A key feature necessary for designing and building complex biological systems is *modularity*. Modularity simplifies the process of engineering, facilitating reuse and abstractions. Biological components should be capable of being built as interchangeable modules.

Requirements for modularity include a module *design* strategy, a module *assembly* strategy, and a module *characterization* strategy. The choice of one of these strategies influences the choice of the others, and so these strategies cannot be selected independently.

The primary focus here is on module characterization, with the goal of simplifying the design and assembly process. Modules will be assumed to be specified entirely as a single linear sequence of DNA. Although other types of modules are possible, DNA modules are currently the easiest to work with in practice.

3.1.1 Characterization

Modules need to be characterized and their behaviors specified in a way useful for module and system design. There are potentially many ways a module can be characterized. Some desirable properties for a module characterization system are:

- *Abstraction.* Module characterization should promote the goals of modularity. A non-abstract characterization would be the DNA sequence of the module itself. Although the sequence, in theory, contains all the information about the module within it, it lacks a useful abstraction layer needed for modular design.

- *Generality.* All modules should be characterized in a similar fashion, independent of what is inside the module. It is not general to characterize one module in terms of protein expression level and another module in terms of DNA binding strength.

  Another way to define generality is that the characterization strategy should not limit the type of modules allowed during module design. For example, characterizing modules as protein levels immediately rules out all modules not expressing proteins but operating at the RNA level.

- *Connectivity.* The modules should be characterized in a way allowing for connections among modules. Connectivity means that an output for one module needs to match the input of another module. This allows for both easy design and assembly of modules.

- *Usefulness.* The characterization of a module must be useful in connecting and designing complex systems. An example of a probably non-useful characterization of a module is the %GC content, as it does not contribute much to the goal of assembling systems from modules.

- *Measurable.* Requiring that modules be characterized with properties that cannot be measured is not useful. To be practical, the measurements should be easy to perform, not time consuming, accurate, and repeatable.

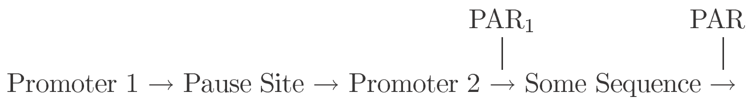


Figure 3.1: *The two polymerase arrival rates in this example integrate information from multiple promoters and pause sites. RNA polymerase can initiate transcription from one of the two promoters and travel from left to right, arriving at the* PAR$_1$ *and* PAR$_2$ *locations on the DNA at some rate.*

3.2 PAR

3.2.1 Definition

The proposed method for characterizing modules is to use RNA polymerase arrival rates as inputs and outputs of the modules. The polymerase arrival rate (PAR) is the number of times a polymerase arrives at a certain position on the DNA per second, with units of polymerase arrivals per second (PAPS).

The polymerase arrival rate integrates information from an entire system, taking into account polymerase binding rates, transcription initiation rates, elongation rates, and unbinding rates. For example, in Figure 3.1, the polymerase arrival rate, PAR$_1$, depends on a variety of factors. Promoter 1 may be strong allowing polymerases to initiate transcription quickly, but the pause site may slow down those polymerases and kick some of them off the DNA. Promoter 2 may also allow some polymerases to bind and initiate. The number of polymerases arriving at PAR$_1$ must take into account the number of polymerases initiating from promoter 1 that pass the pause site and the number of polymerases initiating from promoter 2. The polymerase arrival rate PAR$_2$ at a location further downstream may not be the same as PAR$_1$ depending on the intervening DNA sequence. PAR$_2$ depends on the elongation rate through that intervening sequence and the rate of polymerases falling off, in addition to everything that determined PAR$_1$.

3.2.2 Modules

Figure 3.2 shows the general form for a module containing some number of input polymerase arrivals, INPARs, and some number of output polymerase arrivals, OUTPARs. A module that has i inputs and o outputs will be denoted as an $\{i/o\}$ module. The inputs and outputs of modules should be able to connect to other modules independently of each other. So the two outputs of a $\{1/2\}$ module could be connected to

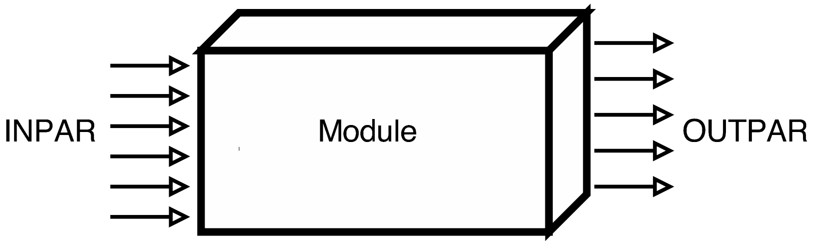


Figure 3.2: *Modules will have some number of inputs and outputs measured in PARs. This is an example of a* $\{6/5\}$ *module.*

the inputs of two different modules or to the inputs of a single module. Some example modules are given below.

Examples

Figure 3.3 shows a $\{1/1\}$ module with a single input and output. The polymerases arriving at its input lead to transcription and production of the transcriptional regulator protein `A`. Protein `A` regulates polymerase binding and initiations from its regulatory region, causing polymerases to leave the module at a certain rate. Note that the polymerases leaving the module at the output are not necessarily the same polymerases that entered the module. The regulatory region for `A` does not necessarily have to, and should not, be physically located immediately after the coding sequence for `A`.

If the regulator protein `A` in Figure 3.3 were a repressor, then the output PAR signal would be inverted from the input PAR, and the module would implement a logical NOT function. If the regulator protein were an inducer instead, then the output PAR could be amplified based on the input, and the module would act as an amplifier.

The important `Insulator` module in Figure 3.4 can be considered a $\{1/0\}$ module. This module contains a transcriptional terminator that fixes the output at zero polymerase arrivals and so can be considered to not have an output. To make the module even more useful, the terminator is made bidirectional so that in both directions, the output PAR is zero. The bidirectional transcriptional terminator could be a standard hairpin terminator with the correct bases on both sides to terminate transcription going in either direction. In effect, polymerases are blocked from going through the module in either direction.

Other $\{1/0\}$ modules may be useful to convert non-transcriptional events into PAR units, relying on non-transcriptional side effects for

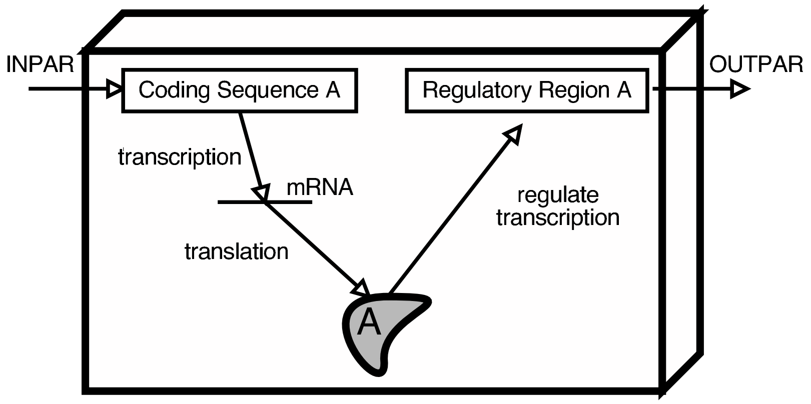

INPAR
Coding Sequence A
Regulatory Region A
OUTPAR
transcription
mRNA
regulate
transcription
translation
A

Figure 3.3: *This simple* {1/1} *module has a regulatory protein mapping the input PAR to the output PAR.*

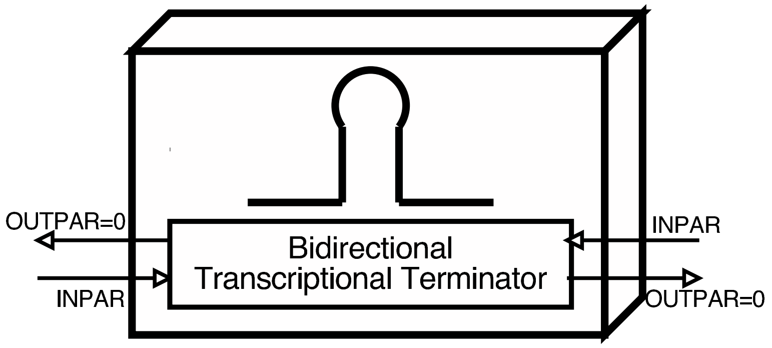

OUTPAR=0
INPAR
Bidirectional
Transcriptional Terminator
INPAR
OUTPAR=0

Figure 3.4: *An example of an* `Insulator` {1/0} *module consisting of a bidirectional transcriptional terminator. The output PAR on both sides is zero regardless of the inputs, thus insulating the modules on either side from each other.*

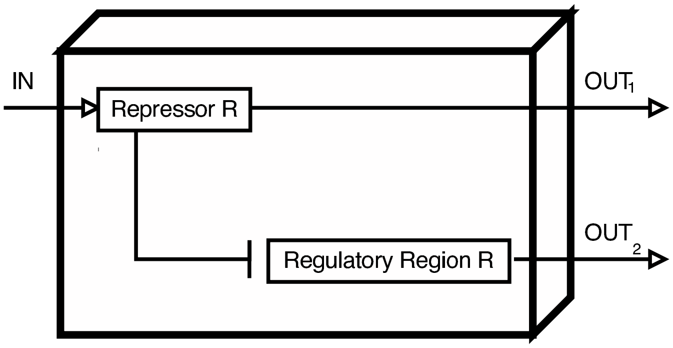


Figure 3.5: *A {1/2} inverting module where* OUT_1 *=IN and* OUT_2 *= NOT(IN).*

their primary function. For example, a {1/0} reporter module may contain only the coding sequence for a reporter protein like GFP. The input PAR is translated into some amount of GFP fluorescence. Modules of this sort are useful in designing systems that need to have an effect that is not purely transcriptional. A PAR-based signal is mapped into some other type of signal, for example protein levels.

A {0/1} module would have no inputs and a single constant output PAR as it is a function of no inputs. The typical {0/1} module consists of a single un-regulated, constitutive promoter connected to the output. The promoter initiates transcription at a fixed rate, leading to a constant output PAR.

Several slightly more complex modules are shown in Figure 3.5 and Figure 3.6. The {1/2} module in Figure 3.5 is an inverter that also passes through the input PAR as an output. Polymerases entering from the input travel through the module and exit as OUT_1, with the OUT_1 PAR roughly equal to the input PAR.The output PAR at OUT_2 is the logical NOT of the input.

Figure 3.6 shows a {2/1} module where the output PAR is the logical NOR of the two inputs. The two inputs both cause the same repressor protein to be transcribed and translated. As the repressor negatively regulates the output, if either input is high, the output will be low. The output will be high only if both inputs are low.

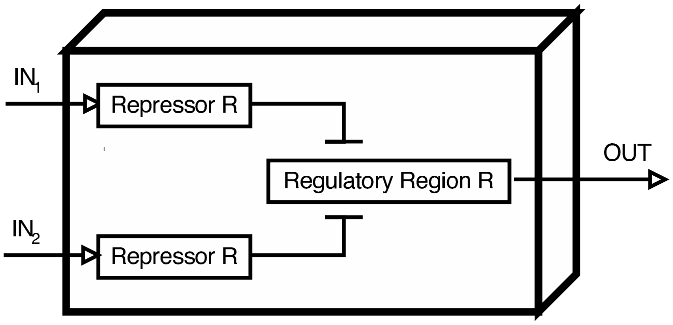


Figure 3.6: *A {2/1} module implementing the NOR logic function.*

3.2.3 Assumptions

Several important assumptions are made in order to effectively use PAR-modules.

- All desired modules can be defined in terms of input and output transcriptional activity, in the form of polymerase arrival rates as shown in Figure 3.2.

- Each molecule of RNA polymerase, at any time, is in one of a finite number of states. For example, the states may include: unbound, bound to promoter, paused, arrested, or stably elongating.

  One state, denoted as s_0, is defined as the normal state. For our purposes, a stably elongating RNA polymerase is the normal state. Module inputs and outputs are defined in terms of the rate of polymerases in state s_0 arriving at the input boundary of the module or leaving at the output boundary.

  We assume polymerases in other states either do not arrive at the inputs and outputs or, if they arrive in another state, they have no effect on the perceived behavior of the module.

- To ensure modules can be arbitrarily connected together, the input and output PARs are assumed to be relatively independent of the sequences coming before and after them. In other words, at the output locations, there should be little look-ahead by the polymerase, with the transcription rates being relatively independent of the downstream DNA sequence. At the inputs, the polymerase should behave independently of the sequence that comes before it.

For example, as it is known that promoter strength can depend on a large number of bases both upstream and downstream of the transcription start, extra bases may need to be inserted before and after some promoters during module design. The extra bases would serve an insulating function for the promoters near the inputs or outputs of the module, guaranteeing that their behavior does not change significantly due to another sequence being attached to the module.

- The rates are assumed to reach a steady-state relatively quickly and maintain that state indefinitely. In addition, an output PAR should be a deterministic function of the input PARs. This is probably not a necessary assumption as one could imagine probabilistic outputs or outputs that vary with time. However, making the output a deterministic function of the inputs simplifies the discussion of the model and is a useful abstraction.

- It is required that the outputs do not depend on any variables other than the inputs, allowing us to treat the module as a black box. Any module can, therefore, be abstracted or defined as a set of transfer curves that specify the outputs as a function of the inputs.

  Of course, in practice, biological systems have little insulation, and so there will be external effects on the system. The problem of cross-talk or interference among modules is an important issue, but it will not be addressed here. The goal is to minimize as many unwanted interactions as possible either by specifying them as an input PARs or by clever design of the module.

3.2.4 Motivation

PAR is defined in terms of a physically well-defined property of the system rather than based on some indirect and relative measurement of transcription. Relative definitions of transcription strength such as those measuring RNA produced or protein expressed often are in arbitrary units such as amount of GFP fluorescence or enzymatic activity relative to another promoter. With relative definitions, measurements of transcription only have meaning when compared with other measurements of the same type. PAR is an intrinsic feature of the biological system that is defined independently of how it is measured.

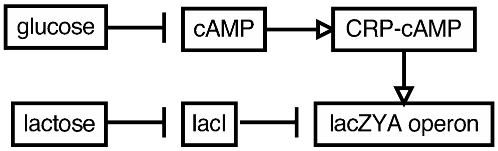


Figure 3.7: *A conventional model representing a biological network of the* lac *operon. The pointed arrows represent positive regulation and the straight line heads represent negative regulation.*

Although the polymerase arrival rate concept uses transcription rates to characterize modules in an unconventional way, specifying modules as PAR provides many advantages for the engineer over the standard biological models.

Conventional Models

Figure 3.7 shows a conventional network diagram for the well-studied *lac* operon and its regulation by glucose and lactose [111]. The boxes in the diagram represent nodes and different arrows are used to represent positive or negative regulation. For example, the repressor LacI negatively regulates the genes in the *lac* operon. High levels of the protein LacI lead to low levels of transcription for the lacZYA genes.

Although these models may be useful to biologists studying a fixed system, it is not at all useful for the engineer trying to build a new system. Conventional biological networks do not use a common unit for specifying connections, so the nodes are not modular and cannot be "rewired" in any meaningful way. The inputs and outputs for the nodes are usually either not specifically defined or are defined in incompatible units, such as in levels of different proteins.

For example, the output for the lacI node is the level of LacI protein and the output for the CRP-cAMP node is the level of the CRP-cAMP complex, making these two nodes not interchangeable even though they both have the same number of inputs and outputs. In addition, the negative regulation connections between lacI and lacZYA and between lactose and lacI may look the same in the diagram, but the underlying mechanism is completely different. Some interactions involve transcriptional regulation, other interactions involve allosteric control, and there is no way to know the difference from the network diagram, making it difficult to reuse the modules in another system.

Another more detailed type of standard model for biological systems involves differential equations with many different rate constants. To obtain values for the rate constants often requires many different types of measurements. For example, obtaining a model for a transcriptional event may require several techniques to measure the binding affinity and the initiation rate. But these techniques are of no use for measuring other events such as translation. As the purpose here is not to understand the detailed mechanism for a process, all these rates add complexity without providing a convenient abstraction layer.

It is preferable to have a layer of abstraction similar to Figure 3.7. For seeing the overall relationships in a biological system, this traditional level of abstraction is sufficient. But for designing and engineering synthetic systems, a different type of model is needed, such as one based on polymerase arrival rates, that provides for consistency and modularity.

Rates vs. Levels

The conventional boundary for a module is at the level of translation and the signal conveying information is often taken to be the level of protein expressed. Based on the biological mechanism, this makes sense, as the amount of protein is usually directly related to the functional expression level. For example, the expression of the *lac* genes depends on the total amount of the repressor LacI and not the rate at which LacI expression is changing.

However, as discussed above, using protein levels leads to incompatible inputs and outputs among modules. One module with an output in terms of one protein level is incompatible with another module that may also be described by protein levels but by a different protein.

A similar problem occurs when using RNA transcript levels as the signal connecting modules. To allow arbitrary connections among modules, inputs and outputs need to be generic and not based on a particular RNA sequence or protein. Thus, instead of levels, the rate that a common molecule, such as RNA polymerase, crosses between modules is used as the signal.

In a dynamic system, rates can convey more meaningful information than an amount at a single point in time. We can think of PAR as analogous to electrical current, with the RNA polymerase analogous to charge. The RNA polymerase is used as the common signal transferred among the modules. The absolute number of RNA polymerase in a module is not important. The rate that polymerases travel between modules provides the critical link needed to connect modules.

Transcription vs. Translation

PAR is based on transcription events as opposed to some other biological event such as translation. Instead of polymerase arrival rates, it is possible to imagine using a similar rate for translation, the ribosome arrival rate. However, there are several reasons why the analogous approach for translation does not work as well as for transcription.

Whatever is used to define modules immediately restricts the type of modules allowed. If modules are defined at the translation level, then everything that does not get translated is immediately ruled out. Transcription, the process of making RNA from DNA is common to the expression of all genes, but not all genes are necessarily translated. In addition, anything that does not operate at the level of translation cannot be used in translational modules. For example, LacI is a transcriptional repressor and so could not be used as an output in a translational module system. None of the parts in the *lac* operon from Figure 3.7 operate at the level of translation and so none of those parts could be utilized as useful modules if modules are defined around translation boundaries.

In a sense, using modules based on translation provides a higher layer of abstraction than modules based on transcription. Translation implicitly assumes transcription happens first, requiring abstracting the details of transcription inside modules. But the lower layer of abstraction provided by transcriptionally-based modules also brings more freedom and power for the module designer, with fewer abstraction requirements.

Another important consideration is what type of modules need to be built and used. For building interesting networks, modules should perform some type of control decision, so we will focus on regulatory modules. Although translation can be regulated, most regulation in existing biological systems appears to be transcriptional. One advantage of translational over transcriptional regulation is faster switching of expression levels. Transcriptional regulation is limited by the mRNA half-life, which can present an undesirable time-lag [138].

The major issue with using translation is that it much more difficult to separate into modules compared with transcription. Most regulators are proteins that themselves have to be transcribed and translated and, for these and other proteins, the desired protein sequence should be expressed. The problem with translation is that most regulation occurs around the start codon, which conflicts with the goal of controlling the exact protein sequence to be translated.

The basic mechanisms of translational regulation can be similar to transcriptional regulation. Translational initiation is the step usually regulated and there are translational operators. For translation to initiate, the ribosome binding site (RBS) needs to be physically near the AUG start codon. There are usually only 5 to 9 nucleotides separating the RBS and the start codon [138].

Translational operator sites that are targets for regulatory proteins often are found immediately upstream of the translation initiation site [138]. Operators can include the RBS and the first codons for the protein to be translated. The mechanism of translational control can involve important changes in the secondary structure of the RNA near and including the initiation site.

With PAR-modules, insulating sequences are inserted to guarantee that the inputs and outputs do not depend on the surrounding sequences. These insulating sequences can lead to an extra transcribed RNA sequence. However, this extra RNA does not affect translation, so the final translated protein is exactly the desired protein. The same cannot be done easily in the case of translational modules. Inserting insulating sequences would add unwanted initial amino acids to the front of the desired protein.

For example, Figure 3.8 shows a connection between two modules. Module X contains a regulatory region that is controlled in some fashion, with the regulatory region overlapping the initiation point shown with the arrow. The output of Module X will be either the initiation of transcription or translation. Module Y contains a protein coding sequence that should be placed under the control of the output from Module X.

The non-buffered scheme for connecting modules in Figure 3.8(a) allows transcription or translation to start exactly at the beginning of the sequence in Module Y. However, as part of the regulatory region for Module X extends past the initiation point, the initial sequence from Module Y becomes part of the regulatory region, affecting the behavior of Module X. This scheme is not usable due to its lack of modularity, as the behavior of Module X is dependent on the module that is connected to it.

Figure 3.8(b) takes care of the modularity problem by adding an extra buffer region between the initiation point and the desired coding sequence. If the modules were transcriptionally-based, then there would be an extra 5' untranslated region of RNA attached to the coding sequence RNA. In most cases, this untranslated region is not important, as the ribosome will find and translate the correct protein coding sequence. However, if the modules are translationally-based, then the

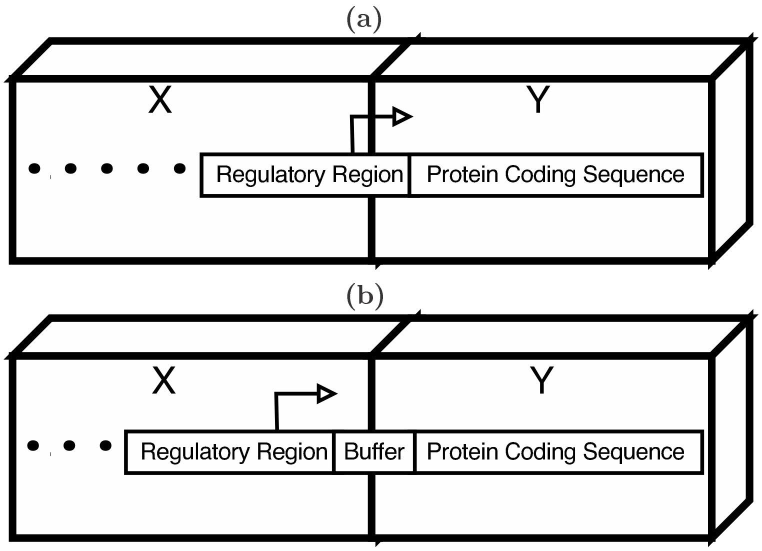

(a)
X
Y
Regulatory Region
Protein Coding Sequence
(b)
X
Y
Regulatory Region
Buffer
Protein Coding Sequence

Figure 3.8: *Two ways of connecting modules X and Y are shown. In (a), Module Y is directly connected to Module X, and in (b), a buffer region is inserted between the two modules. Scheme (a) is not modular, as part of the regulatory region for X depends on the sequence that it is connected to in Y. Scheme (b) is more modular but requires transcribing/translating the extra buffer region in between.*

buffer region would lead to extra amino acids being inserted at the N-terminus of the protein. Thus, to get the desired protein to be translated, while maintaining modularity, it is necessary that the modules work at the transcriptional level.

Modularity

The most important motivating factor for the use of polymerase arrival rates has been the benefits from modularity. PAR has the identical meaning across different types of modules, because a polymerase arriving to transcribe one gene is equivalent to a polymerase arriving to transcribe another gene.

Using polymerase arrival rates satisfies many of the desirable properties for a module characterization system:

- *Abstraction.* Using PAR effectively abstracts away the inner workings of the module, allowing a module to be specified as transfer curves mapping inputs to outputs.
- *Generality.* Transcription is fundamental to most cell processes. The functional parts inside a cell, whether it be protein or RNA, are all transcribed at some point. However, this approach is not completely general as modules can only be defined around transcription boundaries. Protein interactions can not be explicitly specified as inputs or outputs and must be hidden within modules.
- *Connectivity.* Modules need to connect with each other. The easiest way to ensure connectivity among modules is to have the inputs and outputs for modules be in the same units. As all inputs and outputs are defined in terms of the same polymerase arrival rates, with the polymerase always arriving in the same state, one module that outputs x polymerase arrivals per second can be connected to another module that accepts an input of x polymerase arrivals per second.
- *Usefulness.* Several possibilities for designing and modeling systems characterized by polymerase arrival rates are described later.
- *Measurable.* A method to measure polymerase arrival rates forms the majority of the rest of the thesis.

To increase the generality and usefulness of PAR-systems, it is feasible to have non-transcriptional events interact with PAR-modules. For example, although module inputs and outputs were specified as PARs, it would be straightforward to generalize modules to have inputs and outputs that could be in different units in addition to PARs.

For example, a {0/1} module could have an output that depended on the concentration of some molecule in the environment. The module would have no input PAR, but would have an input determined by the concentration of another molecule. These types of modules could increase the generality of PAR-based systems by allowing modules to map non-PAR signals into PAR signals to be used by other modules.

Examples of real biological systems that have been separated into transcriptional modules include the phage λ lysogenic and lysis decision pathway and part of the eukaryotic MAPK signaling pathway [26]. The MAPK pathway is particularly interesting, as it involves many kinases that function non-transcriptionally, but by defining module boundaries around transcription events, useful abstractions can still be formed.

3.2.5 Design and Assembly

Using PAR-modules only makes sense if it facilitates the design and assembly of systems that need to be built. Characterizing modules with polymerase arrival rates influences module design and assembly in a variety of ways.

Design

Modules must be designed around transcriptional events, as inputs and outputs are defined in terms of RNA polymerase arrival rates. Depending on the system that needs to be built, this may limit the inherent modularity possible. As all non-transcriptional events must be inside a module, if there are many non-transcriptional events, then modules may need to be larger, and therefore less modular, than desired.

Also, modules need to be designed carefully to limit undesirable interactions that affect the behavior of the module in an undefined manner. This may require adding insulating sequences within the module or describing the conditions under which a module can be used successfully with the specified behavior.

Assembly

The desired end system needs to be physically assembled into one piece of linear DNA that contains all the modules connected correctly. To facilitate assembly of modules, which may have many inputs and outputs to be connected together, every input and output for a module should be in its own separate tube of DNA *if possible*. Connecting the output of one module to the input of another module would then involve ligating the two corresponding pieces of DNA together using standard methods.

For some modules, it may not be possible to separate an input and output into different tubes. For example, for the module in Figure 3.5, OUT_1 cannot be disconnected from the input as the polymerases arriving at this output come directly from those that arrive at the input. On the other hand, OUT_2, can be separated and placed into its own tube as it does not need to be physically near the repressor DNA for it to function correctly.

This inseparability of some inputs and outputs puts restrictions on how some modules can be connected, due to the limitations of physically assembling the DNA. For example, OUT_1 above cannot be connected back to the module's input, as this would require circularizing the DNA. Circular DNA cannot be ligated with other pieces of DNA or cloned into plasmids so is not practical or useful. More complex cases also exist where the outputs cannot be connected to some inputs, because assembly would require circularizing the DNA.

Aside from the above restriction, modules that have separable inputs and outputs can be designed and assembled at will, leading in the end to several pieces of DNA that can be put together in any order. The module designer ensures that inputs and outputs from different modules can be connected together by inserting insulating sequences within the module. The module assembler, on the other hand, needs to ensure that the behavior of the system is not changed by assembling the pieces of DNA and, so, needs to insert insulating sequences between modules.

To ensure that the partially assembled input and output pieces can be connected together, the `Insulator` from Figure 3.4 or a similar module should be inserted between each piece of DNA. The role of the `Insulator` is to guarantee that polymerase arrivals from one piece do not carry over to the next, allowing the system to be assembled to the design specifications.

The linear piece of DNA representing the assembled system needs to then be cloned and maintained, probably on a plasmid vector. Again, to ensure that the sequences on the plasmid do not affect the behavior of the system, insulating transcriptional terminator sequences are needed on both ends of the system. Figure 3.9 depicts an overview of the assembly process.

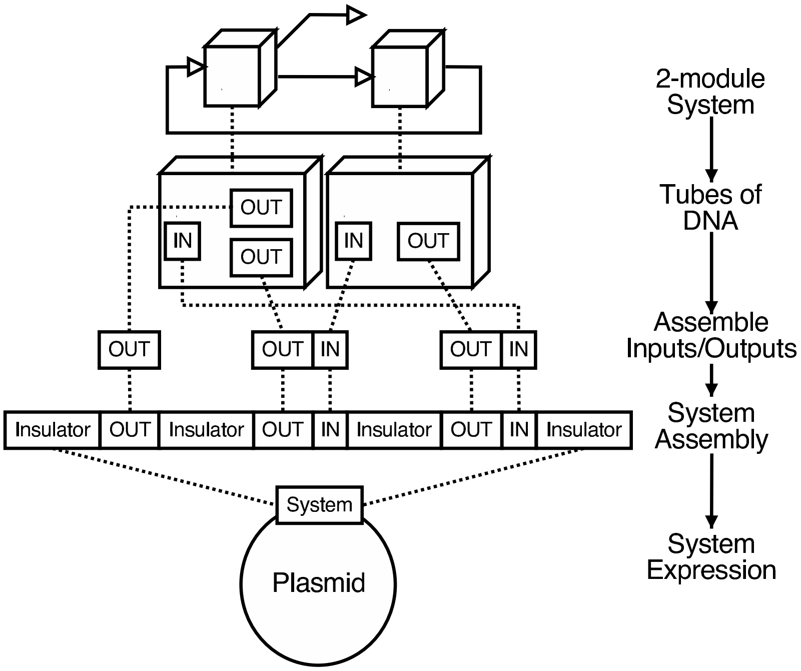

OUT
IN
OUT
IN
OUT
OUT
OUT
IN
OUT
IN
Insulator
OUT
Insulator
OUT
IN
Insulator
OUT
IN
Insulator
System
Plasmid
2-module System
Tubes of DNA
Assemble Inputs/Outputs
System Assembly
System Expression

Figure 3.9: *An example 2-module system is assembled from a schematic of how the modules should be connected. Each module has its inputs and outputs physically separated as tubes of DNA. The connections between modules are made by ligating the corresponding input and output tubes. The pieces are then assembled with additional insulating sequences, cloned into a plasmid, and expressed in a cell.*

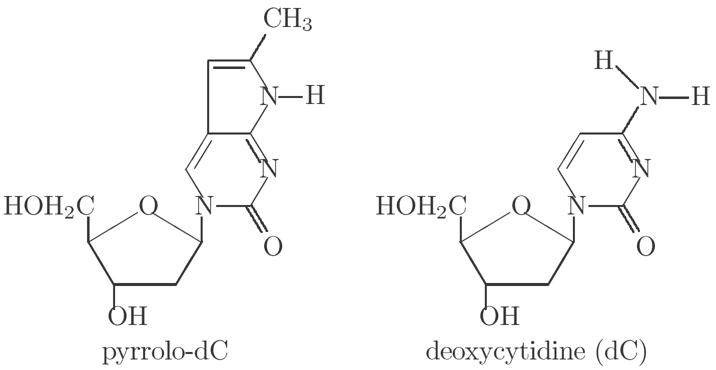


Figure 3.10: *The chemical structure of pyrrolo-dC allows it to substitute for dC. Both dC and pyrrolo-dC can base pair with guanine by forming 3 hydrogen bonds.*

3.3 PARMESAN

For PAR-modules to be practical, a method is needed to measure polymerase arrival rates. PARMESAN is a technique for measuring polymerase arrival rates and stands for Polymerase Arrival Rate Measurements in (En) Standard Assay *iN vitro*. The proposed method involves the use of a fluorescent DNA base, pyrrolo-dC, that is sensitive to the local melting of DNA.

3.3.1 Pyrrolo-dC

Pyrrolo-dC shown in Figure 3.10 is a synthetic base able to form 3 hydrogen bonds with guanine and, thus, is able to substitute for cytosine in DNA [91, 92]. In addition, pyrrolo-dC has different fluorescence properties when it is in single-stranded or double-stranded DNA. This effect may be partly related to hypochromism, the effect of stacked bases in nucleic acids absorbing less light than unstacked bases.

The fluorescence of pyrrolo-dC in ssDNA is approximately double its fluorescence in dsDNA. The fluorescence when pyrrolo-dC is in a mismatched base pair in double-stranded DNA (e.g. paired with adenine) is even higher than in ssDNA. Also, the pyrrolo-dC fluorescence is low when in a RNA:DNA heteroduplex [92]. These fluorescence properties of pyrrolo-dC are summarized in Figure 3.11.

Although 2-aminopurine has similar properties and been used much more extensively than pyrrolo-dC to study proteins that separate the

DNA pairing	*Fluorescence*
dsDNA mismatched	High
ssDNA	Medium
dsDNA or RNA:DNA	Low

Figure 3.11: *The relative fluorescence intensity of pyrrolo-dC in DNA varies depending on its environment.*

strands of DNA, the excitation spectrum of 2-aminopurine overlaps with the intrinsic fluorescence of proteins making subtracting background measurements necessary [148].

Pyrrolo-dC has excitation maxima at 260nm and 350nm and an emission maximum at 460nm, far from the protein and nucleic acid fluorescence range [92]. The shifted spectrum of pyrrolo-dC should in theory lead to less background than 2-aminopurine.

The letter P will be used to represent pyrrolo-dC in sequences.

3.3.2 Transcription Assay

To measure polymerase arriving at a specific location, pyrrolo-dC is incorporated into the desired DNA region. At time 0, RNA polymerase is added and the fluorescence change is measured over time. As RNA polymerase melts the DNA and the transcription bubble passes through the pyrrolo-dC location, the fluorescence of the base will change. Thus, the change in fluorescence measures polymerase arriving at the location in the DNA marked with the pyrrolo-dC.

The method could also be called “polymeracing,” as we are measuring the time it takes for polymerases to reach a certain point, much like in a race.

Theory

The rate of transcription initiation (V) can be written as follows [119]:

$$V = V_{max}\frac{[R]_F}{K_m + [R]_F}$$

$[R]_F$ is the concentration of free, unbound RNA polymerase, K_m is the RNA polymerase concentration when promoter activity is half V_{max}, and V_{max} is the maximum initiation rate with saturating RNA polymerase, when $[R]_F \rightarrow \infty$.

At low polymerase concentrations, when $[R]_F \ll K_m$, $V/[R]_F = V_{max}/K_m$, and this ratio is a measure of a promoter's strength [170]. Therefore, to avoid saturating promoters, a low concentration of polymerase relative to DNA should be used.

To determine actual concentrations to use, the *in vivo* conditions can be approximated. In a cell, estimates of $[R]_F$ range from 10nM to 100nM and a promoter DNA concentration of about 8nM [119, 131]. The effective $[R]_F$ *in vivo* will be much less due to competition with other pieces of DNA.

Measurement Issues

Pyrrolo-dC bases can be inserted on either the template or the non-template strand, leading to potentially large differences in the fluorescence effect due to transcription.

An increase in fluorescence is only seen if a DNA region with pyrrolo-dC is separated and no RNA:DNA hybrid is formed. If the pyrrolo-dC is located on the template strand, then a RNA:DNA hybrid could quench the fluorescence, with increased fluorescence only when the pyrrolo-dC is in the transcription bubble but not paired with RNA.

The length of the RNA:DNA hybrid was determined using pyrrolo-dC for T7 RNA polymerase [92]. The transcription bubble for T7 RNA polymerase was determined to collapse close to the exiting RNA, meaning that the length of the RNA:DNA hybrid is about the same size as the transcription bubble. This means fluorescence from pyrrolo-dC bases on the template strand will almost always be quenched, either being in a DNA:DNA duplex or RNA:DNA heteroduplex.

For *E. coli*, there is disagreement about the length of the RNA:DNA pairing in the transcription bubble, and consequently, the amount of unpaired DNA on the template strand. If the length of the RNA:DNA hybrid is small, then there would be a large number of exposed bases on the template strand in the bubble, providing more time for a fluorescence signal to be seen.

One proposed model for RNA organization during elongation provides for 10 unpaired nucleotides on the non-template strand and only 1 nt on the template strand that is unpaired with either RNA or DNA [72]. Other estimates include 3–6 nt of single-stranded DNA on the template strand upstream of the RNA:DNA hybrid [73].

Unlike the disagreement about bases on the template strand, the bases on the non-template strand in the transcription bubble are oriented away from the enzyme, with the bases being susceptible to nucleases [153]. Thus, pyrrolo-dC bases should be unstacked on the non-template strand, causing a change in fluorescence during strand separation by RNA polymerase. When possible, it is preferable to insert the pyrrolo-dC on the non-template strand rather than the template strand.

3.4 Applications

As part of the motivation driving the development of PARMESAN, applications related to synthetic biology can be facilitated by the use of polymerase arrival rates to characterize modules. Some examples are given here, but the full potential for applications in this area has yet to be thoroughly explored.

3.4.1 Promoter Library

The promoters used here are all from natural biological systems. But one can also imagine creating a collection of random sequences of DNA and characterizing their strength as promoters, thus creating a promoter library. From a promoter library, one could pick out promoters of a desired strength depending on the application.

Simple modules could also be generated using the library of promoters and designing transcriptional factors that bind to those promoters as either repressors or inducers, perhaps by using synthetic zinc-finger proteins [32, 133].

3.4.2 Controlling Gene Expression

A library of promoters and modules could provide biologists more control over the expression of target genes. For simple cases, a gene can be placed under the control of characterized promoters from the promoter library, and, then, the effect of the gene on the biological system can be studied.

For more complex systems, there may be several target genes, each of which needs to be controlled. Building regulatory networks that could control desired genes would provide biologists an incredibly useful tool for studying biological pathways and gene interactions.

The currently existing methods to regulate expression are relatively crude. For example, inducible promoters such as the *lac* promoter are often used to regulate a gene's expression by varying the level of an external inducer, such as IPTG for the *lac* promoter. This is useful when a gene's expression level needs to be varied dynamically and quickly. However, induction may have unintended effects, such as activating the cell's native promoters, and there exist only a handful of known and useful inducible promoters, limiting the applicability of controlling a wide variety of genes simultaneously.

In addition, the level of induction is not quantitatively known or comparable across systems. For example, the level of IPTG used does not provide useful information about the physical expression level of the gene and cannot be compared with the level of another inducer, such as arabinose in a different system. This may partially explain why current experiments mostly involve either the over-expression or the knockout of certain genes, instead of looking at the entire range of expression possible.

It may be preferable to fix the expression of a set of genes to known static levels. A library of constitutive and characterized promoters of constant strength would make this a much easier task. The expression level for any number of genes can be simultaneously fixed, and the promoter strengths would be comparable across genes and systems.

For dynamic control, inducible or other more complex modules can still be used, if they are characterized with PAR as a standard unit of measurement. Standard characterization allows for the ability to compare results under different inducers and systems, as polymerase arrivals for one gene is identical to polymerase arrivals for another gene. Two genes with the same input PAR may be transcribed or translated with different efficiencies, leading to different protein expression levels, but they can still be characterized and controlled in the same way.

3.4.3 Transfer Curves

Provided a module specified as input and output polymerase arrival rates, transfer functions or curves can be calculated to describe the behavior of the module. For any output, a transfer curve exists to specify the output value as a function of the inputs.

It is desirable to have an easy method for measuring the transfer curves that specify a PAR-module. With a direct measure of PAR and a library of promoters of varying strength, it becomes possible to sample points along the curve.

The inputs of a module can be fixed to known values using promoters with characterized PARs. The output PAR of the module can then be measured, obtaining a single point on one transfer curve. The entire transfer curve can be interpolated by measuring as many points as desired, and used for modeling, comparisons, biological insights or for other purposes.

3.4.4 Modeling Biological Networks

Many interesting biological phenomena can be modeled using a transcriptional rate model and by abstracting away all non-transcriptional events into modules [26]. The previous work used the concept of transcription initiations per second (TIPS), which is similar to polymerase arrival rates, but PAR is more general and applicable to a wide range of situations. "Initiations" can be misleading as the polymerase may not be initiating anything.

Using modules such as in Figure 3.2 and with accurate measurement methods for transfer curves, complex systems can be easily modeled. In theory, the transfer curves completely describe the behavior of modules, making simulation straightforward [26]. In addition, as the entire transfer curve can never be known exactly in a real system, simulations can be done relatively accurately even with only limited points sampled from the transfer curve.

3.4.5 Synthetic Biology

A current engineering challenge is to build reliable and functioning synthetic biological networks that perform some function, such as calculating a logic operation or responding to inputs. A critical component missing to facilitate this form of engineering is the knowledge of how to put the right components together.

One way to design systems is to use evolution as a tool for finding the correct parameters [168]. Although evolution is a powerful tool, it would be preferable to be capable of rationally selecting and combining the correct components to make a system work from the start.

To design and build a system that has a reasonable chance of functioning correctly requires fully characterized components. A quantitative characterization of the behavior of all biological components from ribosome binding sites to promoters to protein sequences is essential. PARMESAN is one potential method for standardly quantifying the strength of promoters and general modules by measuring the ability to promote transcription activity.

The benefits of modularity quickly become obvious as more complex systems are being built. One potentially desirable application is the construction of digital logic circuits using biological components. To do this would require a collection of modules that implement a variety of logic functions and the ability for the modules to work correctly when assembled together. PAR-modules provide many benefits for modularity and the construction of such systems.

Chapter 4

Measurement Methods

The focus of the experimental work has been on applying the PARMESAN method to single promoters as a means for measuring polymerase arrival rates.

4.1 Protocol Overview

The following is the standard methodology used to measure fluorescence and transcription levels.

1. Synthesize standard oligos with pyrrolo-dC.
2. Use the standard oligos to incorporate pyrrolo-dC into promoters or other DNA regions to be measured. The two main ways of attaching a pyrrolo-dC are through ligation or through PCR.
3. Perform an *in vitro* transcription reaction by adding purified RNA polymerase.
4. Measure the fluorescence change after addition of the polymerase.
5. Interpret the data as polymerase arrivals at the location of the pyrrolo-dC.

4.1.1 BioBricks

BioBricks is an example of a modular assembly system and a collection of modular parts [15]. BioBricks specifies a standard prefix and suffix,

```
    EcoRI  NotI    XbaI            SpeI    NotI  PstI
5’ GAATTCGCGGCCGCTTCTAGAG-Module-TACTAGTAGCGGCCGCTGCAG 3’
3’ CTTAAGCGCCGGCGAAGATCTC-Module-ATGATCATCGCCGGCGACGTC 5’
```

Figure 4.1: *The BioBricks standard form for biological modules includes several restriction sites, allowing for structured assembly.*

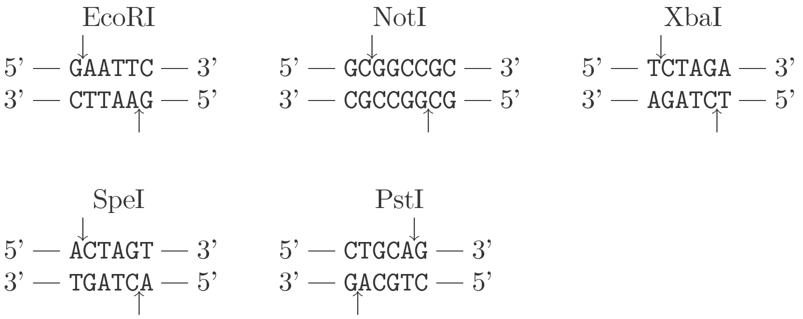


Figure 4.2: *The restriction enzymes in the BioBricks standard form recognize and cut the above DNA sequences.*

allowing for assembly using restriction enzymes and standard molecular cloning techniques.

The BioBricks prefix and suffix contains several restriction enzyme sites as shown in Figure 4.1 and Figure 4.2. Any module conforming to the BioBricks standard must contain the given set of restriction sites before and after the module and not contain any of those sites within the module.

For this work, all promoters and other pieces of DNA to be measured are assumed to be in a standard BioBricks format. Pyrrolo-dC oligos are synthesized that allow for the incorporation of the pyrrolo-dC into all BioBricks-conforming modules.

4.2 Materials

Klenow exo- DNA polymerase, T4 polynucleotide kinase, T4 ligase, restriction enzymes, and bovine serum albumin (BSA) were from New England Biolabs. PCR SuperMix was from Invitrogen.

Purified *E. coli* RNA polymerase holoenzyme was obtained from Epicentre. The holoenzyme contained the major σ^{70} subunit and came in a concentration of about 0.4μg/μl (870nM). NTPs were also from Epicentre.

Fluorescence was measured on a Bio-Tek FL600 fluorescence plate reader using fixed wavelength filters.

4.2.1 RNA Polymerase Purification

Other than using the commercially available polymerase, a higher concentration of *E. coli* RNA polymerase was purified using a histidine-tagged α subunit [144]. The purification protocol involved expressing the plasmid pHTT7f1-NHalpha, containing a hexahistidine tag on the N-terminus of the α subunit, in a BL21(DE3) (Novagen) strain. After induction with IPTG, the α subunit was isolated using a Ni-NTA spin column (Qiagen).

The remaining subunits were isolated separately using the plasmids pHTT7f1-sigma (σ^{70}), pMKSe2 (β), and pT7beta' (β') in the BL21(DE3) strain. After induction, inclusion bodies were collected, washed, and then solubilized with guanidine hydrochloride as a denaturant.

The α, β, and β' subunits were mixed together and dialyzed using a Slide-a-lyzer (Pierce) to remove the denaturant. The σ subunit was dialyzed separately. After dialysis, the σ subunit was mixed with the other subunits and incubated at 30℃ to allow the formation of the holoenzyme. The reconstituted holoenzyme was then purified through a Ni-NTA spin column.

The final solution was concentrated with a Microcon YM-100 (Millipore). Protein concentrations at each step were measured using the Bradford assay (Pierce).

4.3 Oligos

4.3.1 Oligo Synthesis

Oligonucleotides were synthesized on an ABI 394 synthesizer using standard phosphoramidite chemistry. Table 4.1 shows the sequences for several of the oligos synthesized.

For synthesis of oligos containing pyrrolo-dC, anhydrous acetonitrile was added to pyrrolo-dC-CE phosphoramidite (Glen Research) to a 0.05M concentration. A standard 40nmoles CE cycle was used for the synthesis. After synthesis, the oligos were deprotected in ammonium hydroxide at room temperature for 24 hours. The liquid was then evap-

Oligo Name	*Sequence*
BB-F	`5’-GCGCGAATTCGCGGCCGCTTCTAGAG-3’`
BB-R	`5’-GCGCCTGCAGCGGCCGCTACTAGTA-3’`
BB-R-P	`5’-CTGCAGCGGCCGCTAPTAGT-3’`
BB-R-5P-TE	`5’-CCGCAGAAAGGCCCACCCGAAGGTGAGCC ...` `... CTGPAGPGGPPGPTAC-3’`
TTT	`5’-ACACACACACACACACACACGTCTAGAPG-3’`
TTB	`5’-ACACACACACACACACACAPGTCTAGACG-3’`
TT control	`5’-ACACACACACACACACACACGTCTAGACG-3’`
TT complement	`5’-CGTCTAGACGTGTGTGTGTGTGTGTGT-3’`
TT-10P-top	`5’-CTAGCPPPPPPPPPPAAATACAGGAGTGCGG-3’`
TT-10P-bottom	`3’ GGGGGGGGGGTTTATGTCCTCACGCC-5’`
his-rev-2P	`5’-CGCTACTAGTTCTGAATGTCTTPPAGGAGAGATC-3’`

Table 4.1: *Sequences for several synthesized oligos.*

orated at 55℃, the oligo resuspended in 1ml TE (10mM Tris pH 8.0, 1mM EDTA), and stored at -20℃.

4.3.2 BB-R-P

One method of incorporating pyrrolo-dC into sequences is to run a PCR with primers containing pyrrolo-dC. The PCR primer BB-R-P, shown in Table 4.1 includes one pyrrolo-dC and was designed to match the BioBricks suffix. A PCR with BB-R-P and a forward primer incorporates the pyrrolo-dC on to the template strand immediately after the test promoter.

Another primer, BB-R-5P-TE, also matched the BioBricks suffix but contained 5 pyrrolo-dC and the TE terminator from T7. The terminator was to both slow down the polymerase and to terminate transcription normally rather than have the polymerase fall off the end of the DNA.

4.3.3 TTT/TTB

The sequences for Transcription Tester Top (TTT) and Transcription Tester Bottom (TTB) are identical except for the location of a single pyrrolo-dC (Table 4.1). Two other related oligos, a TT control oligo without pyrrolo-dC and an oligo complementary to all of these sequences, were also synthesized.

(a)

```
                        XbaI
5' ACACACACACACACACACACGTCTAGAPG...................
3' ...................GPAGATCTGCACACACACACACACACACA
```

(b)

```
5' CTAGAPGTGTGTGTGTGTGTGTGTGT 3'
3'     TGCACACACACACACACACACA 5'
```

Figure 4.3: *(a) TTT has self-complementary 3' ends allowing it to anneal. Extension of the oligo, shown in dots, forms a completely double stranded oligo. After extension, digestion with XbaI forms two identical pieces shown in (b). By ligating with another piece of DNA, a single pyrrolo-dC can be incorporated on the top strand.*

TTT and TTB were designed to be made double stranded by self-annealing and extension. For example, TTT shown in Figure 4.3(a) has self-complementary 3' ends. To prepare TTT, an extension is done with the self-primed TTT.

Extension was done using the Klenow fragment DNA polymerase and only dTTP and dGTP in the reaction. After extension at 37℃, a digestion with XbaI forms the two identical pieces in Figure 4.3(b), making purification unnecessary.

Starting with TTT, the final double stranded product contains a single pyrrolo-dC on the top strand. TTB is identical to TTT except the pyrrolo-dC in the final product ends up on the bottom strand instead of the top strand.

Note that there was no heat inactivation step after the extension. Heating the oligo before cutting in order to kill the DNA polymerase was experimentally determined to not give good results. As the oligo is perfectly self-complementary, a single strand can fold over on itself, forming a hairpin, making it impossible to cut. After digestion with XbaI, both XbaI and the DNA polymerase are heat inactivated. The DNA polymerase is not able to fill in the overhang during the digest as it requires dCTP. Thus, not inactivating the DNA polymerase before the restriction digest is not a problem.

To attach either TTT or TTB to a promoter, the promoter was cut with SpeI. TTT and TTB have compatible ends and can be ligated to the end of the promoter. After ligation, one pyrrolo-dC will be incorporated either into the non-template or template strand after the promoter DNA.

```
<--------------------- his-top-6P ----------------------->
    65                           bubble               116
    |                        <------------>            |
    v                 ***          * * *  !            v
CTAGCACCATCATCCTGACTAGCGCTTPAGGPGATPTPTPPTGGAAGACATTCAGACG
    GTGGTAGTAGGACTGATCGCGAAGTCCGCTAGAGAGGACCTTCTGTAAGTCTGC
    <------------------ his-bot ------------------------->
                                <----- his-rev-2P ----->
```

Figure 4.4: *The two oligos his-top-6P and his-bot anneal to form a fragment of the* his *pause site with 6 pyrrolo-dC bases on the non-template strand.*

4.3.4 TT-10P

The two strands of TT-10P (Transcription Tester with 10 pyrrolo-dC), were synthesized as the two separate oligos in Table 4.1. An NheI overhang was added to one end and a high GC content was used at the other end to clamp it tightly, reducing the chance of polymerase binding to the ends. The top, non-template strand contains ten pyrrolo-dC bases, and the remaining sequence was randomly chosen.

4.3.5 his-6P

A pause site was synthesized as a means of slowing down the polymerase at a specified location, allowing for easier fluorescence measurements. Two oligos, his-top-6P and his-bot, were created to allow for the insertion of the *his* pause site containing 6 pyrrolo-dC bases.

The sequence for the *his* pause site was from [22]. The bases are numbered as they are in other experiments with +1 referring to the start of the wild-type *his* transcript. The sequence from 65–116 was used and is shown in Figure 4.4.

The pause site, indicated with an exclamation point, is located after the T at 102 and before the G at 103. The sequence included 14 bp downstream of the pause site, as this is the region determined to affect the pause strength [80]. The experimentally determined transcription bubble at the pause site is shown [79].

Several changes were made from the native *his* pause sequence. The mutations from the wild-type sequence are indicated with an asterisk on the top strand and given in Table 4.2. The complementary mutations were made on the bottom strand. The native pause sequence contains a SpeI site in the hairpin loop that was mutated out. Also, mutations were added in the bubble region to increase the number of possible places to insert pyrrolo-dC into the region.

Mutation	Pause Change
TCT →CGC (top of hairpin)	110%
G96 →C	110%
G98 →C	130%
G100→C	110%

Table 4.2: *Mutations were made to the wild-type* his *pause sequence. Each individual mutation has been shown to increase the pause strength [22].*

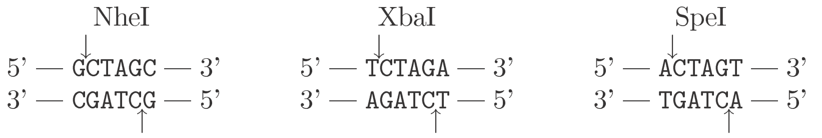


Figure 4.5: *NheI leaves compatible overhangs with XbaI and SpeI.*

It is desirable for the pause site to be as strong as possible. Ideally, RNA polymerase would completely stop at a fixed location containing pyrrolo-dC. Each of the mutations made, at least individually, has been shown experimentally to increase the strength of the pause site [22]. The previously determined change in pause strength are given in Table 4.2.

The NheI restriction enzyme creates compatible overhangs with XbaI and SpeI (Figure 4.5). The his-6P oligos were designed with the NheI site pre-cut with a CTAG overhang, allowing for ligation with another strand of DNA cut with SpeI. In addition, if a his-6P oligo ligates to itself, it would re-form a NheI site, and if the SpeI cut DNA ligates to itself, it would re-form a SpeI site. Therefore, the ligation can be incubated in the presence of both SpeI and NheI to cut any his-6P or DNA that incorrectly ligates to itself.

Another oligo, his-rev-2P (Table 4.1), was synthesized to be used as a reverse PCR primer. After ligation with his-6P, it is possible to PCR with his-rev-2P to incorporate 2 pyrrolo-dC on the template strand at the pause site and the site that follows. If the polymerase pauses at the expected pause site, then the region of DNA around the pause site would be melted and a fluorescence change should be detectable.

A SpeI site was also added to the 5' end of his-rev-2P. After a PCR, the resulting fragment could be cut with SpeI and ligated with another BioBricks compatible component. This could be used to test the effect of downstream DNA regions on transcription rates or pausing.

```
   <------- rrnB-front-top ---------->
  -65                           -35
   |                             |
   v                             v
5’ GCGGTCAGAAAATTATTTTAAATTTCCTCTTGTCA  ...
3’ CGCCAGTCTTTTAATAAAATTTAAAGGAGAACAGT  ...
                            <---------

    <-------- rrnB-top-3P -------->
                       -10         +1                +19
                        |           |                 |
                        v           v                 v
... GGCCGGAATAACTCCCTATAAT GPGPPACC ACTGACACGGAACAACGGC 3’
... CCGGCCTTATTGAGGGATATTA PGPGGTGG TGACTGTGCCTTGTTGCCG 5’
    -rrnB-middle-bottom-><------- rrnB-bot-2P -------->
```

Figure 4.6: *The* rrnB *promoter was synthesized from 4 separate oligos with 3 pyrrolo-dC on the top strand and 2 pyrrolo-dC on the bottom strand. The pyrrolo-dC were placed within the region of DNA expected to be melted when RNA polymerase binds to form the open complex.*

4.3.6 PrrnB

To test directly the effect of pyrrolo-dC due to DNA separation, a promoter was synthesized containing pyrrolo-dC incorporated in the region melted during polymerase binding.

The promoter used was 85 bp from the *E. coli rrnB* promoter. The sequence from −65 to +19 was used. As the melted region in the open complex extends from −10 to +1 [109], several of the cytosines in this region can be replaced with pyrrolo-dC. The entire promoter shown in Figure 4.6 was synthesized as four separate oligos. By synthesizing it in pieces, in principle, variants of the promoter can be easily tested by replacing a part of the promoter with a different oligo.

To form the complete promoter, two oligos, rrnB-top-3P and rrnB-middle-bottom, were first separately phosphorylated on the 5’ end using T4 polynucleotide kinase. Then all four oligos were mixed together, allowed to anneal, and ligated together with T4 ligase. After ligation, Klenow DNA polymerase and dNTPs were added to extend and fill in the rest of the promoter.

4.4 Experiments

To incorporate the above pyrrolo-dC oligos into promoters, promoters were amplified via PCR, and then either another PCR or a ligation was used to insert the pyrrolo-dC oligo into the promoter DNA.

4.4.1 Promoters

Promoter sequences were taken from PromEC and checked for unwanted restriction sites [55]. All promoter sequences were the same length, 100 bp, and have at least 25 bp after the transcription start, ensuring that the initial melted region does not include the downstream pyrrolo-dC DNA. The actual promoters used included PaccA, Pada, PampC, Plac, PribAP2, and Pzwf, with the initial P in the names added to indicate “promoter.”

To amplify the promoters, specific primers were designed with BioBricks ends and a PCR done using *E. coli* genomic DNA as the template. The result was gel purified using Qiagen spin columns.

To test termination, some promoters were attached to the TE transcriptional terminator from bacteriophage T7 via ligation and PCR.

4.4.2 PCR

Standard BioBricks primers, BB-F and BB-R shown in Table 4.1 and corresponding to the forward and reverse BioBricks sequences, were used to PCR BioBricks sequences. After using specific primers to PCR the promoters above, another PCR with the BioBricks primers was done to amplify the promoters again using standard primers.

To incorporate pyrrolo-dC, BB-R-P was substituted for BB-R in the PCR. Thus, the results from a PCR using BB-R can be easily compared with a PCR using BB-R-P, providing a control for the pyrrolo-dC effect.

4.4.3 Ligations

Although the PCR method is an extremely simple and straightforward way of incorporating pyrrolo-dC, it has the disadvantage that the pyrrolo-dC can only be incorporated on the template strand. Ligations can place the pyrrolo-dC on either the template or non-template strands.

Several pyrrolo-dC oligos were designed to be incorporated by ligations. All oligos contained a SpeI compatible overhang. For his-6P and TT-10P, they were first phosphorylated on the 5’ end using T4 polynucleotide kinase.

Promoters were amplified by PCR as above and then cut with SpeI. After the digest, the DNA was purified using Microcon filters (Millipore) to both concentrate the DNA and to remove the short fragments. A ligation reaction was then set up with the SpeI cut promoter, the pyrrolo-dC oligo, and T4 ligase. Most ligations were allowed to run overnight at 16°C. The resulting mixture was filtered again through a Microcon filter and checked for correctness on a 4% agarose gel.

4.4.4 Transcription Experiments

For transcription experiments, the temperature in the plate reader was set to 31°C.

Transcription Buffer

The 4x transcription buffer used for the following results (same as in [33]) contained the following:

- 200mM Tris-HCl (pH 7.9)
- 600mM KCl
- 40mM $MgCl_2$
- 0.4mM EDTA
- 0.4mM DTT

Transcription Reactions

$80\mu l$ transcription reactions were done with the following:

- $20\mu l$ 4x transcription buffer
- $1\mu l$ 25mM each NTP mix
- rest water and DNA

Assay Protocol

The following is an example of a typical protocol used to determine fluorescence changes due to RNA polymerase:

1. The well on the plate is pre-treated with BSA in water and allowed to sit for at least an hour.
2. The well is emptied and allowed to dry.
3. The above $80\mu l$ transcription mix is placed in the well and the baseline fluorescence is measured.
4. $1.5\mu l$ of RNA polymerase holoenzyme from Epicentre is added to the well (about 16nM final concentration).
5. The fluorescence kinetics is measured from the well and compared with the baseline.

Chapter 5

Results

This chapter describes some selected results, many that were unexpected and still not fully explained.

5.1 Fluorescence Detection

Fluorescence was initially measured on a Bio-Tek FL600 fluorescence plate reader using an excitation filter of 360nm/40nm and emission filter of 460nm/30nm (center wavelength/bandwidth). These filters were chosen based on published data about the spectrum of pyrrolo-dC [91] and from the manufacturer (Glen Research). Fluorescence values are in arbitrary units and were usually normalized to the first fluorescence reading.

5.1.1 Plates

Some plates showed uneven background fluorescence between wells. It was discovered that at some wavelengths being used, dust can have a large impact. A wash with water and a drying of the plate could usually reduce the well to well variation in fluorescence.

The type of plate used was found to have an enormous effect on fluorescence readings. All 96-well plates used were black, in order to minimize the background fluorescence. Across different types of plates from different sources, large variations were seen in background fluorescence readings.

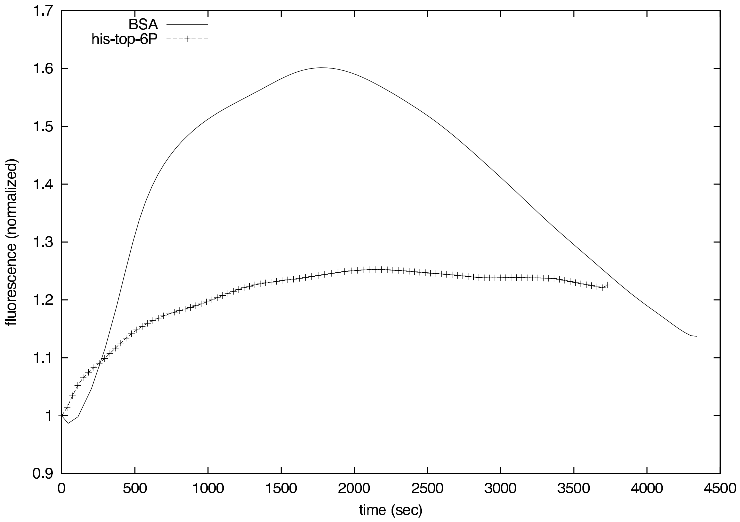


Figure 5.1: *Fluorescence kinetics of bovine serum albumin (BSA) and his-top-6P in water on Dynex high binding plates show unexplained increases and decreases in fluorescence.*

Dynex High Binding Plates

The first plates used were Dynex Microfluor 2 high-binding polystyrene plates. Transcription reactions done on these plates showed a curious effect. Kinetics measurements showed a noticeable bell-shape curve, with both a dramatic increase and decrease in fluorescence over time.

However, the bell-shaped curve was seen even when only RNA polymerase, BSA, or other proteins were placed into wells with only water. An increase in fluorescence but no decrease was seen when the oligo his-top-6P was used (Figure 5.1).

Protein or DNA could have been binding to the wells, leading to the observed increase and decrease in fluorescence. However, the excitation and emission wavelengths used should have been far from any intrinsic protein or DNA fluorescence. It is therefore not known why the fluorescence kinetics appear as they do. By varying the measurement times and durations, the decrease in fluorescence seen in the curve does not appear to be due to photobleaching.

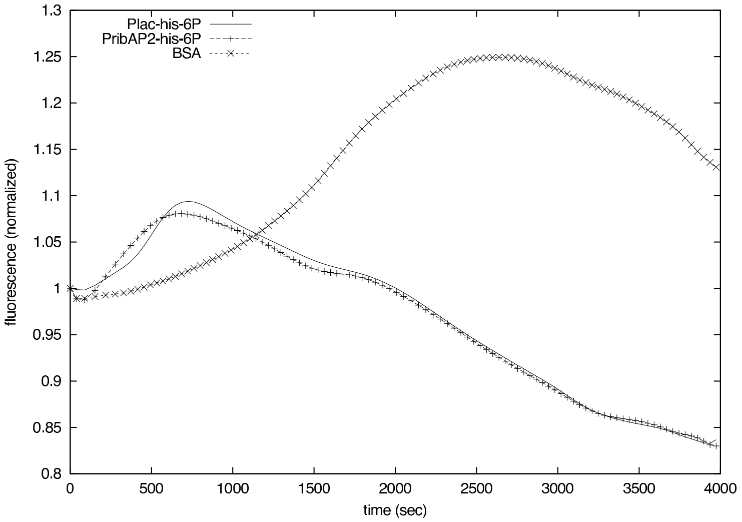


Figure 5.2: *Measurements on Dynex polystyrene non-high-binding plates show the same type of trend as seen on the high-binding plates.*

Dynex Non-treated Polystyrene

Another type of plates tried were Dynex clear bottom polystyrene plates, not treated to be high-binding and believed to be medium binding. With a clear bottom, fluorescence readings could be made from either the top or the bottom. However, bottom readings were much less sensitive, so top measurements were used for all experiments involving clear bottom plates.

Figure 5.2 shows the fluorescence measured for BSA in water. Although the magnitude of the increase was less than for the high-binding plates, a similar shaped curve is present on these medium-binding plates. Transcription reactions done with several promoters attached to his-6P also show upward and downward trends in fluorescence as shown in Figure 5.3.

Less fluorescence change is expected if a transcriptional terminator is added between the promoter and the pyrrolo-dC, due to fewer polymerase arrivals. Most polymerases would be expected to terminate before arriving at the pyrrolo-dC. In Figure 5.4, a promoter, terminator, and his-6P were ligated together and placed into a transcription reaction. However, even with the terminator, there is the characteristic increase and decrease in fluorescence after adding RNA polymerase.

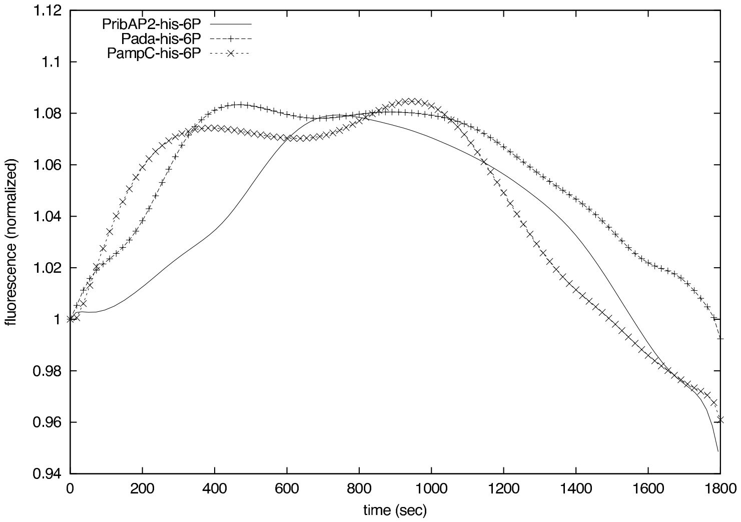


Figure 5.3: *On a polystyrene non-high-binding plate, the fluorescence kinetics of transcription reactions for several promoters attached to his-6P show relatively similar curves.*

The downward trends again appear not to be due to bleaching. In several experiments, decreasing the number of measurements did not significantly change the time to reach the peak, as would be expected if the peak was due to a bleaching effect.

In Figure 5.5, transcription reactions were run lacking either NTPs or the polymerase. In both cases, the fluorescence increases and decreases over time. Also, in Figure 5.6, the his-6P oligo by itself without polymerase shows a similar behavior as on the high-binding plates, with an increase but no decrease in fluorescence.

However, unlike the results on the high-binding plates, controls with only the polymerase or transcription reactions with promoters attached to a single pyrrolo-dC using BB-R-P, did not show the same bell curve but was flat (data not shown). The results indicates that the *his* pause site or the extra pyrrolo-dC signals on the his-6P oligo may have been producing a measurable fluorescence effect, although probably not the expected effect due to transcription.

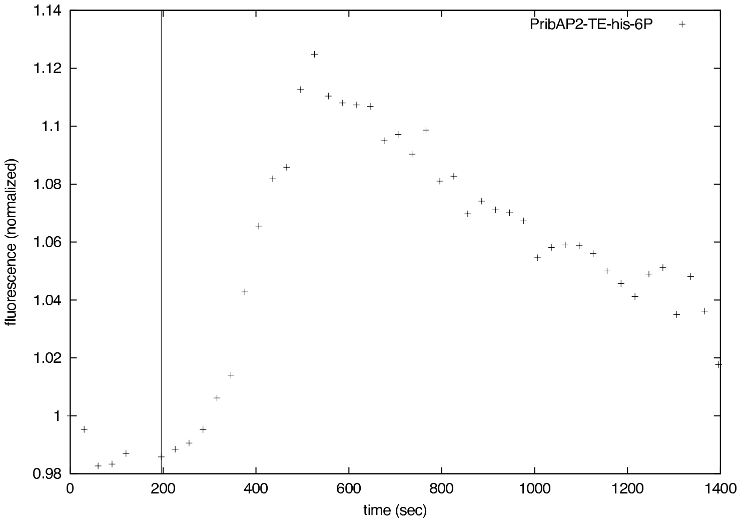


Figure 5.4: *The kinetics of a transcription reaction with a promoter attached to a terminator and his-6P shows an upward trend immediately after addition of RNAP. The vertical line indicates the time of addition of RNAP.*

Costar Polypropylene Plates

To test potential problems with protein binding to wells, polypropylene plates were tried, as the amount of binding on polypropylene was expected to be less than untreated polystyrene. As seen in Figure 5.7, a control well on a Costar clear bottom plate containing RNA polymerase and no DNA shows relatively flat kinetics for a long time before a decrease in fluorescence, presumably due to bleaching.

Transcription reactions with Plac-his-6P and one that includes the TE terminator, Plac-TE-his-6P, were done simultaneously with the previous control. As can be seen in Figure 5.8, the bell-shaped curve was present again but could no longer be attributed to protein binding. Similar results were obtained with another promoter, PribAP2-his-6P and PribAP2-TE-his-6P, as seen in Figure 5.9.

Different curves were seen depending on the promoter and whether there was an attached transcriptional terminator. The increase in fluorescence was faster for the PribAP2 curves compared with the Plac curves. In addition, the terminator for both promoters shifted the peak to the left.

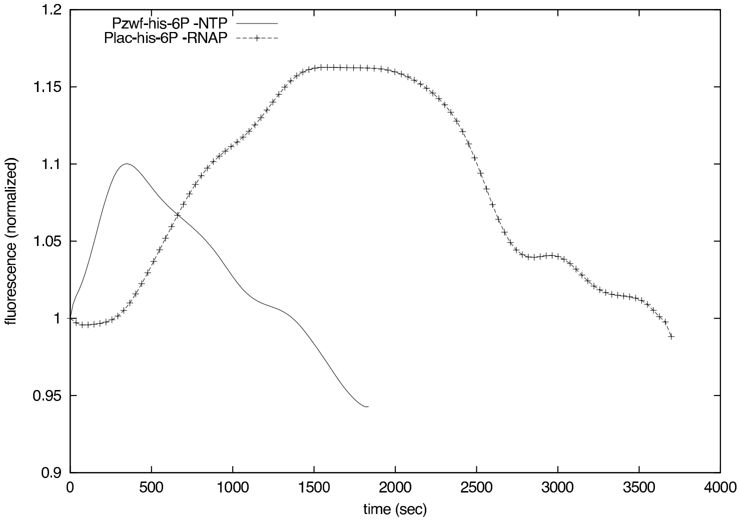


Figure 5.5: *Fluorescence was measured during transcription reactions for promoters attached to his-6P. The reactions were done either without NTPs (Pzwf) or without RNA polymerase (Plac).*

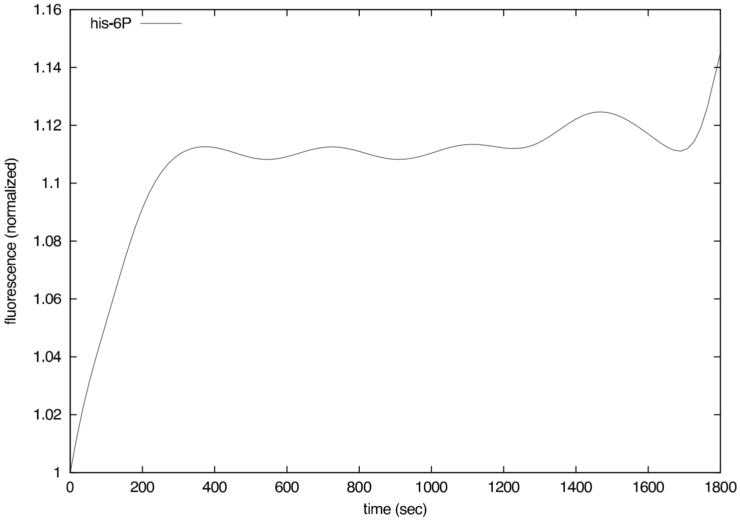


Figure 5.6: *The kinetics of the his-6P oligo in buffer with NTPs and without RNA polymerase shows an increase but no decrease in fluorescence.*

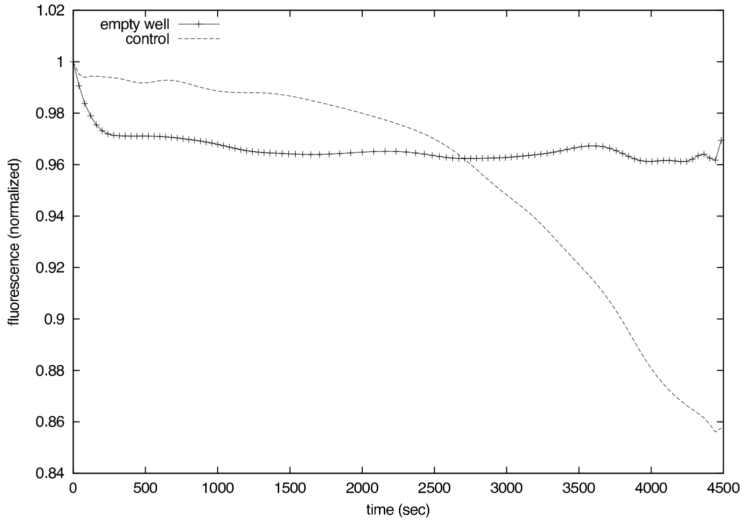


Figure 5.7: *A fluorescence control without DNA on a polypropylene plate shows relatively flat kinetics followed by a decrease. The control well contained buffer, NTPs, and RNA polymerase.*

Corning NBS Plates

The Corning NBS (non-binding surface) treated polystyrene, clear bottom, plates were expected to bind less than either untreated polypropylene or polystyrene plates.

Equal molar amounts of his-top-6P and his-bot (about 100 pmoles each) were added to 80μl of buffer and water and measured in NBS half-well plates. In Figure 5.10, the fluorescence kinetics clearly show a decrease in fluorescence over time, indicating a quenching of fluorescence as the strands anneal.

Unlike the previous plates, transcription reactions and BSA controls on these NBS plates do not show a bell-shaped fluorescence curve and are essentially flat. One example is seen in Figure 5.11 with one promoter. As all measured fluorescence curves were flat, not many useful results could be determined using these plates.

Greiner Fluotrac 200 Plates

The last set of plates tried were Greiner Fluotrac 200, polystyrene, medium binding plates. At the same sensitivity setting, the background

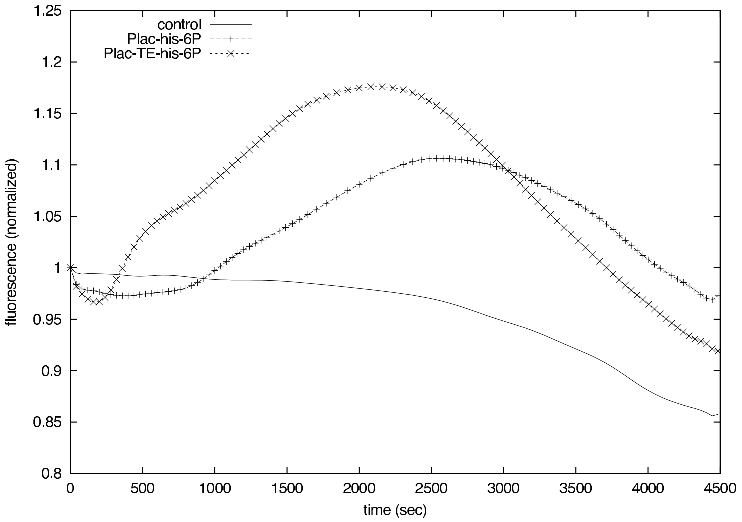


Figure 5.8: *The kinetics of transcription reactions with the Plac promoter with and without a terminator show bell-shaped curves not present in the control. The control well is the same as in Figure 5.7.*

fluorescence of an empty well from a Greiner plate is about half that of a well from a Corning NBS plate.

In Figure 5.12, the fluorescence for BSA measured on these plates showed the now familiar bell curve. After the BSA was measured, the well was emptied and rinsed with water. Then fresh BSA was put into the well and the fluorescence measured again. The second measurements were extremely flat, not showing the same curve.

Thus, the bell-shaped curves seen on the various plates are probably due to binding of protein to the well. After the protein saturates the well, it blocks any further protein from binding, even after a rinse with water.

For all further experiments, these Greiner plates were used, and, before use, the wells were pre-soaked with BSA and allowed to bind to protect the well from binding during the experiment.

The same annealing experiments were done as in Figure 5.10 on the Greiner plates. As seen in Figure 5.13, the fluorescence curve for the annealing condition shows a decrease over time, whereas the control reaction containing only the top oligo does not show an appreciable decrease in fluorescence.

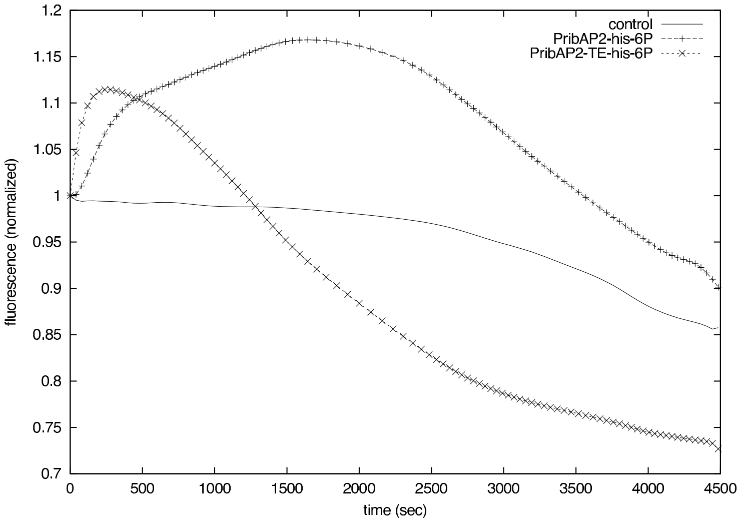


Figure 5.9: *Transcription reactions with the PribAP2 promoter with and without a terminator show similar results to the reactions with Plac above. The control well is the same as in Figure 5.7.*

5.1.2 Machine

Other plate readers were tried to compare their sensitivity and to obtain a fluorescence spectrum, something that was not possible on the BioTek FL600 machine. The most useful results came from a Tecan Safire plate reader.

Plate Background

Using the Safire, the background fluorescence from empty wells of all the previous plates was measured. Excitation was fixed at 360nm/12nm and the emission scanned between 400nm and 500nm. All the plates had extraordinarily high fluorescence compared with the Greiner plates. Figure 5.14 shows the data for a couple of the plates. In addition to a high background, there is an unexplained sharp drop off at 450nm for most of these plates.

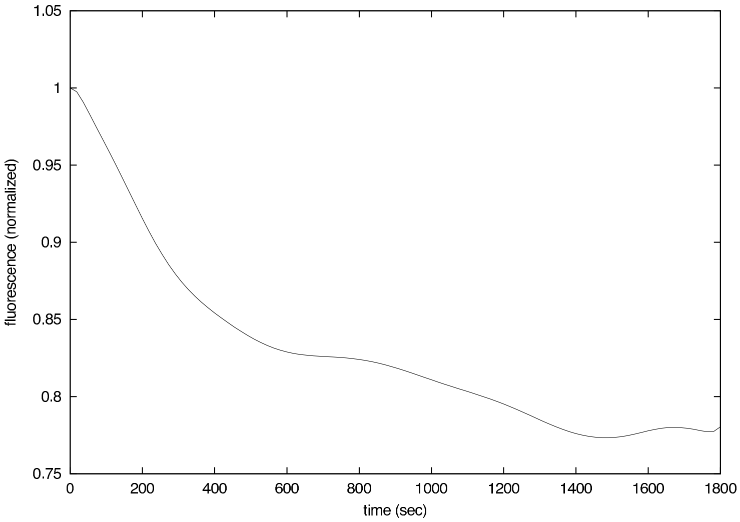


Figure 5.10: *Fluorescence of his-top-6P annealing with his-bot shows a gradual decrease, due to quenching of the pyrrolo-dC fluorescence in double-stranded DNA.*

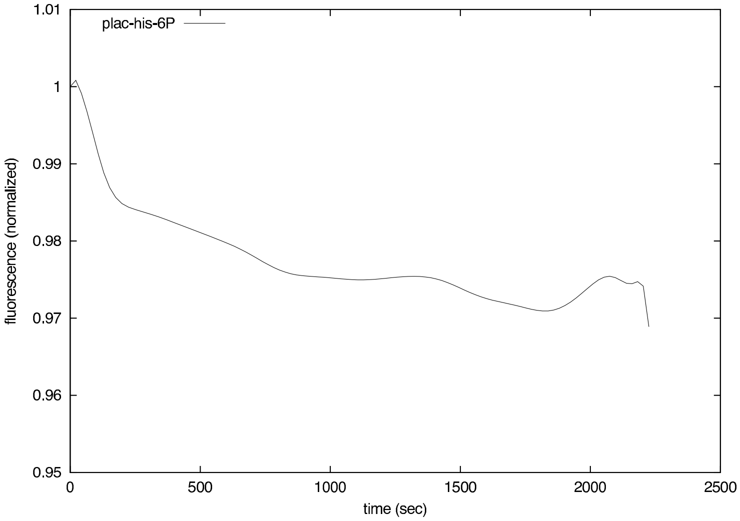


Figure 5.11: *A transcription reaction with plac-his-6P on Corning NBS plates does not show much change in fluorescence over time.*

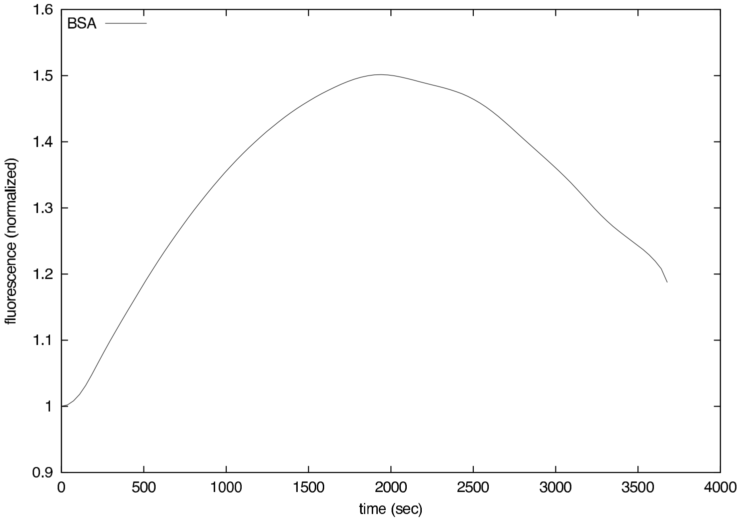


Figure 5.12: *BSA measured on a Greiner Fluotrac 200 plate shows a bell shaped curve similar to other plates.*

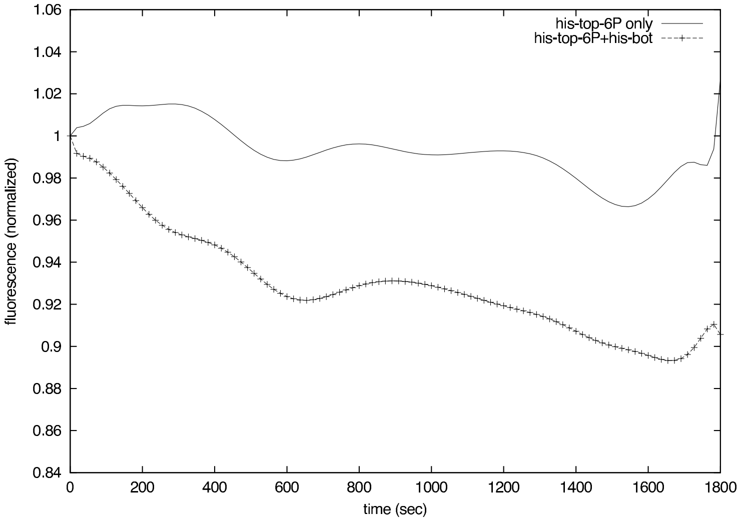


Figure 5.13: *The kinetics of his-top-6P annealing with his-bot on Greiner plates shows a decrease in fluorescence as would be expected. No change in fluorescence is seen with only his-top-6P.*

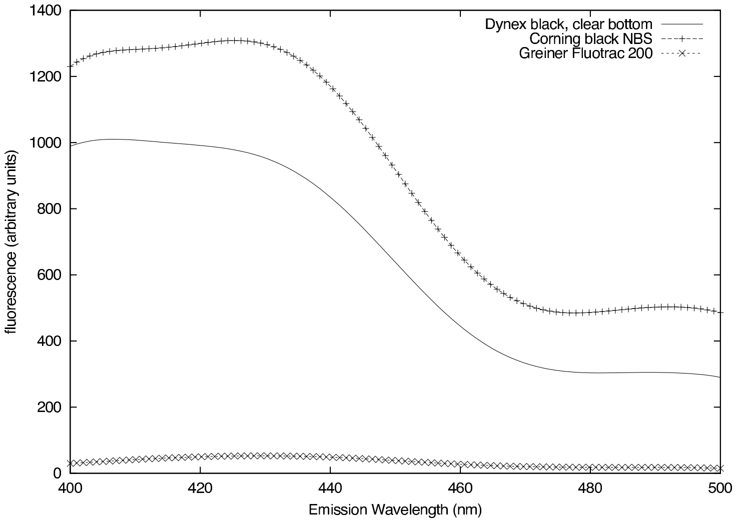


Figure 5.14: *The emission spectrum measured from empty wells of different types of plates show that the Greiner plates have a much lower background fluorescence. Excitation was set at 360nm/12nm.*

Pyrrolo-dC Fluorescence

Using a Greiner plate, the emission spectrum of a sample containing 4μM of a two nucleotide test oligo, `PG`, was measured on the Safire reader. As seen in Figure 5.15, the fluorescence of the pyrrolo-dC is easily detectable above the background.

The exact same plate and wells used in Figure 5.15 was also measured on the Bio-Tek machine. With excitation at 360nm/40nm, the emission was measured using both 460nm/30nm and 485nm/40nm filters. The fluorescence at 460nm was higher than at 485nm, unlike the spectrum measured on the Safire machine. Furthermore, the signal-to-noise ratio was about 1.5 on the Bio-Tek versus 7 on the Safire.

In Figure 5.16, the emission and excitation spectrum of pyrrolo-dC was measured using the two nucleotide test oligo, `PG`. The excitation spectrum used an emission filter at 440nm and was similar to published data. The emission spectrum was measured using an excitation at 350nm/7.5nm, similar to that used in a previously published report [91]. Surprisingly, the emission spectrum shows a peak around 440nm and a *minima* at 460nm. The shape of the spectrum curve was quite different

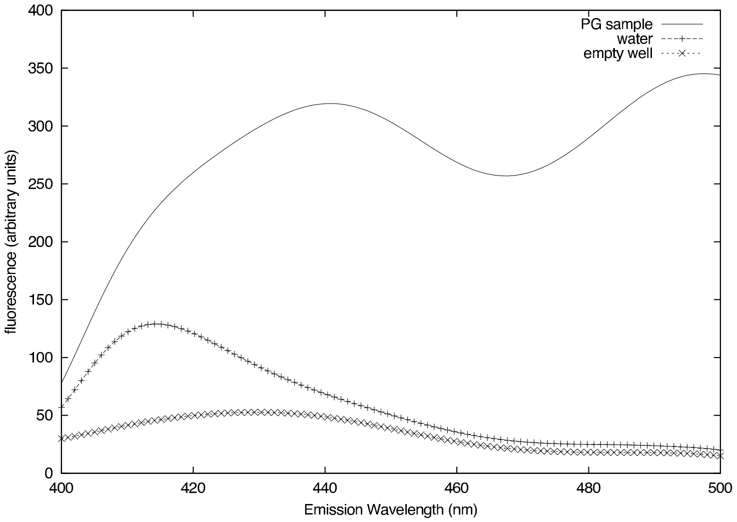


Figure 5.15: *The emission spectrum was measured on a Safire plate reader of a sample containing a test* `PG` *oligo in water. Excitation was set at 360nm/12nm.*

from the published emission spectrum and the data used as the original basis for choosing a 460nm emission filter.

The relative insensitivity of the Bio-Tek fluorescence readings may partially explain some lack of results, especially on the original plates with high background. In addition, the emission filters being used may have been suboptimal. The highest sensitivity with the Bio-Tek machine using available filters was found with the excitation at 360nm and emission measured at 516nm/20nm.

5.2 Transcription Experiments

All following measurements were made on the Bio-Tek machine using Greiner plates with excitation at 360nm and emission at 516nm.

5.2.1 PrrnB

To test that a change in fluorescence is measurable due to strand separation by RNA polymerase, the *rrnB* promoter was synthesized as

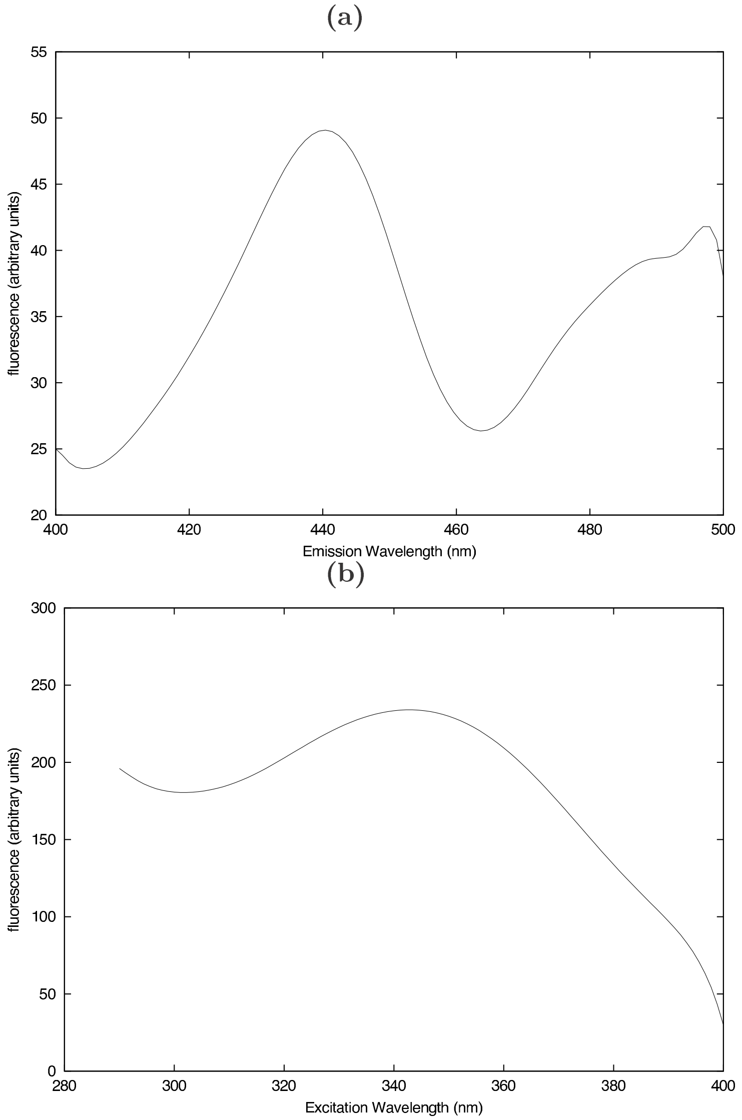


Figure 5.16: *(a) The emission spectrum of pyrrolo-dC was measured with the oligo* `PG` *with excitation at 350nm/7.5nm. (b) The excitation spectrum of pyrrolo-dC was measured with the oligo* `PG` *with emission measured at 440nm/7.5nm.*

	-RNAP	+RNAP	ratio
Plac	1062	833	0.78
Plac-P	812	778	0.96
Plac-5P-TE	766	793	1.03
PribAP2	891	601	0.67
PribAP2-P	903	842	0.93
PribAP2-5P-TE	742	815	1.10

Figure 5.17: *Fluorescence of several constructs with the promoters Plac and PribAP2 before and after addition of RNA polymerase shows a possible effect due to transcribing polymerase. The -TE samples contained the TE transcriptional terminator from T7.*

described in §4.3.6 with 5 pyrrolo-dC bases in the initial melting region.

The fluorescence of only PrrnB was distinctly higher than that of an empty well. After addition of 1.5μl RNA polymerase without any NTPs, the fluorescence noticeably increased (136%) but also decreased after some time in a manner that could not be attributed to bleaching. Under the same conditions, a control containing the RNA polymerase in water showed no change in fluorescence over time.

5.2.2 BB-R-P

Figure 5.17 shows an example of measuring fluorescence during a transcription experiment. Two promoters, Plac and PribAP2 had pyrrolo-dC incorporated on to the template strand via PCR with BB-R-P or BB-R-5P-TE. Controls were also done with BB-R to leave the promoter unlabeled. All PCRs were verified on an agarose gel.

For all promoter constructs, fluorescence was measured before and after addition of RNA polymerase. The ratio of the fluorescence with RNA polymerase to without is calculated. For the controls, the fluorescence with polymerase goes down, giving ratios less than 1. However, the ratios are significantly higher for the promoters with either a single or 5 pyrrolo-dC attached to the end, indicating a potential increase in fluorescence due to the RNA polymerase melting the region around the pyrrolo-dC.

In an independently run experiment, the same methodology used in Figure 5.17 was repeated, with similar final fluorescence ratios shown in Figure 5.18. In this experiment, several additional data points were collected. Fluorescence measured immediately after the addition of

	(a) Empty	(b) -RNAP	(c) +RNAP (-mix)	(d) +RNAP (+mix)	ratio (d/b)
Plac	809	821	812	671	0.82
Plac-P	845	757	742	720	0.95
Plac-5P-TE	821	775	739	772	1.00
PribAP2	845	949	943	699	0.74
PribAP2-P	842	821	748	748	0.91
PribAP2-5P-TE	851	742	739	726	0.98
Empty	861	912	964	949	1.04

Figure 5.18: *A separate experiment gives results similar to Figure 5.17. Column (a) gives the fluorescence of the empty well without the sample. Column (b) shows the fluorescence after adding the sample. Column (c) and (d) show the fluorescence after adding polymerase, with (c) measured immediately after adding polymerase and (d) measured after thorough mixing of the wells. The final column gives the ratio of column (d) to (b). The last row of the table shows a well that was left empty for all measurements.*

RNA polymerase, without mixing, showed little change in fluorescence across the conditions. Fluorescence was measured again after mixing the wells, and the fluorescence for the control promoters went down while all other samples remained relatively constant.

It is unclear why fluorescence decreases for the promoter samples without pyrrolo-dC and does not decrease for the samples with pyrrolo-dC. In addition, the lack of a significant *increase* in fluorescence for the pyrrolo-dC samples is puzzling.

Kinetic measurements were made during separate transcription reactions. Typical results showed up and down swings of fluorescence, possibly due to noise in the machine, and were seen with or without temperature control. The data in Figure 5.19 was measured with the machine temperature set to 31°C. In addition, the fluorescence fluctuations exist even when measuring empty wells.

Although the results in Figures 5.17 and 5.18 could be due to experimental and measurement noise, the similar results in independent experiments indicate some effect probably exists across sample conditions. It is unknown whether the results reflect the expected effect due to transcription or to another mechanism.

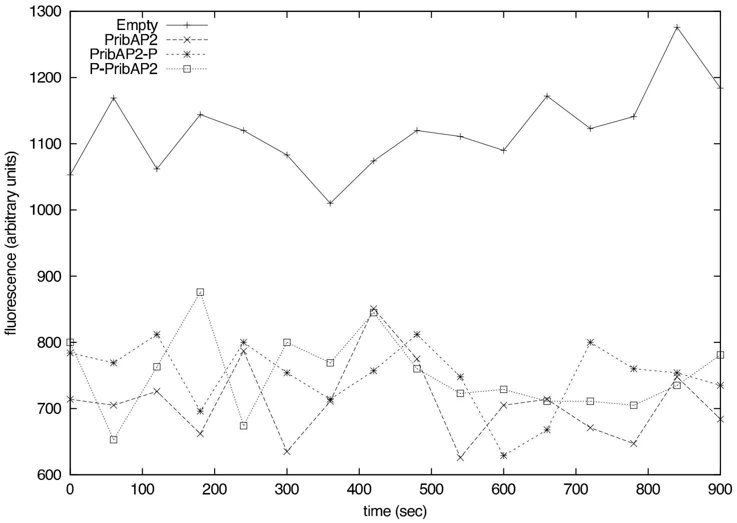


Figure 5.19: *Fluorescence variations were seen during transcription experiments. An empty well was measured along with several promoter constructs.*

Chapter 6

Discussion

We have presented the motivation for using polymerase arrival rates as the standard unit to describe modules. Although PAR-modules provide a useful abstraction layer for engineering biological systems, a method for measuring PAR is also critical.

PAR depends on many things including the strength of RNA polymerase binding, percent of abortive transcripts, promoter clearance rate, elongation rate, dissociation rate, and re-transcription rates. The purpose of many biochemical studies is to separate out many of these effects to obtain individual kinetic rates, but this is not needed nor desired for the measurement of polymerase arrival rates.

The goal of the PARMESAN method is to characterize PAR using a fluorescent assay. Below, we discuss potential problems with PARMESAN, future work, and provide a comparison with other methods.

6.1 Potential Problems

6.1.1 Template DNA

There are several potential issues related to the DNA used as the template for transcription. The linear PCR templates used may have problems with polymerases binding to the ends. The σ subunit of RNA polymerase greatly reduces the amount of non-promoter interactions but does not eliminate non-specificity completely. In particular, the affinity of RNA polymerase holoenzyme for the ends of promoter-free linear DNA fragments is as much as 600 times stronger than random interior sites [104, 152]. Circularizing DNA reduces the binding affinity of holoenzyme, whereas the binding affinity of core enzyme is not affected.

When possible, extra GC bases were added to the ends of the DNA template to make them harder to pull apart. Another way to avoid this potential problem would be to make the DNA circular, but then a strong terminator would need to be inserted to ensure the polymerase does not elongate around the circle multiple times.

There may also be issues involving runoff transcription. At the end of a DNA template, the polymerase presumably just falls off, but this is a boundary condition that may be handled differently than normal transcription. Forcing the polymerase to terminate transcription normally may be a more accurate indicator of its behavior in a real system.

6.1.2 Fluorescence Measurements

The major experimental problems have related to the fluorescence measurement itself, either with the plates or with the plate reader. Although the issues with the plates may be resolved, unresolved issues with the plate reader may still exist. It is not clear how important the sensitivity of the plate reader is to obtain useful results. Boosting the signal may be necessary by increasing the number of fluorescent bases used.

When one uses different fluorescence dyes, their intensities need to be normalized [127]. But even when using a single dye, the fluorescence can change over time due to storage conditions. Controlling for this photodegradation may be necessary. Other fluorescence related problems could include photobleaching or modification of the enzyme or DNA due to prolonged excitation energy, although this has not yet appeared to be a problem.

6.1.3 Reaction Conditions

The factors that affect promoter strength can be grouped into intrinsic and extrinsic factors [20]. Intrinsic factors include the DNA template, RNA polymerase, and transcription factors such as repressors or inducers. Extrinsic factors include reaction conditions such as the buffer or the temperature. It is desirable to measure the effect due to intrinsic factors and control for extrinsic factors.

Previous *in vitro* studies of transcription with both T7 and *E. coli* RNA polymerase have shown a strong and often complex dependence on extrinsic reaction conditions. For example, the initial lag phase due to open complex formation has been shown to be proportional to $[\mathrm{Na}^+]^{12}[\mathrm{RNAP}]^{-1}$ [123]. Other variables such as source of RNAP, temperature, and position of fluorescent bases can also affect this lag [33]. Even the measured *relative* strength of promoters has been shown to vary depending on temperature, salt concentration, and enzyme to DNA ratio [11, 63, 64, 81, 106].

Temperature

The activity of RNA polymerase is affected by temperature, but the effect is not large around the temperature used for the experiments [96].

Salt Concentration

With increasing NaCl or KCl concentrations, the binding of RNA polymerase to DNA drops [95, 98, 123, 141]. However, k_{cat}, a measure of transcription rate, does not show the same dependence on salt concentration. Lowering the ionic strength can have problems of causing the polymerase to aggregate and can greatly increase polymerase binding to nonspecific DNA sites [152].

The salt concentration also affects pausing at the *his* site [23], and higher KCl concentrations and lower magnesium concentrations can lead to increased termination efficiency [160]. Therefore, varying the salt concentration could have a variety of separate effects on transcription.

Salt concentration may also have an effect on the fluorescence readings. Salts have been shown to have effects on 2-aminopurine fluorescence [162] and may also affect pyrrolo-dC.

Enzyme Concentration

The concentration of RNA polymerase almost certainly has an effect on the observed polymerase arrival rates. Although there have been results indicating the rate of polymerase binding to DNA is independent of enzyme concentration [63, 124], much more than binding and open complex formation needs to be measured to determine PAR. In addition, these experiments usually limited the enzyme to one round of transcription.

In other experiments, the elongation rate was dependent on RNA polymerase concentration, due to cooperation among enzyme molecules [41]. As the transcriptional activity being measured is due to many complex factors, including elongation rate, the polymerase concentration is probably a significant factor.

NTP Concentration

Like the concentration of enzyme, the concentration of NTPs as a necessary substrate for transcription, has an effect on the transcription rate. The concentration of the initiating NTP has been shown to affect transcription efficiency for some promoters [47, 50], the concentration of the next nucleotide to be added at a pause site can affect the pause strength [82, 149].

At the *pyrBI* operon, the first six nucleotides are AAUUUG and at high UTP concentrations, the polymerase repeatedly transcribes bases 3 to 5 to synthesize RNA containing 30 or more U residues, effectively stalling transcription [30, 61]. In this case, the observed arrival of polymerase downstream depends greatly on the UTP concentrations.

Other Conditions

Another potential variable is the buffer pH. Increasing the pH decreases the association constant of the RNA polymerase for DNA but increases the rate of formation of the open complex from the closed complex [14].

6.2 Future Work

In theory, the PARMESAN method is a relatively easy and inexpensive way to assay transcription rates. In practice, after many failed experiments, only some decent results were obtained, leaving many open questions. Several possibilities for additional experiments follow.

6.2.1 Method Soundness

The first set of experiments should be to confirm the soundness of using pyrrolo-dC fluorescence as a measure of transcription rates.

Controls with pyrrolo-dC located in other places, for example, before a promoter, can be used to test that fluorescence changes are due to the location of the pyrrolo-dC. In addition, fluorescence changes across promoters should be quantified and compared with known promoter strengths to determine if there is a relationship between fluorescence changes and promoter strength.

Another way to confirm that fluorescence changes are due to transcription is to run the reactions at low temperatures where promoter binding but no strand separation is expected to occur. In experiments with 2-aminopurine at 4℃, no fluorescence change was detected due to the lack of strand separation [142].

6.2.2 Reaction Conditions

As described above, a measurement of transcriptional activity is only meaningful at a given concentration of RNA polymerase, NTP, and buffer salts. Different conditions may lead to different results. Before settling on the transcription buffer in §4.4.4, experiments were initially tried in several other buffers, including an Epicentre buffer designed for T7 RNA polymerase and an *E. coli* buffer, but these gave some misleading fluorescence measurements.

For measurements to be useful, the results should be applicable *in vivo* or in whatever system is being used. To increase the applicability of the *in vitro* PARMESAN method, the buffer and reaction conditions need to be optimized to provide the most useful results. As the goal of PARMESAN is to have standard conditions for measuring transcription, it may be ideal to have conditions as similar to intracellular conditions as possible. The cytoplasm of *E. coli* has been characterized to some extent [19]. For example, K^+ is the primary intracellular solute with a concentration varying from about 140 mM to 760 mM. Even from this type of data, finding the solution conditions to achieve the best results may be a difficult task.

In addition, in a real system, there will be other template DNA floating around, competing for the RNA polymerase. To make more realistic measurements, extra DNA such as pieces of the *E. coli* genome could be added to compete for the RNA polymerase.

6.2.3 Strand Dependence

Although different results are expected depending on whether the pyrrolo-dC is on the template or non-template strand, the current experiments have not been able to provide decisive evidence to show that one strand should be preferred over the other. The non-template strand is a safer choice for pyrrolo-dC incorporation, as the fluorescence of pyrrolo-dC on the non-template strand is not quenched during transcription. However, it is easier to incorporate pyrrolo-dC on the template strand via a PCR reaction, and if the template strand is a workable alternative, then it may be the preferred strategy.

6.2.4 Transcription Regulators

This method is not only useful for measuring promoter strength, but also to characterize various transcriptional signals. For example, the difference in PAR of a promoter followed by a terminator and the promoter alone is a measure of the terminator efficiency. Other transcriptional signals that could be characterized include pause sites or perhaps the speed of transcription through various sequences of DNA.

It has been assumed that downstream DNA has little effect on transcription rates. That is, the polymerase does not look-ahead during elongation. This is likely not a valid assumption, but the conditions when DNA downstream does have an effect can be easily tested using the PARMESAN method. The pyrrolo-dC can be incorporated in the middle of a strand of DNA and, by varying the downstream DNA, the dependence of PAR on the downstream region can be tested.

In addition to the signals located on the DNA template, many proteins and transcription factors influence transcription. The RNA polymerase must integrate all this information in deciding where and when to transcribe [109]. Transcription experiments have been done *in vitro* before to test the effects of repressors and activators [11, 131].

To measure the impact of these transcription factors on polymerase arrival rates, these factors can be added to the *in vitro* transcription reactions. For example, by adding a varying amount of repressor protein and measuring PAR, the effect of the repressor on transcription can be determined.

For more complex modules, a cell-free translation system can be added in addition to the transcription machinery [147]. An *in vitro* transcription and translation system should still be a more controlled environment, leading to more reproducible results, than in a free-living cell.

Including the effects of transcriptional regulators would allow us to obtain transfer curves for many simple types of modules as described in Chapter 3. An output can be measured while varying the inputs, obtaining the desired transfer curve.

6.2.5 Real-Time Kinetics

One benefit of using a non-destructive fluorescent assay with pyrrolo-dC is the potential to obtain real-time kinetic information. Kinetic measurements should be possible, having been done before in other applications with 2-aminopurine [9, 33, 58, 59].

It may be necessary to slow reactions to resolve the signals. Some ways to slow down the polymerase include lowering the temperature, lowering the NTP concentration, and using various drugs and proteins. Streptolydigin is a drug that slows the polymerase elongation rate and NusA enhances polymerase pausing about threefold at the *his* pause site [22]. However, these can affect the measurements in unknown ways, making it unclear whether these manipulations should be done at all.

6.2.6 Pyrrolo-dC Characterization

Not much is known about pyrrolo-dC and few experiments have been done with it. For use as a measurement tool, pyrrolo-dC needs to be characterized and understood better.

The sequence flanking 2-aminopurine has been shown to affect its fluorescence [9]. For example, guanine residues next to 2-aminopurine quenches its fluorescence. A similar effect may occur with pyrrolo-dC. To maximize the signal, it is necessary to determine how the pyrrolo-dC fluorescence depends on the nearby sequence.

Another issue may be unwanted disturbances in the system due to the pyrrolo-dC measurement. The amount of disturbance an artificial base like pyrrolo-dC has on the natural transcription process is not known. Although pyrrolo-dC can pair with guanine, it is possible the artificial base induces some change in the DNA structure that affects the transcription process.

6.2.7 Polymerase States

In a theoretical module, the polymerase arrival rates should be in identical units, allowing them to be interchangeably connected together. To do this, all polymerase arrivals were assumed to be measured at a point when the polymerase is in a fixed standard state.

In the PARMESAN method presented here, the state of the polymerases is not checked. The measurement method cannot distinguish, for example, between polymerase stalled at the pyrrolo-dC and polymerase actively transcribing. Although this method does not necessarily have to distinguish among polymerase states, a separate method may be required to guarantee that a module is valid and usable under the assumptions described in Chapter 3.

In the current experiments, the pyrrolo-dC fluorescence should not be measuring polymerases in the binding state and only measuring elongating polymerases. The pyrrolo-dC labels were placed at least 25 bp downstream of the initiation point, which was assumed to be far enough downstream to not lead to fluorescence changes upon RNA polymerase binding. Promoters with strong binding do not necessarily correlate with strong transcription initiation and elongation, and so it is important to determine that it is not binding being measured.

Another possible problem is the non-directionality of polymerase measurement using the pyrrolo-dC technique. The polymerase could potentially be arriving and moving in the opposite direction from expected. The measurement may actually be directional if a difference exists in the fluorescence depending on whether the pyrrolo-dC is on the template or non-template strands. But separating the effects from the two directions may be difficult.

6.3 Method Comparison

A summary of some advantages and disadvantages of the PARMESAN method are presented below, followed by comparisons with specific methods.

6.3.1 Advantages

- Measurements can be done *in vitro*. It is not needed to make plasmids or transform cells. Plasmid construction is not always an easy task, especially for promoters [159].
- Desired transcriptional activity is isolated from the noise of other events present in a cell.
- No radioactive labeling is needed, simplifying experiments and eliminating radioactive waste.
- Pyrrolo-dC is a non-destructive method for determining single stranded changes in DNA. Potassium permanganate ($KMnO_4$) or dimethyl sulfate (DMS) have been used to probe for single stranded DNA but cannot be used for real-time assays, as they can only permanently modify the DNA once [128, 132]. Similarly, cross-linking techniques covalently modify the DNA and can only be done once per experiment [114].
- The method is not limited to *E. coli* RNA polymerase and any purified polymerase could be substituted.
- The data collection is fast, with the fluorescence change being immediately detectable, providing the potential to look closely at the kinetics of polymerase arrivals.
- It is not required to synthesize new labeled oligos for every piece of DNA to be tested, as the oligos are designed to be general enough to work with existing and future modules.
- The BioBricks assembly scheme is used here, but PARMESAN is relatively independent of the module assembly strategy. Different pyrrolo-dC oligos could be designed for another assembly method.

- There is no need to worry about different RNA half-life times.
- There is minimal disturbance placed on the natural transcriptional system. By not using a modified enzyme or modified NTPs, most of the transcription process should proceed as normal. If the artificial base does have an effect, it likely only has an effect on the region around the site of incorporation. As we are measuring the rate the polymerase arrives at that point and do not care what happens after it passes, the disturbance on the system being measured is probably minimal.
- The method is not limited to pyrrolo-dC. Other fluorescent bases, such as 2-aminopurine, could be substituted for pyrrolo-dC.
- *Simplicity.* The method is theoretically simple and straightforward to do with materials common in molecular biology labs.

6.3.2 Disadvantages

- Results have not yet shown convincingly that the method as described here is a reasonable assay for polymerase arrival rates.
- Measurements need to be done *in vitro*. Using a fluorescent base is not directly adaptable to measurements in live cells, as the non-natural base would be quickly replaced by a non-fluorescent normal base. In addition, the precise incorporation of the fluorescence base into DNA could not be easily controlled in growing cells.
- The *in vitro* measurements may not accurately reflect *in vivo* behavior. The cellular “noise” may be important for the behavior of some systems.
- Pyrrolo-dC may disturb the transcription process.
- A source of purified RNA polymerase is required.
- It may be difficult to measure complex systems *in vitro*.

6.3.3 *in vivo* Methods

The standard method for measuring promoter strength *in vivo* is to clone a promoter in front of some reporter gene into a plasmid. However, strong promoters on plasmids can interfere with cell replication.

An inverse correlation between promoter strength and plasmid copy numbers has been found [1, 75].

Also, the plasmid copy number can vary significantly during cellular growth and can be estimated but not accurately determined. Thus, the measured promoter activity from *in vivo* assays may depend on the growth phase of the culture [150, 159].

Reporter Proteins

Using reporter protein levels has several potential problems. Reporter proteins can easily change the system they are measuring. They build up in the cell and do not measure transcription rates directly, making it difficult to compare results across experiments using different reporters. Also, there is stochastic noise when measuring protein expression levels [39]. Reactions done in defined reactions *in vitro* should be less noisy than in complex living cells.

Although testing promoter strength with reporter proteins is relatively easy, post-transcriptional events may distort the results, requiring precautions to protect the RNA or protein from being selectively degraded [159].

Also, as increased mRNA does not necessarily mean an increase in translation of that mRNA, using protein reporters for measuring transcriptional activity should be avoided if possible.

RNA Assays

It makes sense to measure the product of transcription, the RNA level, in assaying transcriptional activity. However, it is not always clear what is being measured when looking at RNA levels. Is it the complete, entire transcript or does it include aborted transcripts? Is RNA degradation taken into account?

Bases downstream of the start can affect both promoter strength and RNA degradation [66]. Using the same transcribed RNA for different promoters may limit this problem, but this is usually difficult as it requires knowing the exact position of the transcriptional start. This is also not applicable for measuring PAR in a general module that may contain more than a single promoter, as nothing can be assumed about the RNA produced or the degradation rates. In addition, the degradation rate is a function of the RNA concentration and the stage of bacterial growth [13], making RNA levels not an entirely accurate indicator of transcription rate or promoter strength.

Hybridization of mRNA with a probe as used in Northern blots, ribonuclease protection assays, and DNA microarrays are relatively in-sensitive and cannot easily be done in a continuous, real-time assay [127]. PCR-based methods can give greater sensitivity but are also relatively indirect.

Obtaining absolute numbers from fluorescence readings requires standards [127]. The concentration of ribosomal RNA relative to total RNA is relatively constant and useful as a standard. However, the method used in [170] to find the absolute transcription rate of the *rrn* promoters is tedious, and not easily applicable to measuring the PAR for a general module.

RNA assays and reporter proteins can often provide differing measurements of promoter strength. The strength of a T7 promoter and the *lac* promoter have been measured using both a tRNA reporter and a reporter protein [93]. Comparing the tRNA expression level, the T7 promoter was stronger, but using the protein expression level, the *lac* promoter was stronger. Other results have shown that transcriptional and reporter protein assays can be affected in opposite directions due to changes in growth rate [85, 86].

6.3.4 *in vitro* Methods

Differences in promoter strengths have been found between *in vitro* and *in vivo* experiments, undoubtedly due to many elements in living systems we do not yet understand. Even if experiments done *in vitro* do not accurately represent the system in a cell, they represent a system that can be understood more completely.

Inside a cell, many additional reactions affect transcription, many of which cannot be accounted for in a defined manner. For example, DNA polymerase during replication interacts with the transcribing RNA polymerase [90]. Undoubtedly, these are important to real systems, but for measurement and characterization purposes, these are side effects that should be eliminated.

It is also arguably easier to perform experiments *in vitro*. In fact, several attempts to clone promoters into cells failed, partially leading to the development of the PARMESAN method. With *in vitro* experiments, it is not necessary to worry about toxicity to cells due to strong promoters putting an unbearable burden on cells.

Some *in vitro* methods use rifampicin or heparin to limit transcription to a single round. As PAR measurements should include re-initiation rates, techniques that limit transcription cannot be used. Also, it has been determined that trailing RNA polymerase molecules can bump and push forward RNA polymerase molecules in front of it [41]. To ensure the measurement system is similar to the cellular environment, as few non-natural molecules as possible should be used.

Fluorescent Measurements

Most previous experiments using a fluorescent DNA base in a manner similar to the PARMESAN method have been done with 2-aminopurine (e.g. [8, 44, 117, 118, 142, 148, 163]). Usually, previous methods and experiments with either 2-aminopurine and pyrrolo-dC synthesized the fluorescent bases directly into the oligonucleotides to be tested. This is not acceptable as we require a general method for characterizing arbitrary modules.

Fluorescently labeled NTPs can be used to measure RNA formation or the total transcription rate [14, 33]. What is needed, however, is a system to test transcription rate at an arbitrary point in an arbitrary strand of DNA and to ignore all other transcriptional activity. Thus, polymerase arrival rates cannot be measured with techniques measuring total transcription rate.

Other fluorescent measurement methods are possible to measure PAR. In particular, FRET could be used as a potentially more sensitive method. One fluorophore can be attached to the DNA, as done here, and another fluorophone can be attached to either the polymerase or perhaps on to nucleotides. This would allow us to measure polymerase arrivals at specific locations but would require the addition of two fluorescent labels.

6.3.5 Computational Methods

Being able to accurately predict and model biological systems computationally would eliminate or speed up measurements. Algorithms and statistical methods have been used to predict promoter strengths [35, 157]. Current methods mostly involve taking into account the base composition of a piece of DNA. For example, as the GC bond requires a higher melting temperature and DNA needs to be melted during transcription, it may be expected that higher GC content leads to less transcription initiations.

Another method is to use sequence homology scores based on similarity to the consensus sequence. Homology scores have been effective at locating promoter sequences [112]. Although, the homology score for promoters shows high correlation with the association rate, it does not show much correlation with *in vivo* promoter strength [37, 69].

Experiments mutating existing promoters to be more like stronger promoters often have little or the opposite effect [137]. The phage P_L promoter is an extremely strong promoter in *E. coli* yet differs from the *E. coli* consensus sequence for promoters in several highly conserved locations [69].

Even though computational methods are desirable, current methods cannot predict accurately how a biological system will behave. Real measurements of the biological system will probably be necessary for some time and may provide useful data to improve computational approaches.

6.3.6 Single Molecule Techniques

The current PARMESAN method involves measuring the average fluorescence behavior from a large number of molecules and cannot obtain the absolute PAR, as a number of polymerase arrivals per second. To obtain an absolute number, it may be necessary to use single molecule techniques. Measuring fluorescence from single molecules is possible in theory. For example, FRET experiments have been done with single molecules [53]. There are other advanced methods to measure transcription activity from a single polymerase.

Single Molecule Transcription

Single molecules of RNA polymerase moving along DNA have been visualized by attaching a rhodamine fluorophore to the β' subunit of RNA polymerase [62]. The DNA is fixed to a surface in a known direction and the RNA polymerase can be visualized.

The tethered particle motion (TPM) method can be used to observe transcription directly under a light microscope [129]. A nanometer size gold particle is attached to DNA and a stalled RNA polymerase on the DNA is immobilized on a glass surface. The measured Brownian motion of the gold particle is used as an indication of the tether length and increases as the RNA polymerase transcribes the DNA. Both elongation rates and transcription termination can be observed from single RNA polymerase molecules using surface immobilized polymerase [165, 166].

Single molecule studies of *E. coli* RNA polymerase measuring transcription elongation rates have been done using optical-trap and microscopy techniques [2, 29, 46, 155, 167]. Atomic force microscopy (AFM) has been used to observe the activity of polymerase by taking sequential AFM images [67].

Although these methods can show the overall movement of single polymerases, measuring polymerase arrivals at a fixed point is difficult. Most of these single molecule techniques requires fixing either the polymerase or the DNA, and there may be undesirable effects from a non-mobile complex compared with a complex in solution. In addition, a major downside is that the needed equipment is not generally available to almost everyone.

Single Molecule PAR

Using the current pyrrolo-dC method, it may be possible to dilute reactions to a few or single molecules and measure the activity of a single polymerase. There could be detectable fluorescence spikes as a polymerase passes by pyrrolo-dC locations. Absolute polymerase arrival rates would then be measured by counting the spikes over time.

However, single molecule resolution may not be needed or desirable. More variation is found in single molecule measurements and a population average is more desirable for the purpose of module characterization. Studying single molecules may also be inaccurate, as there may be cooperativity among RNA polymerase molecules affecting transcriptional activity [41].

6.4 Conclusion

This thesis has proposed several key ideas:

- Modularity is the ultimate goal.
- All modules are defined by a fixed set of inputs and outputs measured in units of polymerase arrival rates (PAR).
- The fluorescent DNA nucleotide, pyrrolo-dC, shows a fluorescence change in single vs. double stranded DNA.
- PAR is measured by the fluorescence change of pyrrolo-dC as the RNA polymerase locally melts the DNA region during transcription.
- By standardizing the input and output units, modules can be connected and modeled easily.

With modules designed around polymerase arrival rates and the possibility of characterizing them using the PARMESAN method, engineering more complex synthetic biological systems may become a less daunting task.

REFERENCES

[1] Craig W. Adams and G. Wesley Hatfield. Effects of promoter strengths and growth conditions on copy number of transcription-fusion vectors. *Journal of Biological Chemistry*, 259(12):7399–7403, June 1984. 2.3.2, 6.3.3

[2] Karen Adelman, Arthur La Porta, Thomas J. Santangelo, John T. Lis, Jeffrey W. Roberts, and Michelle D. Wang. Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *Proc. Natl. Acad. Sci. USA*, 99(21):13538–13543, October 2002. 6.3.6

[3] Sankar Adhya, editor. *Methods in Enzymology*, volume 274. Academic Press, 1996. 25, 56

[4] Jawed Alam and Julia L. Cook. Reporter genes: Application to the study of mammalian gene transcription. *Analytical Biochemistry*, 188:245–254, 1990. 2.3.2, 2.3.2

[5] Karen M. Arndt and Michael J. Chamberlin. RNA chain elongation by *Escherichia coli* RNA polymerase factors affecting the stability of elongating ternary complexes. *Journal of Molecular Biology*, 213:79–108, 1990. 2.3.3

[6] Irina Artsimovitch and Robert Landick. Interaction of a nascent RNA structure with RNA polymerase is required for hairpin-dependent transcriptional pausing but not for transcript release. *Genes and Development*, 12(19):3110–3122, 1998. 2.2.5

[7] David T. Auble, Todd L. Allen, and Pieter L. deHaseth. Promoter recognition by *Escherichia coli* RNA polymerase: Effects of substitutions in the spacer DNA separating the −10 and −35 regions. *Journal of Biological Chemistry*, 261(24):11202–11206, August 1986. 2.1.2

[8] Rajiv P. Bandwar, Yiping Jia, Natalie M. Stano, and Smita S. Patel. Kinetic and thermodynamic basis of promoter strength: Multiple steps of transcription initiation by T7 RNA polymerase are modulated by the promoter sequence. *Biochemistry*, 41(11):3586–3595, March 2002. 6.3.4

[9] Rajiv P. Bandwar and Smita S. Patel. Peculiar 2-aminopurine fluorescence monitors the dynamics of open complex formation by bacteriophage T7 RNA polymerase. *Journal of Biological Chemistry*, 276(17):14075–14082, April 2001. 6.2.5, 6.2.6

[10] Gil Bar-Nahum and Evgeny Nudler. Isolation and characterization of σ^{70}-retaining transcription elongation complexes from *Escherichia coli*. *Cell*, 106(4):443–451, August 2001. 2.2.3

[11] Melanie M. Barker, Tamas Gaal, Cathleen A. Josaitis, and Richard L. Gourse. Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation *in vivo* and *in vitro*. *Journal of Molecular Biology*, 305(4):673–688, January 2001. 2.1.3, 2.2, 2.3.2, 6.1.3, 6.2.4

[12] Michael Becker-André and Klaus Hahlbrock. Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a *P*CR *a*ided *t*ranscript *t*itration assa*y* (PATTY). *Nucleic Acids Research*, 17(22):9437–9446, 1989. 2.3.2

[13] Joel G. Belasco, Gisela Nilsson, Alexander von Gabain, and Stanley N. Cohen. The stability of E. coli gene transcripts is dependent on determinants localized to specific mRNA segments. *Cell*, 46:245–251, July 1986. 2.3.2, 6.3.3

[14] Elisabeth Bertrand-Burggraf, Jean François Lefèvre, and M. Daune. A new experimental approach for studying the association between RNA polymerase and the *tet* promoter of pBR322. *Nucleic Acids Research*, 12(3):1697–1706, February 1984. 2.3.3, 6.1.3, 6.3.4

[15] BioBricks. `http://biobricks.ai.mit.edu/`. 4.1.1

[16] Michael Brunner and Hermann Bujard. Promoter recognition and promoter strength in the *Escherichia coli* system. *EMBO Journal*, 6(10):3139–3144, 1987. 2.2, 2.2.2

[17] Tom Burr, Jennie Mitchell, Annie Kolb, Steve Minchin, and Steve Busby. DNA sequence elements located immediately upstream of the -10 hexamer in *Escherichia coli* promoters: a systematic study. *Nucleic Acids Research*, 28(9):1864–1870, 2000. 2.1.2

[18] Sandhya Callac and Tomasz Heyduk. Conformation and DNA binding properties of a single-stranded DNA binding region of σ^{70} subunit from *Escherichia coli* RNA polymerase are modulated by an interaction with the core enzyme. *Biochemistry*, 37:3312–3320, 1998. 2.1.2, 2.3.3

[19] Scott Cayley, Barbara A. Lewis, Harry J. Guttman, and M. Thomas Record, Jr. Characterization of the cytoplasm of

Escherichia coli K-12 as a function of external osmolarity: Implications for protein-DNA interactions *in vivo*. *Journal of Molecular Biology*, 222:281–300, 1991. 6.2.2

[20] Michael J. Chamberlin and Lilian M. Hsu. RNA chain initiation and promoter escape by RNA polymerase. In Lin and Lynch [88], chapter 2, pages 7–25. 2.1.2, 2.2.3, 6.1.3

[21] Michael J. Chamberlin, William C. Nierman, Janey Wiggs, and Norma Neff. A quantitative assay for bacterial RNA polymerases. *Journal of Biological Chemistry*, 254(20):10061–10069, October 1979. 2.2.4, 2.3.3

[22] Cathleen L. Chan and Robert Landick. Dissection of the *his* leader pause site by base substitution reveals a multipartite signal that includes a pause RNA hairpin. *Journal of Molecular Biology*, 233:25–42, 1993. 2.2.5, 4.3.5, 4.2, 4.3.5, 6.2.5

[23] Cathleen L. Chan and Robert Landick. Effects of neutral salts on RNA chain elongation and pausing by *Escherichia coli* RNA polymerase. *Journal of Molecular Biology*, 268(1):37–53, 1997. 6.1.3

[24] Cathleen L. Chan, Daguang Wang, and Robert Landick. Multiple interactions stabilize a single paused transcription intermediate in which hairpin to 3' end spacing distinguishes pause and termination pathways. *Journal of Molecular Biology*, 268(1):53–68, 1997. 2.2.6

[25] Dipankar Chatterji and Vijaya Gopal. Fluorescence spectroscopy analysis of active and regulatory sites of RNA polymerase. In Adhya [3], pages 456–478. 2.3

[26] Austin Che and Bogdan Fedeleş. The modular modeling of biological networks using transcription rates. `http://austin.che.name/docs/modular_modeling.pdf`. 3.2.4, 3.4.4

[27] Christina M. Collis, Peter L. Molloy, Gerald W. Both, and Horace R. Drew. Influence of the sequence-dependent flexure of DNA on transcription in *E. coli*. *Nucleic Acids Research*, 17(22):9447–9468, 1989. 2.1.2

[28] Jean-Luc Darlix and Mireille Horaist. Existence and possible roles of transcriptional barriers in T7 DNA early region as shown by electron microscopy. *Nature*, 256:288–292, July 1975. 2.2.4

[29] R. John Davenport, Gijs J. L. Wuite, Robert Landick, and Carlos Bustamante. Single-molecule study of transcriptional pausing and arrest by *E. coli* RNA polymerase. *Science*, 287:2497–2500, 2000. 6.3.6

[30] Pieter L. deHaseth, Margaret L. Zupancic, and M. Thomas Record, Jr. RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *Journal of Bacteriology*, 180(12):3019–3025, June 1998. 2.1.1, 6.1.3

[31] Ulrich Deuschle, Wolfgang Kammerer, Reiner Gentz, and Hermann Bujard. Promoters of *Escherichia coli*: a hierarchy of *in vivo* strength indicates alternate structures. *EMBO Journal*, 5(11):2987–2994, 1986. 2.2.3, 2.3.2

[32] Birgit Dreier, Roger R. Beerli, David J. Segal, Jessica D. Flippin, and Carlos F. Barbas III. Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *Journal of Biological Chemistry*, 276(31):29466–29478, August 2001. 3.4.1

[33] Kevin S. Dunkak, Michael R. Otto, and Joseph M. Beechem. Real-time fluorescence assay system for gene transcription: Simultaneous observation of protein/DNA binding, localized DNA melting, and mRNA production. *Analytical Biochemistry*, 243(2):234–244, 1996. 2.3.3, 4.4.4, 6.1.3, 6.2.5, 6.3.4

[34] R.H. Ebright. RNA polymerase-DNA interaction: Structures of intermediate, open, and elongation complexes. In *Cold Spring Harbor symposia on quantitative biology*, volume 63, pages 11–10, 1998. 2.2

[35] *E. coli* promoter strength prediction. `http://wwwmgs.bionet.nsc.ru/mgs/programs/acts2/ec_pbla.html`. 6.3.5

[36] *E. coli* database collection – ECDC. `http://www.uni-giessen.de/~gx1052/ECDC/ecdc/ec_prom.htm`. 2.1.2

[37] Thomas Ellinger, Detlev Behnke, Hermann Bujard, and Jay D. Gralla. Stalling of *Escherichia coli* RNA polymerase in the +6 to +12 region *in vivo* is associated with tight binding to consensus promoter elements. *Journal of Molecular Biology*, 239(4):455–465, June 1994. 2.1.2, 6.3.5

[38] Michael B. Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, January 2000. 1.1

[39] Michael B. Elowitz, Arnold J. Levine, Eric D. Siggia, and Peter S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, August 2002. 6.3.3

[40] B.E. Enger-Valk, J. van Rotterdam, A. Kos, and P.H. Pouwels. Construction of new vectors for cloning of promoters. *Gene*, 15(4):297–305, December 1981. 2.3.2

[41] Vitaly Epshtein and Evgeny Nudler. Cooperation between RNA polymerase molecules in transcription elongation. *Science*, 300(5620):801–805, May 2003. 6.1.3, 6.3.4, 6.3.6

[42] Shawn T. Estrem, Tamas Gaal, Wilma Ross, and Richard L. Gourse. Identification of an UP element consensus sequence for bacterial promoters. *Proc. Natl. Acad. Sci. USA*, 95(17):9761–9766, August 1998. 2.1.2

[43] Hyone-Myong Eun. Marker/reporter enzymes. In *Enzymology Primer for Recombinant DNA Technology*, chapter 8. Academic Press, 1996. 2.3.2

[44] Andrew M. Fedoriw, Huaying Liu, Vernon E. Anderson, and Pieter L. deHaseth. Equilibrium and kinetic parameters of the sequence-specific interaction of *Escherichia coli* RNA polymerase with nontemplate strand oligodeoxyribonucleotides. *Biochemistry*, 37(34):11971–11979, August 1998. 2.1.2, 2.2, 2.3.1, 6.3.4

[45] Robert F. Fisher, Anathbandhu Das, Roberto Kolter, Malcolm E. Winkler, and Charles Yanofsky. Analysis of the requirements for transcription pausing in the tryptophan operon. *Journal of Molecular Biology*, 182:397–409, 1985. 2.2.5

[46] Nancy R. Forde, David Izhaky, Glenna R. Woodcock, Gijs J. L. Wuite, and Carlos Bustamante. Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by *Escherichia coli* RNA polymerase. *Proc. Natl. Acad. Sci. USA*, 99(18):11682–11687, September 2002. 6.3.6

[47] Tamas Gaal, Michael S. Bartlett, Wilma Ross, Charles L. Turnbough, Jr., and Richard L. Gourse. Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. *Science*, 278(5346):2092–2097, December 1997. 6.1.3

[48] Howard B. Gamper and John E. Hearst. A topological model for transcription based on unwinding angle analysis of E. coli RNA polymerase binary, initiation and ternary complexes. *Cell*, 29:81–90, May 1982. 2.2.4

[49] Timothy S. Gardner, Charles R. Cantor, and James J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403(6767):339–342, January 2000. 1.1

[50] R. L. Gourse, T. Gaal, S.E. Aiyar, M.M. Barker, S.T. Estrem, C.A. Hirvonen, and W. Ross. Strength and regulation without transcription factors: lessons from bacterial rRNA promoters. In *Cold Spring Harbor symposia on quantitative biology*, volume 63, pages 131–139, 1998. 6.1.3

[51] Elizabeth J. Grayhack, Xianjie Yang, Lester F. Lau, and Jeffrey W. Roberts. Phage lambda gene Q antiterminator recognizes RNA polymerase near the promoter and accelerates it through a pause site. *Cell*, 42(1):259–269, August 1985. 2.3.3

[52] Ivan Gusarov and Evgeny Nudler. The mechanism of intrinsic transcription termination. *Molecular Cell*, 3(4):495–504, 1999. 2.2.6

[53] T. Ha, Th. Enderle, D. F. Ogletree, D. S. Chemla, P. R. Selvin, and S. Weiss. Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. USA*, 93:6264–6268, June 1996. 6.3.6

[54] John D. Helmann and Pieter L. deHaseth. Protein–nucleic acid interactions during open complex formation investigated by systematic alteration of the protein and DNA binding partners. *Biochemistry*, 38(19):5959–5967, May 1999. 2.2.1, 2.2.2

[55] Ruti Hershberg, Gill Bejerano, Alberto Santos-Zavaleta, and Hanah Margalit. PromEC: An updated database of *Eschericia coli* mRNA promoters with experimentally identified transcriptional start sites. *Nucleic Acids Research*, 29(1):277, 2001. `http://bioinfo.md.huji.ac.il/marg/promec/`,. 2.1.2, 4.4.1

[56] Tomasz Heyduk, Yuexing Ma, Hong Tang, and Richard H. Ebright. Fluorescence anisotropy: Rapid, quantitative assay for protein-DNA and protein-protein interaction. In Adhya [3], pages 492–503. 2.3.1

[57] David R. Hodge, Delores M. Thompson, Alexandra Panayiotakis, and Arun Seth. Transcriptional activation analysis by the chloramphenicol acetyl transferase (CAT) enzyme assay. In Tymms [147], chapter 28. 2.3.2

[58] Yiping Jia, Amarendra Kumar, and Smita S. Patel. Equilibrium and stopped-flow kinetic studies of interaction between T7 RNA polymerase and its promoters measured by protein and 2-aminopurine fluorescence changes. *Journal of Biological Chemistry*, 271(48):30451–30458, November 1996. 2.3.1, 2.3.3, 6.2.5

[59] Yiping Jia and Smita S. Patel. Kinetic mechanism of GTP binding and RNA synthesis during transcription initiation by bacteriophage T7 RNA polymerase. *Journal of Biological Chemistry*, 272(48):30147–30153, November 1997. 6.2.5

[60] Yiping Jia and Smita S. Patel. Kinetic mechanism of transcription initiation by bacteriophage T7 RNA polymerase. *Biochemistry*, 36(14):4223–4232, 1997. 2.2

[61] Ding Jun Jin and Charles L. Turnbough, Jr. An *Escherichia coli* RNA polymerase defective in transcription due to its overproduction of abortive initiation products. *Journal of Molecular Biology*, 236:72–80, 1994. 6.1.3

[62] Hiroyuki Kabata, Osamu Kurosawa, Ichiro Arai, Masao Washizu, Stefanie A. Margarson, Robert E. Glass, and Nobuo Shimamoto. Visualization of single molecules of RNA polymerase sliding along DNA. *Science*, 262(5139):1561–1563, 1993. 6.3.6

[63] Thomas R. Kadesch, Steven Rosenberg, and Michael J. Chamberlin. Binding of *Escherichia coli* RNA polymerase holoenzyme to bacteriophage T7 DNA: Measurements of binding at bacteriophage T7 promoter A_1 using a template competition assay. *Journal of Molecular Biology*, 155:1–29, 1982. 2.2, 2.3.3, 6.1.3, 6.1.3

[64] Masayuki Kajitani and Akira Ishihama. Determination of the promoter strength in the mixed transcription system. II. Promoters of ribosomal RNA, ribosomal protein S1, and *recA* protein operons from *Escherichia coli*. *Nucleic Acids Research*, 11(12):3873–3888, 1983. 6.1.3

[65] Masayuki Kajitani and Akira Ishihama. Determination of the promoter strength in the mixed transcription system: Promoters of lactose, tryptophan and ribosomal protein L10 operons from *Escherichia coli*. *Nucleic Acids Research*, 11(3):671–686, 1983. 2.3.3

[66] Wolfgang Kammerer, Ulrich Deuschle, Reiner Gentz, and Hermann Bujard. Functional dissection of *Escherichia coli* promoters: information in the transcribed region is involved in late steps of the overall process. *EMBO Journal*, 5(11):2995–3000, November 1986. 2.1.2, 2.2, 6.3.3

[67] Sandor Kasas, Neil H. Thomson, Bettye L. Smith, Helen G. Hansma, Xingshu Zhu, Martin Guthold, Carlos Bustamante, Eric T. Kool, Mikhail Kashlev, and Paul K. Hansma. *Escherichia* coli RNA polymerase activity observed using atomic force microscopy. *Biochemistry*, 36(3):461–468, 1997. 6.3.6

[68] Robin-Ann Klotsky and Ira Schwartz. Measurement of cat expression from growth-rate-regulated promoters employing beta-lactamase activity as an indicator of plasmid copy number. *Gene*, 55(1):141–146, 1987. 2.3.2

[69] R. Knaus and H. Bujard. Principles governing the activity of *E. coli* promoters. In Fritz Eckstein and David M.J. Lilley, editors, *Nucleic Acids and Molecular Biology*, volume 4, pages 110–122. Springer-Verlag, 1990. 2.2, 6.3.5

[70] Natalia Komissarova and Mikhail Kashlev. RNA polymerase switches between inactivated and activated states by translocating back and forth along the DNA and the RNA. *Journal of Biological Chemistry*, 272(24):15329–15338, June 1997. 2.3.3

[71] Natalia Komissarova and Mikhail Kashlev. Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc. Natl. Acad. Sci. USA*, 94:1755–1760, March 1997. 2.2.5

[72] Natalia Komissarova and Mikhail Kashlev. Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc. Natl. Acad. Sci. USA*, 95:14699–14704, December 1998. 2.2.4, 3.3.2

[73] Nataliya Korzheva, Arkady Mustaev, Maxim Kozlov, Arun Malhotra, Vadim Nikiforov, Alex Goldfarb, and Seth A. Darst. A structural model of transcription elongation. *Science*, 289(5479):619–625, July 2000. 3.3.2

[74] Sidney R. Kushner. mRNA decay. In Neidhardt [113], pages 849–860. 2.3.2

[75] Angus I. Lamond and Andrew A. Travers. Requirement for an upstream element for optimal transcription of a bacterial tRNA gene. *Nature*, 305(5931):248–250, September 1983. 6.3.3

[76] Robert Landick. RNA polymerase clamps down. *Cell*, 105:567–570, June 2001. 2.2.4

[77] Robert Landick, Charles L. Turnbough, Jr., and Charles Yanofsky. Transcription attenuation. In Neidhardt [113], pages 1263–1286. 2.2.5

[78] Robert Landick and Charles Yanofsky. Isolation and structural analysis of the *Escherichia coli trp* leader paused transcription complex. *Journal of Molecular Biology*, 196:363–377, 1987. 2.2.6

[79] Donna N. Lee and Robert Landick. Structure of RNA and DNA chains in paused transcription complexes containing *Escherichia coli* RNA polymerase. *Journal of Molecular Biology*, 228(3):759–777, December 1992. 2.2.5, 4.3.5

[80] Donna N. Lee, Le Phung, Judith Stewart, and Robert Landick. Transcription pausing by *Escherichia coli* RNA polymerase is modulated by downstream DNA sequences. *Journal of Biological Chemistry*, 265(25):15145–15153, September 1990. 2.2.5, 2.2.6, 4.3.5

[81] S. Leirmo and M. Thomas Record, Jr. Structural, thermodynamic and kinetic studies of the interaction of Eσ^{70} RNA polymerase with promoter DNA. In Fritz Eckstein and David M.J. Lilley, editors, *Nucleic Acids and Molecular Biology*, volume 4, pages 123–151. Springer-Verlag, 1990. 6.1.3

[82] Judith R. Levin and Michael J. Chamberlin. Mapping and characterization of transcriptional pause sites in the early genetic region of bacteriophage T7. *Journal of Molecular Biology*, 196(1):61–84, July 1987. 6.1.3

[83] Jennifer C. Lewis, Agatha Feltus, C. Mark Ensor, Sridhar Ramanathan, and Sylvia Daunert. Applications of reporter genes. *Analytical Chemistry*, 70(17):579A–585A, September 1998. 2.3.2

[84] S. Liang, M. Bipatnath, Y. Xu, S. Chen, P. Dennis, M. Ehrenberg, and H. Bremer. Activities of constitutive promoters in *Escherichia coli*. *Journal of Molecular Biology*, 292(1):19–37, September 1999. 2.3.2

[85] S.-T. Liang, Y.-C. Xu, P. Dennis, and H. Bremer. mRNA composition and control of bacterial gene expression. *Journal of Bacteriology*, 182(11):3037–3044, June 2000. 6.3.3

[86] Sung-Tzu Liang, Patrick P. Dennis, and Hans Bermer. Expression of *lac*Z from the promoter of the *Eschericia coli spc* operon cloned into vectors carrying the W205 *trp-lac* fusion. *Journal of Bacteriology*, 180(23):6090–6100, December 1998. 2.3.2, 6.3.3

[87] David MJ Lilley and Timothy J Wilson. Fluorescence resonance energy transfer as a structural tool for nucleic acids. *Current Opinion in Chemical Biology*, 4(5):507–517, 2000. 2.3.3

[88] E. C. C. Lin and A. Simon Lynch, editors. *Regulation of Gene Expression in* Escherichia coli. R. G. Landes Company, 1996. 20, 122, 138

[89] Shlomit Lisser and Hanah Margalit. Compilation of *E. coli* mRNA promoter sequences. *Nucleic Acids Research*, 21(7):1507–1516, 1993. 2.1.2

[90] Bin Liu, Mei Lie Wong, Rachel L. Tinker, E. Peter Geiduschek, and Bruce M. Alberts. The DNA replication fork can pass RNA polymerase without displacing the nascent transcript. *Nature*, 366:33–39, November 1993. 6.3.4

[91] Cuihua Liu and Craig T. Martin. Fluorescence characterization of the transcription bubble in elongation complexes of T7 RNA polymerase. *Journal of Molecular Biology*, 308(3):465–475, May 2001. 3.3.1, 5.1, 5.1.2

[92] Cuihua Liu and Craig T. Martin. Promoter clearance by T7 RNA polymerase: Initial bubble collapse and transcript dissociation monitored by base analog fluorescence. *Journal of Biological Chemistry*, 277(4):2725–2731, January 2002. 3.3.1, 3.3.1, 3.3.2

[93] Pascal J. Lopez, Isabelle Iost, and Marc Dreyfus. The use of a tRNA as a transcriptional reporter: The T7 late promoter is extremely efficent in *Escherichia coli* but its transcripts are poorly expressed. *Nucleic Acids Research*, 22(7):1186–1193, 1994. 2.3.2, 6.3.3

[94] Rolf Lutz and Hermann Bujard. Independent and tight transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I_1-I_2 regulatory elements. *Nucleic Acids Research*, 25(6):1203–1210, 1997. 2.3.2

[95] Walter F. Mangel and Michael J. Chamberlin. Studies of ribonucleic acid chain initiation by *Escherichia coli* ribonucleic acid polymerase bound to T7 deoxyribonucleic acid. II. The effect of alterations in ionic strength on chain initiation and on the conformation of binary complexes. *Journal of Biological Chemistry*, 249(10):3002–3006, May 1974. 6.1.3

[96] Walter F. Mangel and Michael J. Chamberlin. Studies of ribonucleic acid chain initiation by *Escherichia coli* ribonucleic acid polymerase bound to T7 deoxyribonucleic acid. III. The effect of temperature on ribonucleic acid chain initiation and on the conformation of binary complexes. *Journal of Biological Chemistry*, 249(10):3007–3013, May 1974. 6.1.3

[97] Craig T. Martin and Joseph E. Coleman. Kinetic analysis of T7 RNA polymerase-promoter interactions with small synthetic promoters. *Biochemistry*, 26:2690–2696, 1987. 2.3.3

[98] Maribeth Maslak and Craig T. Martin. Effects of solution conditions on the steady-state kinetics of initiation of transcription by T7 RNA polymerase. *Biochemistry*, 33(22):6918–6924, June 1994. 6.1.3

[99] WT Mason, editor. *Fluorescent and Luminescent Probes for Biological Activity: A Practical Guide to Technology for Quantitative Real-Time Analysis*. Biological Techniques. Academic Press, 2nd edition, 1999. 2.3.1

[100] William R. McClure. Rate-limiting steps in RNA chain initiation. *Proc. Natl. Acad. Sci. USA*, 77(10):5634–5638, October 1980. 2.3.3

[101] William R. McClure. Mechanism and control of transcription initiation in prokaryotes. *Annual Review of Biochemistry*, 54:171–204, 1985. 2.1.1, 2.2, 2.2.3, 2.3.3

[102] William R. McClure and Carol L. Cech. On the mechanism of rifampicin inhibition of RNA synthesis. *Journal of Biological Chemistry*, 253(24):8949–8956, December 1978. 2.1.1

[103] William R. McClure, Carol L. Cech, and David E. Johnston. A steady state assay for the RNA polymerase initiation reaction. *Journal of Biological Chemistry*, 253(24):8941–8948, December 1978. 2.3.3

[104] Paul Melançon, Richard R. Burgess, and M. Thomas Record, Jr. Direct evidence for the preferential binding of *Escherichia coli* RNA polymerase holoenzyme to the ends of

deoxyribonucleic acid restriction fragments. *Biochemistry*, 22(22):5169–5176, 1983. 6.1.1

[105] Sandra Milan, Linda D'Ari, and Michael J. Chamberlin. Structural analysis of ternary complexes of *Escherichia coli* RNA polymerase: Ribonuclease footprinting of the nascent RNA in complexes. *Biochemistry*, 38(1):218–225, 1999. 2.2.4

[106] Jacqueline S. Miller and Richard R. Burgess. Selectivity of RNA chain initiation *in vitro*. 3. Variables affecting initiation of transcription. *Biochemistry*, 17(11):2064–2069, May 1978. 6.1.3

[107] Jeffrey H. Miller. Assays of the *lac* operon enzymes. In *Experiments in Molecular Genetics*. Cold Spring, 1972. 2.3.2, 2.3.2

[108] Akiko Miura, Judy Heilig Krueger, Seiga Itoh, Herman A. de Boer, and Masayasu Nomura. Growth-rate-dependent regulation of ribosome synthesis in E. coli: Expression of the *lacZ* and *galK* genes fused to ribosomal promoters. *Cell*, 25(3):773–82, September 1981. 2.3.2

[109] Rachel Anne Mooney, Irina Artsimovitch, and Robert Landick. Information processing by RNA polymerase: Recognition of regulatory signals during RNA chain elongation. *Journal of Bacteriology*, 180(13):3265–3275, 1998. 2.2.2, 2.2.4, 2.2.6, 4.3.6, 6.2.4

[110] Jayanta Mukhopadhyay, Achillefs N. Kapanidis, Vladimir Mekler, Ekaterine Kortkhonjia, Yon W. Ebright, and Richard H. Ebright. Translocation of σ^{70} with RNA polymerase during transcription: Fluorescence resonance energy transfer assay for movement relative to DNA. *Cell*, 106(4):453–463, August 2001. 2.2.3, 2.3.3

[111] Benno Müller-Hill. *The* lac *operon: a short history of a genetic paradigm*. Walter de Gruyter, 1996. 3.2.4

[112] Martin E. Mulligan, Diane K. Hawley, Robert Entriken, and William R. McClure. *Escherichia coli* promoter sequences predict *in vitro* RNA polymerase selectivity. *Nucleic Acids Research*, 12(1):789–800, 1984. 6.3.5

[113] Frederick C. Neidhardt, editor. Escherichia coli *and* Salmonella*, Cellular and Molecular Biology*. ASM Press, 1996. 74, 77, 119, 121

[114] Evgeny Nudler, Ivan Gusarov, Ekaterina Avetissova, Maxim Kozlov, and Alex Goldfarb. Spatial organization of transcription elongation complex in *Escherichia coli*. *Science*, 281:424–428, July 1998. 6.3.1

[115] Evgeny Nudler, Arkady Mustaev, Evgeny Lukhtanov, and Alex Goldfarb. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell*, 89:33–41, April 1997. 2.2.4

[116] John R. Palmer and Charles J. Daniels. A transcriptional reporter for in vivo promoter analysis in the archaeon *Haloferax volcanii*. *Applied and Environmental Microbiology*, 60(10):3867–3869, October 1994. 2.3.2

[117] Edward L. Rachofsky, Roman Osman, and J. B. Alexander Ross. Probing structure and dynamics of DNA with 2-aminopurine: Effects of local environment on fluorescence. *Biochemistry*, 40:946–956, 2001. 6.3.4

[118] Kevin D. Raney, Lawrence C. Sowers, David P. Millar, and Stephen J. Benkovic. A fluorescence-based assay for monitoring helicase activity. *Proc. Natl. Acad. Sci. USA*, 91(14):6644–6648, July 1994. 2.3.1, 6.3.4

[119] M. Thomas Record, Jr., William S. Reznikoff, Maria L. Craig, Kristi L. McQuade, and Paula J. Schlax. *Escherichia coli* RNA polymerase ($E\sigma^{70}$), promoters, and the kinetics of the steps of transcription initiation. In Neidhardt [113], pages 792–821. 2.1.1, 2.1.2, 2.1.3, 2.2, 2.2.1, 2.2.3, 2.3.3, 3.3.2

[120] William S. Reznikoff, Deborah A. Siegele, Deborah W. Cowing, and Carol A. Gross. The regulation of transcription initiation in bacteria. *Annual Review of Genetics*, 19:355–387, 1985. 2.1.1

[121] John P. Richardson and Jack Greenblatt. Control of RNA chain elongation and termination. In Neidhardt [113], pages 822–848. 2.1.1, 2.2.4, 2.2.5

[122] Jeffrey W. Roberts. Transcription termination and its control. In Lin and Lynch [88], chapter 3, pages 27–45. 2.2.6

[123] Jung-Hye Roe and M. Thomas Record, Jr. Regulation of the kinetics of the interaction of *Escherichia coli* RNA polymerase with the λP_R promoter by salt concentration. *Biochemistry*, 24(18):4721–4726, August 1985. 6.1.3, 6.1.3

[124] Steven Rosenberg, Thomas R. Kadesch, and Michael J. Chamberlin. Binding of *Escherichia coli* RNA polymerase holoenzyme to bacteriophage T7 DNA: Measurements of the rate of open complex formation at T7 promoter A_1. *Journal of Molecular Biology*, 155:31–51, 1982. 6.1.3

[125] Wilma Ross, Sarah E. Aiyar, Julia Salomon, and Richard L. Gourse. *Escherichia coli* promoters with UP elements of different strengths: Modular structure of bacterial promoters. *Journal of Bacteriology*, 180(20):5375–5383, October 1998. 2.1.2

[126] Wilma Ross, Khoosheh K. Gosink, Julia Salomon, Kazuhiko Igarashi, Chao Zou, Akira Ishihama, Konstantin Severinov, and Richard L. Gourse. A third recognition element in bacterial promoters: DNA binding by the α subunit of RNA polymerase. *Science*, 262(5135):1407–1413, November 1993. 2.1.2

[127] Charles M. Roth. Quantifying gene expression. *Current Issues Molecular Biology*, 4:93–100, 2002. 2.3.2, 6.1.2, 6.3.3

[128] Selina Sasse-Dwight and Jay D. Gralla. $KMnO_4$ as a probe for *lac* promoter DNA melting and mechanism *in vivo*. *Journal of Biological Chemistry*, 264(14):8074–8081, May 1989. 6.3.1

[129] Dorothy A. Schafer, Jeff Gelles, Michael P. Sheetz, and Robert Landick. Transcription by single molecules of RNA polymerase observed by light microscopy. *Nature*, 352:444–448, August 1991. 6.3.6

[130] Joseph G. Schlageck, Michael Baughman, and L. R. Yarbrough. Spectroscopic techniques for study of phosphodiester bond formation by *Escherichia coli* RNA polymerase. *Journal of Biological Chemistry*, 254(23):12074–12077, December 1979. 2.3.3

[131] Paula J. Schlax, Michael W. Capp, and M. Thomas Record, Jr. Inhibition of transcription initiation by *lac* repressor. *Journal of Molecular Biology*, 245(4):331–350, January 1995. 2.1.1, 2.1.3, 2.3.3, 3.3.2, 6.2.4

[132] Bernard Schmitt and Claude Reiss. Kinetic study *in vitro* of *Escherichia coli* promoter closure during transcription initiation. *Biochemical Journal*, 306:123–128, 1995. 6.3.1

[133] Takashi Sera and Carla Uranga. Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table. *Biochemistry*, 41(22):7074–7081, 2002. 3.4.1

[134] E. Brooks Shera, Newton K. Seitzinger, Lloyd M. Davis, Richard A. Keller, and Steven A. Soper. Detection of single fluorescent molecules. *Chemical Physics Letters*, 174(6):553–557, November 1990. 2.3.1

[135] Igor Sidorenkov, Natalia Komissarova, and Mikhail Kashlev. Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Molecular Cell*, 2:55–64, July 1998. 2.2.4

[136] Ulrich Siebenlist, Robert B. Simpson, and Walter Gilbert. *E. coli* RNA polymerase interacts homologously with two different promoters. *Cell*, 20(2):269–281, June 1980. 2.1.2

[137] Nicole Sommer, Vida Salniene Egle Gineikiene, Rimas Nivinskas, and Wolfgang Rüger. T4 early promoter strength probed *in vivo* with unribosylated and ADP-ribosylated *Escherichia coli* RNA polymerase: a mutation analysis. *Microbiology*, 146:2643–2653, 2000. 2.1.2, 2.3.2, 6.3.5

[138] Mathias Springer. Translational control of gene expression in *E. coli* and bacteriophage. In Lin and Lynch [88], chapter 6, pages 85–126. 3.2.4

[139] James E. Stefano and Jay D. Gralla. Spacer mutations in the *lac* p^s promoter. *Proc. Natl. Acad. Sci. USA*, 79(4):1069–1072, February 1982. 2.1.2

[140] Vasily M. Studitsky, Konstantin L. Brodolin, Ye Liu, and Andrei D. Mirzabekov. Topography of *lac*UV5 initiation complexes. *Nucleic Acids Research*, 29(3):854–861, 2001. 2.2.1

[141] Won Chul Suh, Sigrid Leirmo, and M. Thomas Record, Jr. Roles of Mg^{2+} in the mechanism of formation and dissociation of open complexes between *Escherichia coli* RNA polymerase and the λP_R promoter: Kinetic evidence for a second open complex requiring Mg^{2+}. *Biochemistry*, 31(34):7815–7825, September 1992. 6.1.3

[142] Jennifer J. Sullivan, Keith P. Bjornson, Lawrence C. Sowers, and Pieter L. deHaseth. Spectroscopic determination of open complex formation at promoters for *Escherichia coli* RNA polymerase. *Biochemistry*, 36:8005–8012, 1997. 2.3.1, 6.2.1, 6.3.4

[143] Synthetic Biology. `http://syntheticbiology.org/`. 1.1

[144] Hong Tang, Younggyu Kim, Konstantine Severinov, and Alex Goldfarb Richard H. Ebright. *Escherichia coli* RNA polymerase holoenzyme: Rapid reconstitution from recombinant α, β, β', and σ subunits. In Sankar Adhya, editor, *Methods in Enzymology*, volume 273, pages 130–134. Academic Press, 1996. 4.2.1

[145] Dmitry Temiakov, Pamela E. Mentesana, Kaiyu Ma, Arkady Mustaev, Sergei Borukhov, and William T. McAllister. The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance. *Proc. Natl. Acad. Sci. USA*, 97(26):14109–14114, December 2000. 2.2.4

[146] James A. Triccasa, Warwick J. Britton, and Brigitte Gicquel. Isolation of strong expression signals of *Mycobacterium tuberculosis*. *Microbiology*, 147:1253–1258, 2001. 2.3.2

[147] Martin J. Tymms, editor. *In Vitro Transcription and Translation Protocols*, volume 37 of *Methods in Molecular Biology*. Humana Press, 1995. 2.3.3, 6.2.4, 57

[148] Andrea Újvári and Craig T. Martin. Thermodynamic and kinetic measurements of promoter binding by T7 RNA polymerase. *Biochemistry*, 35:14574–14582, 1996. 2.3.1, 3.3.1, 6.3.4

[149] S.M. Uptain, C.M. Kane, and M.J. Chamberlin. Basic mechansims of transcript elongation and its regulation. *Annual Review of Biochemistry*, 66:117–72, 1997. 2.2.4, 2.2.5, 6.1.3

[150] Ulla Vogel and Kaj Frank Jensen. The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *Journal of Bacteriology*, 176(10):2807–2813, May 1994. 2.3.2, 6.3.3

[151] Peter H. von Hippel. An integrated model of the transcription complex in elongation, termination, and editing. *Science*, 281(5377):660–665, July 1998. 2.2.4, 2.2.6

[152] Peter H. von Hippel, David G. Bear, William D. Morgan, and James A. McSwiggen. Protein-nucleic acid interactions in transcription: A molecular analysis. *Annual Review of Biochemistry*, 53:389–446, 1984. 2.2.3, 6.1.1, 6.1.3

[153] Daguang Wang and Robert Landick. Nuclease cleavage of the upstream half of the nontemplate strand DNA in an *Escherichia coli* transcription elongation complex causes upstream translocation and transcriptional arrest. *Journal of Biological Chemistry*, 272(9):5989–5994, February 1997. 3.3.2

[154] Daguang Wang, Timothy I. Meier, Cathleen L. Chan, Guohua Feng, Donna N. Lee, and Robert Landick. Discontinuous movements of DNA and RNA in RNA polymerase accompany formation of a paused transcription complex. *Cell*, 81(3):341–350, May 1995. 2.2.5

[155] Michelle D. Wang, Mark J. Schnitzer, Hong Yin, Robert Landick, Jeff Gelles, and Steven M. Block. Force and velocity measured for single molecules of RNA polymerase. *Science*, 282(5390):902–907, October 1998. 6.3.6

[156] Ron Weiss. *Cellular Computation and Communications using Engineered Genetic Regulatory Networks*. PhD thesis, MIT, 2001. 1.1

[157] Klausdieter Weller and Rolf-Dieter Recknagel. Promoter strength prediction based on occurrence frequencies of consensus patterns. *Journal of Theoretical Biology*, 171(4):355–359, December 1994. 6.3.5

[158] William B. Whitten, J. Michael Ramsey, Stephen Arnold, and Burt V. Bronk. Single-molecule detection limits in levitated microdroplets. *Analytical Chemistry*, 63(10):1027–1031, 1991. 2.3.1

[159] Kai Wilkens and Wolfgang Rüger. Characterization of bacteriophage T4 early promoters *in vivo* with a new promoter probe vector. *Plasmid*, 35(2):108–120, March 1996. 2.3.2, 6.3.1, 6.3.3, 6.3.3

[160] Kevin S. Wilson and Peter H. von Hippel. Stability of *Escherichia coli* transcription complexes near an intrinsic terminator. *Journal of Molecular Biology*, 244:36–51, 1994. 2.2.5, 2.2.6, 6.1.3

[161] Malcolm E. Winkler and Charles Yanofsky. Pausing of RNA polymerase during in vitro transcription of the tryptophan operon leader region. *Biochemistry*, 20(13):3738–3744, 1981. 2.2.5

[162] Da-Guang Xu and Thomas M. Nordlund. Sequence dependence of energy transfer in DNA oligonucleotides. *Biophysical Journal*, 78:1042–1058, February 2000. 6.1.3

[163] Daguang Xu, Kervin O. Evans, and Thomas M. Nordlund. Melting and premelting transitions of an oligomer measured by DNA base fluorescence and absorption. *Biochemistry*, 33(32):9592–9599, 1994. 6.3.4

[164] L. R. Yarbrough, Joseph G. Schlageck, and Michael Baughman. Synthesis and properties of fluorescent nucleotide substrates for DNA-dependent RNA polymerases. *Journal of Biological Chemistry*, 254(23):12069–12073, December 1979. 2.3.3

[165] Hong Yin, Irina Artsimovitch, Robert Landick, and Jeff Gelles. Nonequilibrium mechanism of transcription termination from observations of single RNA polymerase molecules. *Proc. Natl. Acad. Sci. USA*, 96(23):13124–13129, November 1999. 6.3.6

[166] Hong Yin, Robert Landick, and Jeff Gelles. Tethered particle motion method for studying transcript elongation by a single RNA polymerase molecule. *Biophysical Journal*, 67:2468–2478, December 1994. 6.3.6

[167] Hong Yin, Michelle D. Wang, Karel Svoboda, Robert Landick, Steven M. Block, and Jeff Gelles. Transcription against an applied force. *Science*, 270(5242):1653–1657, December 1995. 6.3.6

[168] Yohei Yokobayashi, Ron Weiss, and Frances H. Arnold. Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. USA*, 99(26):16587–16591, December 2002. 1.1, 3.4.5

[169] X. Zhang, P. Dennis, M. Ehrenberg, and H. Bremer. Kinetic properties of *rrn* promoters in *Escherichia coli*. *Biochimie*, 84(10):981–996, October 2002. 2.3.2

[170] Xiangyang Zhang and Hans Bremer. Effects of fis on ribosome synthesis and activity and on rRNA promoter activities in *Escherichia coli*. *Journal of Molecular Biology*, 259:27–40, 1996. 2.3.2, 3.3.2, 6.3.3